2019

Latent Variable Interactions and Quadratics

Robert Ping
Wright State University

Robert Ping

Latent Variable Interaction and Quadratic Research

(Displays best with Microsoft IE/Edge or Mozilla Firefox)

(HOME)

Please note: The entire site is now under construction.

Please send me an email at rping@wright.edu if something isn't working, especially the endogenous interaction workaround--I can provide it via email.

FOREWARD--This web site contains research concerning:

- **o** Latent variables (LV's), and LV interactions (XZ) and quadratics (XX) in theoretical (hypothesis testing) models involving real world survey data,

  and

- **o** Testing these models.

It is intended for Ph.D. students and researchers who are somewhat familiar with LV's and structural equation modeling (SEM) software such as LISREL, EQS, AMOS, SIMPLIS, etc., and who may be just getting started with estimating substantive (theory-testing) models with one or more LV interactions (XZ) or quadratics (XX); or for example, having difficulties with model-to-data fit, a measure with unacceptable Average Variance Extracted (AVE), estimating truly categorical variables (e.g., gender) in a structural equation model with LV's, a 3-way LV interaction (e.g., XZW), etc.

INTRODUCTION**--**This update emphasizes two perhaps surprising topics:

1) The **importance** of XZ and XX **reliability and validity** in theory testing with LV's, and

2) a procedure for deleting measure items to attain **model-to-data fit** that typically produces at least 2 **"about-6-item" subsets** that **fit the data**, and have **acceptable reliability and Average Variance Extracted (AVE)**.

These topics are important in LV theory testing because interaction and quadratic reliability and AVE (a facet of validity) are typically ignored--perhaps because they are sometimes unacceptably low. However, they frequently can be improved by simply reitemizing X and Z.

Interaction and quadratic reliability and validity are important in LV theory tests because if XZ is unreliable, it is *per se* invalid. Further, lack of XZ Convergent Validity, for example, may threaten any conclusion based on XZ's significance. Specifically, a significant XZ-Y association may be NS (falsely disconfirmed) in a subsequent study because the present study's significant XZ-Y association may be an artifact (the result) of XZ's measurement error. Further, lack of XZ's Content (Face) Validity, may call into question whether or not XZ's moderation hypothesis was actually tested. To explain, if XZ is not Face Valid (it is not itemized as

$$XZ = (x1 + x2 +...+ xn)*(z1 + z2 +...+ zm)$$

$$= x1z1 + x1z2 +... + x1zm$$
$$+ x2z1 + x2z2 +,,, + x2zm$$
$$+...+ xnz1 + xnz2 +... + xnzm) ,$$

where xizj is the product of the indicators Xi and Zj, the hypothesized X moderation of Z-Y in the model

Y=... + aX + bZ + cXZ + ...

= ... + [a+cX]Z + bz + ...

(i.e., the coefficient of Z depends on--is moderated by--values of X in the coefficient [a+cX]) (see Aiken and West 1991) has not been adequately tested by the model because XZ cannot be factored into [a+cX]Z--the Z in bZ is not itemized the same way as the Z in cXZ.

To find out more see "**W**hy are reviewers asking about reliability and validity in my XZ interaction, and is there any way to improve XZ reliability and validity?" in ***QUESTIONS OF THE MOMENT*** below.

FINDING CONSISTENT, VALID AND RELIABLE ITEMIZATIONS--As a colleague once quipped, "what has happened to more-than-three-item measures these days?" The answer may be simple: Any three-item measure is *a priori* consistent (i.e., it always fits the data) and one only needs to find three items that are valid and reliable.

Kidding aside, theory testers know that while there are procedures to find a reliable subset of items, finding one that is 1) reliable, 2) face valid, 3) convergent and discriminant valid, plus 4) consistent can be a challenge.

Fortunately, there is a procedure that can find two or more subsets of items that usually attains 1) through 4) jointly. Specifically, experience suggests the resulting item subsets all "fit the data," they usually have near-maximum reliability and AVE, plus they typically are larger than 3 items.

This procedure is discussed further in ***QUESTIONS OF THE MOMENT*** below under

" **I**s there any way to speed up 'item weeding' and find a set of items that jointly fits the data, contains more than 3 items, and is valid and reliable?"

NEWS:

"The EXCEL spreadsheet for specifying XZ, XX and/or ZZ has been extensively revised to speed up XZ, etc. specification and estimation. In addition, as some have noticed the spreadsheet can be used to compute X and Z reliability and AVE, even when an interaction or quadratic is not being specified. (Please see ***EXCEL TEMPLATES*** below.)

"Workarounds" for an unbiased estimate of an "endogenous" interaction or quadratic with AMOS (and SIMPLIS) are available by e-mail. (AMOS and SIMPLIS do not allow the free (non zero) correlation(s) between an interaction, XZ, and its endogenous constituent latent variable X (and/or Z) (i.e., X and/or Z has an antecedent) that are required to adequately estimate an interaction or quadratic in real-world data).

A paper about **reusing one's data set** to create a second theory-test paper is available. It turns out that an editor may not object to a theory-testing paper that reuses data which has been used in a previously published theory-testing paper, if the new paper's theory/model is "interesting" and materially different from the previously published paper. The paper discusses how submodels from a previous paper might be found for a second paper that may not require collecting new data (to help reduce the "time between papers"). Please see "Notes on Used Data--Reusing a Data Set to Create A Second Theory-Test Paper" in ***SELECTED PAPERS...*** below.

A working-paper about **peer-revewing** a theory-test paper is available by e-mail. Its suggestions could be used as a check-list to help with peer-reviewing a (survey data) theory-testing paper. It might also be useful as a framework for a dissertation that tests a model using survey data, or for a theory-test paper prior to its submission for review.

A warning about **relying on "standardized loadings"** (latent variable (LV) loadings that are all free, so the resulting LV has a variance of unity) for significance estimates is available. (Standardized loadings may produce an incorrect t-value for a structural coefficient in real-world data, and any interpretation or implication of its significance or nonsignificance may be risky.) Please see "**W**hy are reviewers complaining about my use of standardized loadings?" in ***QUESTIONS OF THE MOMENT*** below.

Suggestions for estimating a "moderated quadratic" LV (i.e., XXZ) are available by email. LV's such as XXZ may seem unlikely, but for an LV (X) that exhibits diminishing returns (XX), its rate of return also may vary with the level of a moderator. (For example, the Satisfaction-Exiting association can be quadratic--when Satisfaction is high, further increases in Satisfaction are likely to produce fewer and fewer reductions in the likelihood of Exiting. However, this quadratic (i.e., upside-down horseshoe) Satisfaction-Exiting relationship also may depend on the availability of Alternatives: with lots of Alternatives, increases in Satisfaction may produce near-linear reductions to Exiting--i.e., the Satisfaction-Exiting association may no longer be quadratic--it may be linear.

The procedure for estimating **(truly) categorical variables** in a mixed model with (multiple indicator) LV's has been expanded. Please see "**H**ow does one estimate (truly) categorical variables in theoretical model tests using structural equation analysis?" in ***QUESTIONS OF THE MOMENT*** below.)

An annotated suggested format for substantive papers that may reduce the likelihood of paper rejection is available by email. (For example, a paper's customary Abstract-Intro-Lit Review, etc. can be insufficient to "hook" the reviewers, and pique their interest in the rest of the paper. The format suggests, in some cases, subtle changes in wording and content for these and other paper sections.)

A suggested approach for improving a **problem**

measure--one with a low Average Variance Extracted, or one that has been excessively weeded or with low model-to-data fit, etc.--also has been expanded. Please see "**H**ow are "Formative" (indicators-pointing-to) Latent Variables estimated with LISREL, EQS, AMOS, etc.?" in ***QUESTIONS OF THE MOMENT*** below.)

A suggested approach is available for estimating a 3-way interaction (e.g., XZW) that has some surprising results.

A three-way interaction may seem unlikely, but a proper test of hypotheses such as "H0: X-->Y is moderated by Z" and "H1: X-->Y also is moderated by W" is to specify XZ, XW **and XZW** in the model. (Z and W may moderate X-Y jointly via ZWX rather than separately (via ZX and WX). Stated differently, ZX and WX could be NS while ZWX is significant (ZW moderates X).

However, there are at least 4 nonequivalent specifications of XZW. And, the recommended specification used in regression is typically NS in real-world theory tests, while one or more of the other three 3-way specifications can be significant, and this difficulty appears to occur in latent variable specification as well. (Please see "**W**hy are the Hypothesized Associations not Significant? A *Three-Way Interaction*?" in ***SELECTED PAPERS...*** below.)

The suggestions for **improving AVE** have been expanded, also with some surprising results (e.g., some of the popular "discriminant validity" tests are not trustworthy in theory tests with real-world survey data). (Please see "**I**s there any way to improve Average Variance Extracted (AVE) in a Latent Variable X?" in ***QUESTIONS OF THE MOMENT*** below.)

A suggested approach is provided for remedying a "not Positive Definite" message (an "Ill Conditioned" message in exploratory factor analysis) that typically occurs when a full (unweeded) measure is factored, specified etc. for the first time. (Please see "**H**ow does one remedy a "not Positive

Definite" message?" in
QUESTIONS OF THE MOMENT.)

Comments on the use of PLS in theory tests involving survey data are available.
(Please see "**W**hy are reviewers complaining about the use of PLS in my paper?
in ***QUESTIONS OF THE MOMENT***.)

Specifications for an interaction form other than XZ are available by e-mail. This may not sound like much, but XZ is ***not the only form an interaction can take***--
X/Z and XZ^2 are also interactions. So, a nonsignificant XZ may not mean an hypothesized moderation (interaction) is disconfirmed.
Specifically, if the XZ-->Y association is NS there still may be a significant interaction--it just doesn't have the form "X times Z". Experience suggests that X/Z, XZ^2 or other interaction forms may be significant when the XZ is not.

Comments on the use of regression to test an hypothesized interaction are available.
(Please see "**W**hy are reviewers complaining about the use of moderated multiple regression in my paper? in ***QUESTIONS OF THE MOMENT***.)

The suggestions for estimating an endogenous interaction have changed. An EXCEL template for estimating the indirect effect of an endogenous interaction is available by e-mail. (Please see "**Please Note:** If you are estimating an interaction involving an **endogenous** variable..." in the ***INTRODUCTION***.)

A paper on hypothesizing interactions is available (i.e., what evidence suggests the presence of an interaction before data is collected, how an interaction might be justified (argued for), etc.). (Please see "Interactions May Be the Rule Rather than the Exception, But..." in ***SELECTED PAPERS...***below.)

A web site relocation has obsoleted many citations on this web site (and published papers) that reference http://home.att.net/~rpingjr/... Since that internet address no longer exists, these citations should be changed to http://www.wright.edu/~robert.ping/...

Coming Attractions:

- Suggestions regarding reviewer comments about "Implications," and "(Management) Recommendations" in a theory testing paper. (Please e-mail me for a progress report.)
  (Comments on handling reviewer comments in general are available on the Higher Education web page--please click on "Home" above, then click on "Higher Education.")

- An EXCEL template that provides a simple and straightforward tabular interpretation of a significant Interaction (similar to crosstabs) that does not require graphs (please e-mail me for a draft template).

Recent Additions and Changes:

- Suggestions for estimating a mixed SEM model with the customary "Reflective" (arrows-to-indicators) LV's, plus "Formative" (indicators-pointing-to) LV's are provided.
  This sounds like it would never be useful. Some may have been told this can not, or should not, be done with SEM (e.g., with LISREL, AMOS, EQS, etc.). However, a new measure (or an older, well established, measure developed before the advent of SEM) frequently requires substantial weeding in order to attain model-to-data fit. And, the weeded measure may be missing so many items that a reviewer might judge it to be no longer adequately face valid. Or, no itemization can be found with adequate Average Variance Extracted (AVE), or adequate discriminant validity, etc. Perhaps surprisingly, such measures usually can be re-specified as Formative--typically using the full (unweeded) measure--and estimated using

LISREL, AMOS, EQS, etc. in order to remedy these difficulties.

- **o** The "Why is my hypothesized interaction or quadratic nonsignificant?" paper (below) is being revised to account for interaction forms besides XZ;

- **o** A paper on using regression to test an hypothesized interaction that links to a paper titled "What is Structural Equation Analysis?" is provided.

- **o** The cubics paper is revised, and an EXCEL template for specifying cubics is provided; and

- **o** Several EXCEL templates calculate reliability and Average Variance Extracted for XZ, XX and ZZ. (The quadratic XX, for example, could be viewed as the interaction of X with itself.).

Don't forget to Refresh: If you have visited this web site before, and the latest "**Updated**" date (at the top of the page) seems old, or, if you are actively estimating an interaction, etc., you may want to be sure you are viewing the current version of this web page. For Internet Explorer, click on "Tools" (above), then click on "Delete Browsing History," "check" "Temporary Internet Files," uncheck everything else, and click on delete. After that, close this browser window, then re-launch it. (The procedures for Firefox and Chrome are discussed next in refreshing WORD documents.)

In adddition, many of the links on this web site are in Microsoft WORD.
If you have viewed one or more of them before, the procedure to view the latest
(refreshed) version of them is tedious ("Refresh" does not work for WORD documents
on the web). With my apologies for the tediousness, to refresh any (and all) Word
documents in Internet Explorer, please follow the above procedure.

To refresh all WORD documents in Chrome, please click on the 3 "bars" in the upper
right-hand corner of the screen, then click on "History." Next, click on "Clear
browsing data," check "Cashed images and files," and uncheck everything else.
Finally, click on "Clear browsing data." After that, close this browser window,
then re-launch it so the latest versions of all the WORD documents are forced to
download.

To refresh WORD documents in Firefox, please click on "Tools" on the browser toolbar
(above), then click on "Options." Next, in the "Advanced" tab, find the "Network" tab,
and under "Cached Web Content" click on "Clear Now" Then, click "OK." After
that, close this browser window, then re-launch it.
If several browsers are used, they all should be refreshed.

Your questions are encouraged; just send an e-mail to me at rping@wright.edu. Don't
worry about being an expert in latent variables, structural equation modeling or
structural equation analysis, or using "correct" terminology or perfect English.

A **Table of Contents** or **Index** to this website is not available. With my apologies,
please use your browser's search capability instead. For example, to find the EXCEL
templates try Ctrl+F. Or, click on "Edit" on the browser toolbar (above) and click on
"Find" (or click on the three horizontal "bars" in the upper right-hand corner of Chrome,
then click "Find..."), and type the word "excel" in the find box.

REMARKS...

Please Note: If you are estimating an interaction involving one or more **endogenous** variables please click here--things are different in this case.

ALSO, please remember that interactions and quadratics should be **hypothesized before data is ever collected**. Please be aware that the only situations where one should search for significant interactions or quadratics after the data is collected are:

1) when one wishes to explain a non-significant (NS) model association
(e.g., is there an unexpected interaction or quadratic suppressing the hypothesized
Z-->Y association?), or

2) where one wishes to "probe" an important significant association(s) to see
if it is also conditional (i.e., moderated by some other variable), and provide
a "finer grained" interpretation. (ANOVA for example, routinely provides "all
possible interactions to help interpretations of significant effects. Although
this is rarely done in LV model tests, such "post hoc" probing could suggest,
for example, that while Z-->Y was significant in the model test, the
association was stronger (or weaker) at various levels of X, which might
add "interest" to a paper's Discussion, Implications and Conclusions sections.)

In the first case, any significant "suppressor" interaction XZ or quadratic ZZ found after the data is collected could be offered as a possible explanation for the hypothesized but observed NS Z-->Y association in the Discussion section of the paper.
In the second case, any significant association that is "discovered" to actually conditional (moderated) after the data is collected could be the basis for noting that the significant Z-->Y association was "supported" only for some levels of X in the study (see "Interpreting Latent Variable Interactions" in the ***SELECTED PAPERS...***section below).

Any of these moderated Z-->Y association(s) discovered after the fact could be hypothesized in the Discussion section, for testing in a follow-up study, or replication (to investigate whether or not the "after-the-fact" moderation simply was significant by chance in the present study) (see "Hypothesized Associations and Unmodeled Latent Variable Interactions/Quadratics: An F-Test..." in the ***SELECTED PAPERS...***section below for more).

In the absence of situations 1 or 2 above, hunting for significant interactions or quadratics that were not hypothesized before the data was collected could lead to "poor science." It can tempt one to add an interaction or quadratic hypothesis to the model as though it were hypothesized before the data was collected. This changes a confirmatory study into an exploratory study. Specifically, adding to a model an interaction or quadratic that was discovered after the data was collected, changes one's "hypotheses-before-first-test" model-test (confirmatory) study into an exploratory study where part of the model was unknown before data collection. Again, the proper approach is to put any interaction(s)/quadratic(s) discovered after the data was collected in the Discussion section, noting that they were discovered after the data was collected, and perhaps arguing for their disconfirmation in the next study.

QUESTIONS OF THE MOMENT

"**W**hat about the alternative specifications for a Latent Variable (LV) interaction?"

An informal review in 2005 of substantive Social Science journal articles written since Kenny and Judd's (1984) seminal proposal for specifying LV interactions and quadratics found that the most frequently encountered specifications for LV interactions in substantive articles were: Jaccard and Wan (1995) (which specifies a 4-product-indicator subset[1] of the Kenny and Judd interaction (product) indicators to avoid the model-to-data fit problems that occur when all of the Kenny and Judd indicators are specified--46 citations); Mathieu, Tannenbaum and Salas (1992) (which has not been formally evaluated for possible bias and inefficiency (see Cortina, Chen and Dunlap 2001 for other difficulties)--51 citations); and Ping (1995) (41 citations). (See FAQ's A, B and C in the ***FREQUENTLY ASKED QUESTIONS*** section below for more.)

---

[1] Unfortunately, in theoretical model tests, deleting ("weeding" out) all but 4 of the Kenny and Judd product indicators to attain model-to-data fit raises many questions, including, what is the reliability and validity of the resulting 4-item interaction or quadratic? Reliability is necessary for the validity of all the LV's in a model, but the reliability of an interaction specified with nearly all of its indicators absent is unknown. (The formula for the reliability of the interaction of X and Z assumes XZ is operationally (unweeded) X times (unweeded) Z.)

The face- or content-validity of a 4-item interaction or quadratic also is questionable--if nearly all the indicators of X and Z are not represented in the itemization XZ, is XZ still the LV (unweeded) X times (unweeded) Z?.

Further, it is easy to show that in real-world data a weeded XZ's structural coefficient varies with its indicators. Unfortunately, the "best" set of four indicators is unknown.

Finally, an interaction with weeded Kenny and Judd product indicators cannot be "factored" to produce detailed interpretation because XZ is no longer (unweeded) X times (unweeded) Z.

"**I**s there an example that shows all the steps involved in estimating a latent variable interaction/quadratic?"

(Please click here for more.)

"**W**hy are reviewers complaining about my use of standardized loadings?"

It turns out that standardized loadings (latent variable (LV) loadings specified as all free so the resulting LV has a variance of unity) may produce incorrect t-values for some parameter estimates, including structural coefficients. This presents a problem for theory testing: An incorrect (biased) t-value for a structural coefficient means that any interpretation of the structural coefficient's significance or nonsignificance versus its hypothesis may be risky.
(Please click here for more.)

"**H**ow does one estimate (truly) categorical variables in a model with LV's?"

In the popular structural equation analysis modeling (SEM) software (e.g., LISREL, EQS, Amos, etc.), the term "categorical variable" usually means an ordinal variable (e.g., an attitude measured by Likert scales), rather than a "truly" categorical (i.e., nominal) variable (e.g., Marital Status, with the categories Single, Married, Divorced, etc.), and typically there is no provision for "truly" categorical variables. In regression, a (truly) categorical variable is estimated using "dummy" variables with regression through the origin, but a similar approach currently is not available using the popular SEM software. However, a "mixed SEM" approach for estimating categorical variables and LV's with their measurement errors is available here. (There also is a working paper available via e-mail with a "workaround" alternative to the paper on the "hot spot" just provided (categorical3.doc) that might be useful for a theory test.)

"**W**hy are reviewers complaining about the use of moderated multiple regression in my paper?"
(Please click here for comments on this matter.)

"**H**ow should PRELIS or similar "preprocessor" software be used with LISREL, EQS, AMOS, etc. to create interactions/quadratics?"
(Please click here for suggestions.)

"**W**hy should Applied Researchers be interested in interactions/quadratics?"

Contrary to customary practice, interactions and quadratics also may be important in applied model-building/research--model building in econometrics, epidemiology, market response models, biostatistics, etc.--not to explain additional variance in a target variable, but to better understand, explain and predict important relationships in the model.

Anecdotally, it may not be widely appreciated in applied research that important model predictors may not be "unconditional" (e.g., Y is unconditionally increased/decreased with X in the study). Instead, these effect may have been "conditional" (e.g., Y

increased with X when Z was at a high level (strong), but Y was unrelated to X, or it decreased with X, when Z was at a lower level (weaker)).
(Please click here for more.)

"**W**hy are reviewers asking about reliability and validity in my interaction, XZ, and is there any way to improve XZ reliability and validity?"
(Please click here for more.

"**W**hen theory proposes that Z moderates the X-->Y association, but theory is mute about a Z-->Y association, why does one still include the Z-->Y association, in addition to the X-->Y and XZ-->Y associations, in the model?"

Excluding the Z-->Y (or the X-->Y) association when XZ-->Y is hypothesized to be significant, biases all structural coefficients and standard errors in the proposed model EVEN WHEN THE Z-Y ASSOCIATION SHOULD NOT EXIST IN THE POPULATION.
(Please click here for more.)

"**H**ow is a cubic latent variable (LV) estimated?"

Please see the paper on estimating a LV cubic in the ***SELECTED PAPERS...*** section (below). The "EXCEL templates..." section (below) also includes a template to assist in calculating LV cubic loadings, error variances, etc.

"**W**hy is my hypothesized interaction or quadratic nonsignificant?"
(Please click here for a paper on this topic that is being revised to include other interaction forms (e.g., X/Z)--please e-mail me for more.)

"**I**s there any way to improve Average Variance Extracted (AVE) in a Latent Variable X?"
(Please click here for a paper on this matter. Also, please consider reading "How are Formative Latent Variables estimated with LISREL, EQS, AMOS, etc.?", below, for more suggestions for "troubled" LV's--unacceptable AVE's, etc.)

"**I**s there any way to speed up "item weeding" and find a set of items that jointly fits the data, contains more than 3 items and is valid and reliable?"
(Please click here for suggestions.)

"**H**ow might a 'mixed interaction' XZ, where X is a manifest/observed/continuous/single-indicator, etc. variable (not a latent variable), be estimated?"
(Please click here for suggestions.)

"**W**hy is my hypothesized interaction significant using a 'median split' of the data, or a '2-group analysis,' but not

significant when specified in my model?"
(Please click here for comments on this subject.)

"**W**hy are most (or all) of my hypothesized interactions not significant?"
(Please click here for more on this matter.)

"**W**hat is the Average Variance Extracted (AVE) for a Latent Variable Interaction (or Quadratic)?"
(Please click here for comments on this topic, then please e-mail me--I have more
suggestions.)

"**W**hat is the 'validity' of a Latent Variable Interaction (or Quadratic)?"
(Please click here for more on this subject.)

"**H**ow does one remedy a "not Positive Definite" message?"
(Please click here for suggestions.)

"**W**hy are reviewers complaining about the use of PLS in my paper?"
(Please click here for comments on this topic.)

"**H**ow are "Formative" (indicators-pointing-to) Latent Variables estimated with LISREL, EQS, AMOS, etc.?"

> **T**his may seem uninteresting--formative variables are rare in the Social Sciences. But, a formative re-specification of a difficult measure (e.g., one with inadequate AVE; or, excessive weeding was required to attain model-to-data fit and it could now be judged to be face invalid; or, it has unacceptable discriminant validity; etc.) may be the only alternative to abandoning the measure (and deleting the LV from the model).
> This approach also may be useful when using older, well established, measures--developed before the advent of SEM--as they were intended (i.e., without weeding-out most of the items).
> (Please click here for more.)

FREQUENTLY ASKED QUESTIONS

(CLICK ON A RED DOT)

FREQUENTLY ASKED QUESTIONS (FAQ'S) ABOUT LATENT VARIABLE INTERACTIONS AND QUADRATICS IN SURVEY DATA

E.g., The answer to FAQ D, "How does one test hypothesized interactions or
quadratics?" contains step-by-step instructions for Ph.D. students, and
theoretical or applied researchers, interested in estimating their first Latent

Variable Interaction or Quadratic in a theoretical model test using survey data.

EXCEL TEMPLATES

Spreadsheets for expediting the specification of Latent Variable Interactions,
Quadratics and cubics; for "weeding" measures to attain model fit; for Latent Variable
Regression, etc.

(CLICK ON A RED DOT)

For **specifying** a single indicator LV **Interaction** or Quadratic using Direct
(LISREL 8, CALIS, etc.) estimation, or "2-Step" estimation using LISREL,
EQS, AMOS, etc. (See Ping 1995, *JMR*, and Ping 1996, *Psych Bull.*--revised
versions of which appear below in ***SELECTED PAPERS...*** .

The template also calculates LV Interaction or Quadratic **Reliability** and
Average Variance Extracted (AVE). More about the template.

For **specifying** a single indicator LV **Cubic** using "2-Step" estimation and
LISREL, EQS, AMOS, etc.) (please also see "Notes on Estimating Cubics and other
'Powered' Latent Variables" below).

For deleting items from a multi-item measure so it fits the data
(i.e., **finding a set of items that fits the data,** so the measure is internally
consistent).

In real-world data, there frequently are many subsets of a multi-item
measure that will fit the data, and this begs the question, which of these
subsets is "best" from a validity standpoint? The template helps find at least
one subset of items, usually with a maximal number of items (typically
different from the one found by maximizing reliability, for example), that will
fit the data.
The template then can be used to search for additional subsets of items
that will also fit the data, and thus it may help to find the "best" face- or
content-valid subset of items in a measure.
More about the template.

For Kenny and Judd (1984) multiple indicator specification with LISREL, EQS,
AMOS, etc. (see Ping 1996, *Psych. Bull.;* a revised version appears below).

This approach is useful with a consistent subset of product indicators (see
Chapter VIII.--SxA Unidimensionalization, in the monograph, *LATENT*
VARIABLE INTERACTIONS... below).
More about the template .

For Latent Variable Regression, a measurement-error-adjusted regression
approach to Structural Equation Analysis, for situations where regression
is useful (e.g., to estimate nominal/categorical variables with LV's) (see
Ping 1996, *Multiv. Behav. Res.*, a revised version appears below).
More about the template.

BIBLIOGRAPHY

A **Bibliography** on Latent Variable Interactions and Quadratics.
(Now somewhat out of date, it is being revised.)

ON-LINE MONOGRAPHS

(CLICK ON A RED DOT)

LATENT VARIABLE INTERACTIONS AND QUADRATICS IN SURVEY DATA: A SOURCE BOOK FOR THEORETICAL MODEL TESTING (2nd Edition)

The on-line monograph concernes Latent Variable Interactions and Quadratics, and their estimation, with examples. Potentially of interest to Ph.D. students and researchers who conduct or teach theoretical model (hypothesis) testing using survey data. It includes a "fast start" section on estimating a latent variable interaction, a section on estimating multiple interactions and quadratics, how to interpret a significant interaction or quadratic, and pedagogical examples (232 pp.).

The APA citation for this on-line monograph is Ping, R.A. (2003). *Latent variable interactions and quadratics in survey data: a source book for theoretical model testing, 2nd edition*. [on-line monograph]. http://www.wright.edu/~robert.ping/intquad2/toc2.htm.

Ping (2001) *LATENT VARIABLE INTERACTIONS AND QUADRATICS IN SURVEY DATA: A SOURCE BOOK FOR THEORETICAL MODEL TESTING* (Edition 1)

TESTING LATENT VARIABLE MODELS WITH SURVEY DATA (2nd Edition)

The on-line monograph presents the results of a large study of theoretical model (hypothesis) testing practices using survey data, with critical analyses, suggestions and examples. Potentially of interest to Ph.D. students and researchers who conduct or teach theoretical model testing using survey data. Its contents include the six steps in theoretical model (hypothesis) testing using survey data; a useful, but little-used research design that combines an experiment with a survey, scenario analysis; alternatives to dropping items from a measure to attain its model-to-data fit; inadmissible structural model solutions with remedies; interactions and quadratics; and pedagogical examples (177 pp.).

Of particular interest recently is how to efficiently and effectively delete items to attain a consistent measure (see STEP V, PROCEDURES FOR ATTAINING...).

The APA citation for this on-line monograph is Ping, R.A. (2004). *Testing latent variable models with survey data, 2nd edition.* [on-line monograph]. http://www.wright.edu/~robert.ping/lv1/toc1.htm .

- Ping (2002) *TESTING LATENT VARIABLE MODELS WITH SURVEY DATA* (Edition 1)

SELECTED PAPERS ON LATENT VARIABLES, AND THEIR INTERACTIONS AND QUADRATICS

(CLICK ON A RED DOT)

"Notes on 'Used Data'--Reusing a Data Set to Create A Second Theory-Test Paper"
(An earlier version of "Notes on 'Used Data'" (Ping 2013), *Am. Mktng. Assoc. (Summer) Educators' Conf. Proc.*).

The paper critically discusses an intriguing possibility: that one's data set might be re-used in a second theory-test paper (which could reduce the time and expense associated with gathering new data for an additional paper). It turns out that an editor might not object to reviewing a paper that is based on data which has been used in a previously published paper. It also turns out that a published model is likely contain at least one submodel that might be a candidate for an additional paper. The paper on reusing data discusses how "interesting" sub-models from a previous paper might be found, and issues that might arise.

"Why are the Hypothesized Associations

not Significant? A *Three-Way Interaction?*" (An earlier version of Ping 2010, *Am. Mktng. Assoc. (Winter) Educators' Conf. Proc.*).

The paper suggests en approach for estimating a 3-way interaction (e.g., XZW) with some surprising results. While a three-way interaction may seem unlikely, a proper test of hypotheses such as "H0: X-->Y is moderated by Z" and "H1: X-->Y also is moderated by W" is to specify XZ, XW **and** XZW in the model. (ZX and WX could be NS while ZW**x**X could be significant (ZW moderates X). The surprises include that there are at least 4 nonequivalent specifications of XxZxW (the recommended regression specification) is typically NS in real-world theory tests (while one or more of the other three 3-way specifications, such as Xx(ZxW), can be significant), and this (overlooked) regression difficulty may occur in latent variables as well.

- "But what about Categorical (Nominal) Variables in Latent Variable Models?" (An earlier version of Ping 2009, *Am. Mktng. Assoc. (Summer) Educators' Conf. Proc.*).

In part because categorical variables almost always are measured in surveys in the Social Sciences (e.g., "Demographics"), the paper suggests a procedure for estimating nominal (categorical) variables in a structural equation model that also contains latent variables.

- "Interactions May Be the Rule Rather than the Exception, But...: A Note on Issues in Estimating Interactions in Theoretical Model Tests" (An earlier version of Ping 2008, *Am. Mktng. Assoc. (Summer) Educators' Conf. Proc.*).

The paper critically addresses theory-testing matters concerning conceptualizing, estimating and interpreting interactions in survey data. For example, (before data is collected) what evidence suggests that an interaction should be hypothesized in a proposed model? Is an interaction a construct, or a mathematical form, or both? Is specifying the interaction between X and Z, for example, as XZ a sufficient disconfirmation test (i.e., are there other plausible interaction forms besides XZ?)

- "On the Maximum of About Six Indicators per Latent Variable with Real-World Data." (An earlier version of Ping 2008, *Am. Mktng. Assoc. (Winter) Educators' Conf. Proc.*).

The paper suggests an explanation and remedies for the puzzling result that Latent Variables in theoretical model testing articles frequently have a

maximum of about
6 indicators.

- "Second-Order Latent Variable Interactions, and Second-Order Latent Variables." (An earlier version of Ping 2007, *Am. Mktng. Assoc. (Winter) Educators' Conf. Proc.*).

  The paper proposes several specifications for a Second-Order Latent Variable interaction.

- "Notes on Estimating Cubics and other 'Powered' Latent Variables." (An earlier version of Ping 2007, *Am. Mktng. Assoc. (Summer) Educators' Conf. Proc.*).

  The paper discusses "satiation" and "diminishing returns," infrequently explored topics in theoretical model tests, along with a Latent Variable (LV) that is related to a quadratic, a cubic. The paper suggests a specification for this difficult-to-specify LV.

- "Estimating Latent Variable Interactions and Quadratics: Examples, Suggestions and Needed Research" (An earlier version of Ping 1998, in *Interaction...*, revised December 2006).

  The paper provides estimation examples, including LISREL and EQS code. The revision also corrects several errors.

- "Pseudo Latent Variable Regression: an Accessible Estimation Technique for Latent Variable Interactions," (An earlier version of Ping 2003, *2003 Acad. of Mktng. Sci. Conf. Proc.*, Miami: Acad. of Mktng Sci., revised October, 2003).

  The paper proposes a reliability-based regression estimator for Latent Variable Interactions and Quadratics.

- "Improving the Detection of Interactions in Selling and Sales Management Research" (An earlier version of Ping 1996, *J. of Personal Selling and Sales Mgt.*, revised October 2003).

  Using Monte Carlo simulations, the paper evaluates non-structural equation analysis approaches to detecting a Latent Variable Interaction such as median splits.

- "Interpreting Latent Variable Interactions" (An earlier version of Ping 2002, *Am. Mktng. Assoc. (Winter) Educators' Conf. Proc.*, revised June 2002).

The paper suggests an approach to developing detailed interpretation of a significant Latent Variable Interaction or Quadratic that is suitable for model tests--reveals the regions of the domain of a moderated variable where it is significant and non significant.

- "A Parsimonious Estimating Technique for Interaction and Quadratic Latent Variables" (An earlier version of Ping 1995, *JMR,* revised July 2001).

  The paper proposes a single indicator specification for Latent Variable Interactions and Quadratics that addresses the model-to-data fit problem associated with specifying these variables in real-world data with all their Kenny and Judd interaction items and thus impairing model-to-data fit; the proposed specification can be used with LISREL, EQS, AMOS, CALIS, etc.

- "Latent Variable Interaction and Quadratic Effect Estimation: A Two-step Technique Using Structural Equation Analysis" (An earlier version of Ping 1996, *Psych. Bull.,* revised July 2001).

  The paper proposes a "2-step" Kenny and Judd (1984) estimation approach for a Latent Variable Interaction (XZ) or a Quadratic (WW) with LISREL, EQS, AMOS, CALIS, etc. This approach is useful when X, Z and W each fit their single construct measurement model (i.e., they are internally consistent). (Otherwise, experience suggests that in real-world data the approach produces a structural coefficient for XZ, for example, that has an inflated standard error.)

- "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients" (An earlier version of Ping 1996, *Multiv. Behav. Res.,* revised July 2001).

  The paper proposes a measurement-error-adjusted regression technique for Latent Variables, including Interactions and Quadratics (the Standard Error is explained in the paper below, "A Suggested Standard Error..."). This approach is useful in situations where regression is useful. These situations may include model building (e.g., in market research, econometrics, epidemiology, biostatistics, etc.) where many candidate models are evaluated using "stepwise" and "backward elimination" procedures to determine the model(s) that explains the most variance in the calibration data; and, theoretical model (hypothesis) testing of a model that combines nominal (categorical) variables with other latent variables (see for example, "But what about Categorical (Nominal) Variables in Latent Variable Models?" above).

- "A Suggested Standard Error for Interaction

Coefficients in Latent Variable Regression" (An earlier version of Ping 2001, *Acad. Mktng. Sci. Proc.*, revised September 2001).

The paper suggests a Standard Error term for Latent Variable Regression (see "Latent Variable Regression: A Technique..." above).

WORKING PAPERS MENTIONED ELSEWHERE ON THE WEB SITE

(CLICK ON THE RED DOT)

"Hypothesized Associations and Unmodeled Latent Variable Interactions/ Quadratics: An F-Test, Lubinski and Humphreys Sets, and Shortcuts Using Reliability Loadings."

The paper proposes an approach to post-hoc probing for Latent Variable Interactions and Quadratics in order to explain an hypothesized association that is non significant.

The APA citation for this on-line paper is Ping, R.A. (2006). "Hypothesized Associations and Unmodeled Latent Variable Interactions/Quadratics: An F-Test, Lubinski and Humphreys Sets, and Shortcuts Using Reliability Loadings." [on-line paper]. http://www.wright.edu/~robert.ping/Ftest10.doc .

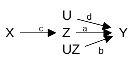
(HOME)

ESTIMATING ENDOGENOUS INTERACTIONS

(The **APA citation** for this paper is Ping, R.A. (2009). "Estimating endogenous interactions." [on-line paper]. http://www.wright.edu/~robert.ping/ Endog_ints2.doc)

(Earlier versions of this paper, Ping, R.A. (2008). "Estimating endogenous interactions." [on-line paper]. http://www.wright.edu/~robert.ping/ Endog_ints1.doc, and Ping, R.A. (2005). "A status report on estimating endogenous interactions." [on-line paper]. http://www.wright.edu/~robert.ping/Endog_ints.doc, are available here.)

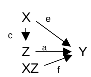
In the structural model with endogenous effects,



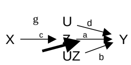
,

(Figure 1)

UZ could be called an "endogenous interaction" because one of its constituent LV's, Z, is endogenous. Endogenous interaction models also include,



(Figure 2),
and



.

(Figure 3)

These (sub)models could also be part of a larger model.

There may be four (additional) interaction considerations for an endogenous interaction in theory testing using structural equation software (e.g., LISREL, EQS, Amos, etc.): theoretical, specification, and estimation software considerations; and methodological and interpretational considerations.

The **theoretical considerations** will be discussed first. Anecdotally, some researchers believe that a moderator LV, Z, can not also be a mediator LV. Specifically, in Figure 2 Z should not moderate the X --> Y association with the XZ interaction, and also mediate the X --> Y association.

However, Z in Figure 2, as both a mediator and a moderator of the X-Y association, for example, is theoretically plausible. For example, Relationship satisfaction (SAT) and the attractiveness of alternative relationships (ALT) are well known in several social science literatures to be antecedents of relationship exiting (EXIT) (SAT --> EXIT and ALT --> EXIT, see Figure 4) (e.g., Ping 1993).

SAT

ALT EXIT

SATxALT

(Figure 4)

However, Johnson and Rusbult (1989) proposed that satisfaction "devalues" or reduces the attractiveness of alternatives (i.e., SAT --> ALT). In addition, Ping (1994) argued that satisfaction and alternative attractiveness interact in their effect on relationship exiting (i.e., that alternative attractiveness suppresses the satisfaction association with relationship exiting) (SATxALT --> EXIT). Thus, with mediation, satisfaction is likely to reduce alternative attractiveness, which then is likely to reduce exiting. With moderation, the interaction, increased alternative attractiveness decreases the strength of the (indirect) satisfaction-exiting association.

Thus, alternative attractiveness (ALT) may be both a mediator and a moderator of satisfaction's (SAT) effect on relationship exiting (EXIT), and in at least one model, joint mediation and moderation (and an endogenous interaction) are theoretically plausible. Moreover, I have tested this model in an unpublished paper using survey data, and all four paths were significant. Thus, mediation and moderation are both theoretically and empirically plausible.

Regarding **specification considerations**, in Figure 1 the interaction UZ should be specified as correlated with its constituent LV's, U and Z to reflect the fact that in real-world (non-normal) survey data, UZ is usually significantly correlated with U and Z, even when U and Z are mean- or zero-centered.[1] Failure to do so reduces model-to-data fit, and biases structural model parameter estimates (e.g., variances, covariances and structural coefficients). The same is true for XZ and UZ in Figures 2 and 3. However, it is not possible to correlate UZ in Figure 1, for example, directly with endogenous Z in SEM. In this case, a correlation specified between UZ and Z actually specifies a correlation between UZ and the structural disturbance for Z, ζ_Z.

However, experience suggests that not specifying the correlation between ζ_Z and Z also reduces model-to-data fit, and it biases the structural model's parameter estimates in real world data. Stated differently, experience suggests the UZ-ζ_Z correlation is typically significant in real world data, and not specifying it usually misspecifies the structural model (the misspecification can be verified using the "Modification Indices for PSI" (LISREL) for the UZ-ζ_Z correlation). In addition, experience suggests that any misspecification bias increases as the R^2 ("Squared Multiple Correlations for Structural Equations" in LISREL) of the interaction's predicted LV (e.g., Y in Figures 1, 2 and 3) declines.

[1] If X and Z are normally distributed, their correlations with XxZ are zero.

Parenthetically, specifying the UZ-ζ_Z correlation, for example, could be argued to be equivalent to correlating UZ with itself (some of the variance of UZ is included in ζ_Z). However, experience suggests that this does not appear to matter in an interaction, UZ for example, with centered constituent LV's U and Z. Specifically, experience suggests that the UZ-ζ_Z correlation, for example, typically are comparatively moderate (e.g., in the 0.10 to 0.20 range in absolute value).

In addition, specifying an exogenous correlation with endogenous structural disturbance is plausible in theoretical model tests with survey data. The LV's in these models typically are all correlated, and it is reasonable to assume that unmodeled LV's (i.e., those in ζ_Z) also should be correlated with the other model LV's.

However, correlations with structural disturbances in general must be specified with caution to avoid violating the assumption that a predictor is not (directly or indirectly) correlated with its predicted variable's prediction error (its structural disturbance). Specifically, in Figure 1, X should not be specified as both a predictor, and correlated, with Z (ζ_Z actually), and Y (ζ_Y actually) should not be specified as correlated with Z, U or UZ.

Similarly, in Figure 2, XZ cannot be directly correlated with Z in SEM, and in Figure 3, UZ cannot be directly correlated U or Z in SEM. However, because they are typically significant, and not specifying them reduces model-to-data fit and introduces structural coefficient bias, in Figure 2 the XZ-Z (XZ-ζ_Z actually) correlation should be specified. For similar reasons, in Figure 3 the UZ-U (UZ-ζ_U actually) and UZ-Z (UZ-ζ_Z actually) correlations also should be specified.

In summary, an interaction should be correlated with its constituent LV's. If one or both constituent LV's are endogenous, the interaction should be correlated with its constituent LV structural disturbances (ζ's, or PSI's in LISREL).

Regarding **estimation software considerations**, from e-mails I have received, AMOS may not allow the specification of the correlation an exogenous interaction and its endogenous LV that is required in real-world survey data. For example, in Figure 1 the UZ-Z (UZ-ζ_Z actually) correlation is not allowed. ~~At present, this restriction cannot accurately be "worked around," [2] and~~ LISREL, EQS, etc. ~~must~~ should be used to avoid the difficulties mentioned above when an interaction ~~is~~ can not be correlated with its endogenous constituent LV's structural disturbance(s). Alternatively, a "workaround" is available by email.

Parenthetically, correlations between exogenous and endogenous LV's in LISREL require the use of Submodel 3B.

Regarding **methodological considerations**, the reader should consider reading the remarks in "FAQ D" under "FREQUENTLY ASKED QUESTIONS" on the previous web page, and "Is there an example that shows all the steps involved in estimating a latent variable interaction/quadratic?" under "QUESTIONS OF THE MOMENT," also on the previous web

[2] Obvious strategies such as "2 group analysis" are well known to be untrustworthy for testing an interaction in survey-data theory tests because of, for example, reduced of statistical power in each of the structural coefficients, and difficulty gauging and interpreting non-interaction effects because they are frequently not identical between the two groups.

page. In particular, successfully estimating an interaction requires attention to centering, model specification, admissible values, etc. In addition, with endogenous interactions, experience suggests that not manually providing starting values for the structural model *may* produce unreasonable parameter estimates. Stated differently, experience suggests that an important step in successfully estimating one or more endogenous interactions, is to verify the resulting structural model parameter estimates. In particular, structural model parameter estimates (loadings, measurement error variances, LV variances and covariances) should approximate their measurement model values. In addition, a regression model of the interaction (e.g., XZ, X, Z, for example, and their predicted variable, Y) should produce unstandardized regression coefficients that are interpretationally equivalent to their corresponding structural coefficients in the SEM interaction model (corresponding t-values should have the same sign and roughly the same magnitude) (see FAQ D for more).

Regarding **interpretational considerations**, in Figure 1 for example, the association between Z and Y can be written as (1) $Y = dU + aZ + bUZ + error = dU + (a + bU)Z + error$, where d, a and b are unstandardized structural coefficients.[3] Thus, the (unstandardized or standardized) structural coefficient of Z's moderated association with Y is the coefficient $(a + bU)$.

Using path analysis (see Wright 1934), the (unstandardized or standardized) structural coefficient of X's association with Y via or mediated by Z is the product of the (unstandardized or standardized) structural coefficient on the X-Z path, c, with the (unstandardized or standardized) moderated structural coefficient on the Z-Y path, $(a + bU)$, which equals $c(a + bU)$ (= $ca + cbU$). Unfortunately, structural equation estimation packages will report only the unmoderated (indirect) X-Y association that has the coefficient ca. As a result, the coefficient of the (indirect) X-Y associations and their significances (e.g., X's indirect association with Y depending on the level of U in the coefficient $ca + cbU$) must be computed manually. (Click here for an EXCEL spreadsheet to expedite that process.)

REFERENCES

Johnson, D. J. and C. E. Rusbult (1989), "Resisting Temptation: Devaluation of Alternative Partners as a Means of Maintaining Commitment in Close Relationships," *J. of Personality and Social Psychology*, 57 (6), 967-980.

Ping, R. A. (1993), "The Effects of Satisfaction and Structural Constraints on Retailer Exiting, Voice, Loyalty, Opportunism, and Neglect," *J. of Retailing*, 69 (3), 320-352.

Ping, R. A. (1994), "Does Satisfaction Moderate the Association between Alternative Attractiveness and Exit Intention in a Marketing Channel?" *J. of the Academy of Marketing Science*, 22 (4), 364-371.

[3] In the second part of Equation 1 the partial derivative of Y with respect to Z is $(a + bU)$.

QUESTIONS of the MOMENT...

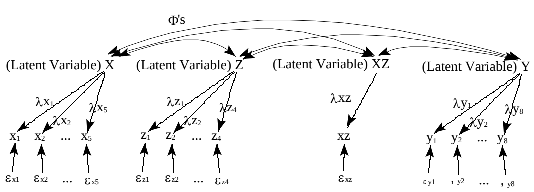
"Is there an example that shows all the steps in estimating a latent variable interaction/quadratic?"

(The **APA citation** for this paper is Ping, R.A. (2007). "Is there an example that shows all the steps in estimating a latent variable interaction/quadratic?" [on-line paper]. http://www.wright.edu/~robert.ping/yes.doc)

Yes, beginning with Chapter VIII of the monograph *Interactions and Quadratics in Survey Data:*... on this web site. The steps are few and simple; but they are tedious, and burdened with unfamiliar terminology and notation. In summary, after mastering the terminology, some of the survey data must be adjusted (mean centered) to facilitate interaction/quadratic estimation, the interaction/quadratic indicator must be created in the data set, and the interaction/quadratic must be specified in the model.

Beginning with the terminology, in structural equation modeling (SEM) an unobserved (Latent) variable (LV) (e.g., capital X in Figure 1--diagrams are used extensively in SEM) is assumed to be "connected" to (the "cause" of) one or more observed variables or "indicators" (items in a scale or measure) (e.g., small x's) by paths from (the correlations between) the LV to the indicator. The actual correlation or "path coefficient" on this path between the LV and an indicator is called a "loading" (the loading of the indicator on its LV) and it is represented by the Greek letter lambda (λ).

Figure 1--A Measurement Model



Every indicator (e.g., x_1) is assumed to be measured with error. This measurement error variable is denoted by a Greek epsilon (ε) (and its variance is denoted by $\theta\varepsilon$, and termed “theta”) and the correlation or "path coefficient" on the path between an observed indicator and its measurement error is assumed to be 1.

Because LV's, indicators, and measurement errors are variables they each have variance. The variance of an LV is represented by Phi (Φ_{LV}), the variance of an indicator has no special name or representation, and the variance of a measurement error is represented by $\theta\varepsilon$ (and termed “theta,” or theta epsilon). LV's are assumed to be correlated, this correlation is diagrammed with a two-headed arrow, and this correlation is also represented by Phi (e.g., $\Phi_{X,Z}$) (the particular meaning of a "Φ" is usually clear in context). Measurement errors are assumed to be uncorrelated, and the correlations (direct paths) among the indicators are not shown (they are accounted for by the paths to and between LV's).

Regression terminology is not used in SEM, and regression equations are infrequently written out in SEM. When they are, the familiar regression symbols are changed. For example, the regression equation

$$Y = b_0 + b_1X + b_2Z + b_3XZ + e \qquad (1$$

is written

$$Y = \beta_1X + \beta_2Z + \beta_3XZ + \xi \qquad (2$$

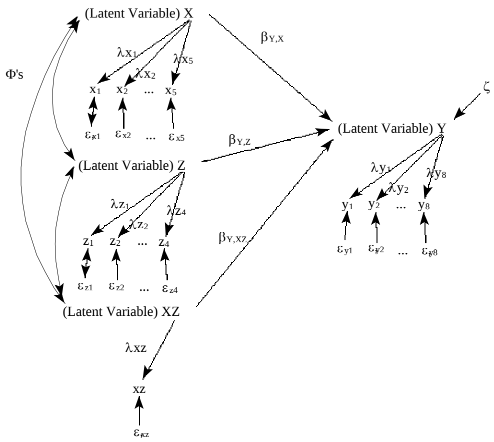
or

$$Y = \gamma_1X + \gamma_2Z + \gamma_3XZ + \xi$$

(b_0 is assumed to be zero), and is called a structural equation. The regression coefficients (b's in Equation 1) are called structural coefficients in SEM, and the regression forecast or estimation error (e) is called a structural disturbance (ζ). The dependent variable Y is called an *endo*genous variable and the independent variables X, Z and XZ are called *exo*genous variables.

Instead of being written out, regression relationships are usually diagrammed in SEM, using paths among the LV's, and regression relationships among the LV's are represented by single-headed arrows sometimes called structural paths (see Figure 2). The diagram of regression relationships is called a structural model (a diagram with no regression relationships, just correlations among the LV's, is called a measurement model).

Figure 2--A Structural Model



Note the arrow from XZ to Y in Figure 2. SEM has no ability to model an arrow from X, for example, pointing to, for example, an arrow from Z to Y (as sometimes done in texts). Stated differently, an interaction is "just another LV."

Because SEM is related to factor analysis, everything (λ's, θε's, etc.) is called a "parameter" and everything has its own matrix. For example, LV variances and the correlations among them belong to their PHI matrix, the structural disturbances (ζ 's) and the correlations among them belong to their PSI matrix, and the structural coefficients belong to their beta or gamma matrix.

The terms "estimation" and "specification" are common in SEM. "Estimation" means producing estimates of the model parameters--in regression the b's in Equation 1 are "estimated" by SPSS, SAS, etc. "Specification" means “to show.” For example, “specification” can mean diagramming the relationships in the model (e.g., Figures 1 and 2 are specifications). The relationships among X, Z and Y are specified by Equation 1. They are also "specified" in their SPSS, SAS, etc. regression code. The relationships between x_1, x_2, ... , x_5 and X are specified by the arrows from X to x_1, x_2, ... , x_5 in Figure 1.

In a measurement model all the LV's are assumed to be correlated. In a structural model, however, only exogenous (independent) LV's are assumed to be correlated. Dependent (endogenous) LV's are assumed not to be correlated (the correlations between the exogenous and endogenous variables are accounted for by structural paths). If there are multiple endogenous variables, the structural disturbances are usually assumed to be correlated.

To estimate the simple regression model specified as

$$Y = \beta_1 X + \beta_2 Z + \beta_3 XZ + \xi \qquad (1$$

where Y has indicators y_1, y_2, ... , y_8, x has indicators x_1, x_2, ... , x_5, z has indicators z_1, z_2, ... , z_4 (XZ and its indicator will be discussed shortly), the indicators (x_1, x_2, etc., z_1, z_2, etc.) of the constituents of XZ, X and Z, must be replaced by "mean centered" indicators (also called centered or zero-centered indicators). This is accomplished by subtracting the mean of x_1 from x_1, subtracting he mean of x_2 from x_2, etc., and the mean of z_1 from z_1, the mean of z_2 from z_2, etc. (see "Questions of the Moment," "How should PRELIS..." on this web site for more on mean centering). The resulting centered x's (xc_1, xc_2, ... , xc_5) and centered z's (zc_1, zc_2, ... , zc_4) are used in place of x_1, x_2, ... , x_7 and z_1, z_2, ... , z_6 from this point on.

Then, the single averaged indicator of XZ, $zx = [(xc_1+xc_2+xc_3+xc_4+xc_5)/5]*[(zc_1+zc_2+zc_3+zc_4)/4]$ is created (see "Questions of the Moment," "How should PRELIS..." on this web site for details)

Skipping over the XZ reliability and validity steps, which are covered elsewhere on this web site (e.g., FAQ D, beginning at "In Summary"), and assuming X, Z and Y are internally consistent (strongly unidimensional--a single construct measurement model with just X, for example, fits the data, ditto Z, and ditto Y) a "full" measurement model with X, Z and Y (but not XZ) (MM1) is run to estimate the parameters that are used to specify the loading and measurement error variance of xz (Barbara Byrne has written several books with measurement model examples using LISREL 8, EQS and AMOS). (Exhibit A shows the LISREL 8 commands for the Figure 1 measurement model—in SEM software with simplified commands such as LISREL's SIMPLIS, or a graphical interface such as AMOS, "drawing," or coding the paths, in Figure 1 specifies its measurement model). An EXCEL spreadsheet is provided on this web site to combine this full measurement model's parameter estimates into the loading (λxz—"lambda") and measurement error variance ($\theta\varepsilon_{xz}$—"theta") of XZ.

Exhibit A--LISREL 8 Figure 1 Measurement Model Commands

```
Example Measurement Model
DA NI=18 NO=200
LA
 x1 x2 x3 x4 x5
 z1 z2 z3 z4
 xz
```

```
 y1 y2 y3 y4 y5 y6 y7 y8
RA FI=raw.dat
SE
 x1 x2 x3 x4 x5
 z1 z2 z3 z4
!xz
 y1 y2 y3 y4 y5 y6 y7 y8
/
MO NY=17 ne=3 ly=fu,fi te=di,fr be=fu,fi ps=sy,fr
LE
 X
 Z
!XZ
 Y
pa ly
 *
 1 0 0 !x1
 0 0 0 !x2
 1 0 0 !x3
 1 0 0 !x4
 1 0 0 !x5
 0 0 0 !z1
 0 1 0 !z2
 0 1 0 !z3
 0 1 0 !z4
 0 0 1 !y1
 0 0 1 !y2
 0 0 1 !y3
 0 0 0 !y4
 0 0 1 !y5
 0 0 1 !y6
 0 0 1 !y7
 0 0 1 !y8
ma ly
 *
 1 0 0 !x1
 1 0 0 !x2
 1 0 0 !x3
 1 0 0 !x4
 1 0 0 !x5
 0 1 0 !z1
 0 1 0 !z2
 0 1 0 !z3
 0 1 0 !z4
 0 0 1 !y1
 0 0 1 !y2
 0 0 1 !y3
 0 0 1 !y4
 0 0 1 !y5
 0 0 1 !y6
 0 0 1 !y7
 0 0 1 !y8
pa te
 *
 17*1
ma te
 *
```

```
 17*.2
pa ps
 *
 1      !X
 1 1    !Z
 1 1 1 !Y
!X Z Y
ma ps
 *
 1      !X
 1 1    !Z
 1 1 1 !Y
!X Z Y
OU all nd=5 it=300 ad=300
```

Then, I specify another "full" measurement model containing the interaction/quadratic XZ and all the other latent variables (MM2--see Figure 1). This is done by adding the XZ to MM1, and using its fixed EXCEL spreadsheet loading and fixed EXCEL spreadsheet measurement error variance. The LISREL 8 commands for MM2 is shown in Exhibit B.

Exhibit B--LISREL 8 XZ Measurement Model Commands

```
Example Measurement Model
DA NI=18 NO=200
LA
 x1 x2 x3 x4 x5
 z1 z2 z3 z4
 xz
 y1 y2 y3 y4 y5 y6 y7 y8
RA FI=raw.dat
SE
 x1 x2 x3 x4 x5
 z1 z2 z3 z4
 xz
 y1 y2 y3 y4 y5 y6 y7 y8
/
MO NY=18 ne=3 ly=fu,fi te=di,fr be=fu,fi ps=sy,fr
LE
 X
 Z
 XZ
 Y
pa ly
 *
 1 0 0 0  !x1
 0 0 0 0  !x2
 1 0 0 0  !x3
 1 0 0 0  !x4
 1 0 0 0  !x5
 0 0 0 0  !z1
 0 1 0 0  !z2
 0 1 0 0  !z3
 0 1 0 0  !z4
 0 0 0 0  !xz
 0 0 0 1  !y1
```

```
 0 0 0 1  !y2
 0 0 0 1  !y3
 0 0 0 0  !y4
 0 0 0 1  !y5
 0 0 0 1  !y6
 0 0 0 1  !y7
 0 0 0 1  !y8
ma ly
 *
 1 0 0 0  !x1
 1 0 0 0  !x2
 1 0 0 0  !x3
 1 0 0 0  !x4
 1 0 0 0  !x5
 0 1 0 0  !z1
 0 1 0 0  !z2
 0 1 0 0  !z3
 0 1 0 0  !z4
!The xz loading was calculated using the
!EXCEL spreadsheet on this web site
 0 0 0.80640 0  !xz
 0 0 0 1  !y1
 0 0 0 1  !y2
 0 0 0 1  !y3
 0 0 0 1  !y4
 0 0 0 1  !y5
 0 0 0 1  !y6
 0 0 0 1  !y7
 0 0 0 1  !y8
pa te
 *
 9*1 0 8*1
ma te
 *
!The xz measurement error variance was calculated using the
!EXCEL spreadsheet on this web site
 9*.2 0.03903 8*.2
pa ps
 *
 1        !X
 1 1      !Z
 1 1 1    !XZ
 1 1 1  1 !Y
!X Z XZ Y
ma ps
 *
 1       !X
 1 1     !Z
!The xz variance is a starting value that LISREL cannot estimate
!and it was calculated using the EXCEL spreadsheet on this web site
 1 1 0.57040 !XZ
 1 1 1  1 !Y
!X Z XZ Y
OU all nd=5 it=300 ad=300
```

In Exhibit B, note that raw data was used as input and a covariance matrix was analyzed. Also note that the variance of the interaction XZ was free (the starting value shown is optional in small models, but frequently essential in larger models). The interaction XZ was also allowed to correlate with X and Z by freeing the correlational (PHI) paths between them. If there had been other exogenous variables, XZ would be allowed to correlate with them as well.

If MM1 fits the data, this measurement model also should fit the data, and the parameter estimates for X and Z (loadings, variances, and measurement error variances) should be the same as they were in MM1, in at least the first two decimal places. If not, use these measurement model values in the EXCEL spreadsheet to recompute the xz loading and measurement error variance.

Next, the structural model containing the interaction/quadratic XZ and all the other latent variables is estimated (see Figure 2). This is done in by altering MM2 by fixing the correlations between Y and the LV's X, Z and XZ in MM2 to zero, and by specifying the beta paths between Y and the LV's X, Z and XZ. The LISREL 8 commands for the structural model is shown in Exhibit C. (Again, in SEM software with simplified commands such as LISREL's SIMPLIS, or a graphical interface such as AMOS, "drawing," or coding the paths, in Figure 2 specifies its structural model).

Exhibit C--LISREL 8 Figure 2 Structural Model Commands

```
Example Structural Model
DA NI=18 NO=200
LA
 x1 x2 x3 x4 x5
 z1 z2 z3 z4
 xz
 y1 y2 y3 y4 y5 y6 y7 y8
RA FI=raw.dat
SE
 x1 x2 x3 x4 x5
 z1 z2 z3 z4
 xz
 y1 y2 y3 y4 y5 y6 y7 y8
/
MO NY=18 ne=3 ly=fu,fi te=di,fr be=fu,fi ps=sy,fr
LE
 X
 Z
 XZ
 Y
pa ly
 *
 1 0 0 0  !x1
 0 0 0 0  !x2
 1 0 0 0  !x3
 1 0 0 0  !x4
 1 0 0 0  !x5
 0 0 0 0  !z1
 0 1 0 0  !z2
```

```
 0 1 0 0  !z3
 0 1 0 0  !z4
 0 0 0 0  !xz
 0 0 0 1  !y1
 0 0 0 1  !y2
 0 0 0 1  !y3
 0 0 0 0  !y4
 0 0 0 1  !y5
 0 0 0 1  !y6
 0 0 0 1  !y7
 0 0 0 1  !y8
ma ly
 *
 1 0 0 0  !x1
 1 0 0 0  !x2
 1 0 0 0  !x3
 1 0 0 0  !x4
 1 0 0 0  !x5
 0 1 0 0  !z1
 0 1 0 0  !z2
 0 1 0 0  !z3
 0 1 0 0  !z4
!The xz loading was calculated using the
!EXCEL spreadsheet on this web site
 0 0 0.80640 0  !xz
 0 0 0 1  !y1
 0 0 0 1  !y2
 0 0 0 1  !y3
 0 0 0 1  !y4
 0 0 0 1  !y5
 0 0 0 1  !y6
 0 0 0 1  !y7
 0 0 0 1  !y8
pa te
 *
 9*1 0 8*1
ma te
 *
!The xz measurement error variance was calculated using the
!EXCEL spreadsheet on this web site
 9*.2 0.03903 8*.2
pa be
 *
 0 0 0  0 !X
 0 0 0  0 !Z
 0 0 0  0 !XZ
 1 1 1  0 !Y
!X Z XZ Y
pa ps
 *
 1       !X
 1 1     !Z
!The xz variance is a starting value that LISREL cannot estimate
!and it was calculated using the EXCEL spreadsheet on this web site
 1 1 0.57040 !XZ
 0 0 0  1 !Y
!X Z XZ Y
```

```
ma ps
 *
 1       !X
 1 1     !Z
 1 1 1   !XZ
 0 0 0  1 !Y
!X Z XZ Y
OU all nd=5 it=300 ad=300
```

Now, go on to FAQ D, skip down to "In Summary," and verify that the structural model results are "trustworthy" using the remarks beginning with "Once the model is estimated, check..." on p. 10.

QUESTIONS of the MOMENT...

"Why are reviewers complaining about my use of standardized loadings?"

(The **APA citation** for this paper is Ping, R.A. (2013). "Why are reviewers complaining about my use of standardized loadings?" [on-line paper]. http://www.wright.edu/~robert.ping/stdLoad.doc)

Jöreskog (1996, "LISREL 8 … Reference Guide, p.35) warned that standard errors (in LISREL), among other statistics, may be incorrect when correlations are analyzed (without standard deviations, etc.) in structural equation models. This presents a problem in theory testing—an incorrect (biased) standard error for a structural coefficient means that its t-value is incorrect, and any interpretation of the observed structural coefficient's significance or nonsignificance versus its hypothesis may be risky.

While I have yet to find equivalent warnings about correlations and standard errors in documentation for EQS or AMOS, other authors have warned against analyzing correlations (see the citations in Bentler 2006, "EQS 6 … Program Manual," p. 11). As a result, it may be prudent to avoid analyzing correlations in theory tests involving structural equation analysis.

However, it is easy to show using real-world data that covariances and "standardized loadings" (latent variable (LV) loadings specified as all free—so the resulting LV has a variance of 1) may produce incorrect t-values for parameter estimates, including structural coefficients. Specifically, the t-values of the resulting structural coefficients (which are now standardized estimates) may be different from those produced by the preferred "unstandardized loadings" LV specification, where one loading of each LV is fixed at 1, and each LV's estimated (error-disattenuated) variance is different from 1 (e.g., Jöreskog 1996).

Thus, it also may be prudent to avoid using standardized loadings in theoretical model tests involving structural equation analysis. (If standardized coefficient estimates are required, standardized and unstandardized estimates could be requested, and standardized values could be reported with unstandardized t-values.)

Parenthetically, one procedure for specifying unstandardized LV loadings is to specify each LV with its first indicator fixed at 1. To simplify any subsequent interpretation of loadings, I then respecify each LV by fixing the largest loading of each LV to 1, and freeing each LV's first indicator if it is not the largest (to avoid having two indicators fixed at 1), and reestimate the model.

QUESTIONS of the MOMENT...

"How does one estimate categorical variables in theoretical model tests using structural equation analysis?"

(The **APA citation** for this paper is Ping, R.A. (2010). "How does one estimate categorical variables in theoretical model tests using structural equation analysis?" [on-line paper]. http://www.wright.edu/~robert.ping/categorical3.doc)

(An earlier version of this paper, Ping, R.A. (2008). "How does one estimate categorical variables in theoretical model tests using structural equation analysis?" [on-line paper]. http://www.wright.edu/~robert.ping/categorical1.doc, is available here.)

In structural equation analysis software (e.g., LISREL, EQS, Amos, etc.), the term "categorical variable" usually means an ordinal variable (e.g., an attitude measured by Likert scales), rather than a nominal or "truly categorical" variable (e.g., Marital Status, with categories such as Single, Married, Divorced, etc.), and there is no provision for truly categorical variables.

In regression, a (truly) categorical variable is estimated using "dummy" variables. (For example, while the cases in Marital Status, for example, might have the values 1 for Single, 2 for Divorced, 3 for Married, etc., a new variable, Dummy_Single, is created, with cases that have the value 1 if Marital Status = Single, and 0 otherwise. Dummy_Married is similar with cases equal to 1 or 0, etc.) The same approach might be used in structural equation modeling (SEM).

However, nominal variables present difficulties in SEM that are not encountered in dummy variable regression. For example, dummy variables violate several important assumptions in SEM: dummy variables are not continuous, and they do not have normal distributions. While ordinal variables from (multi-point) rating scales (e.g., Likert scales), that are not continuous and not normally distributed, are routinely analyzed in theoretical model (hypothesis) tests using SEM, dummy variables with only two values are very non-normal, and the covariances that are customarily analyzed in SEM, are formally incorrect. Point-biserial, tetrachoric, polychoric, etc. correlations are more appropriate.

In addition, Maximum Likelihood (ML) structural coefficient estimates are preferred in theoretical model tests, and this estimator also assumes multivariate normality. While ML structural coefficients are believed to be robust to departures from normality (see the citations in Chapter VI. RECENT APPROACHES TO ESTIMATING INTERACTIONS AND QUADRATICS), it is believed that standard errors are not robust to departures from normality (Bollen 1995). (There is some evidence to the contrary (e.g., Ping 1995, 1996), and EQS, for example, provides Maximum Likelihood Robust estimates of structural coefficients and standard errors that appear to be robust to departures from normality (see Cho, Bentler and Satorra 1991).)

In "interesting" models (ones with more than a few latent variables) there could be several (truly) categorical variables, each with several categories. Because most SEM software requires at least 1 case per estimated parameter (10 are preferred) (some authors prefer a stricter criterion from regression: multiple cases per covariance matrix entry), the number of dummy variables can empirically overwhelm the model unless the sample is large and the number of dummy variables is comparatively small. For example, in a model that adds 2 categorical variables each with 2 categories to 2 latent variables, 11 additional parameters are required for the additional correlations and/or structural coefficients due to the dummy variables--the loadings and measurement error variances of dummies are assumed to be 1 and 0 respectively. This would require up to 110 addition cases to safely estimate these parameters (5 times more additional cases would be required for the stricter regression criterion involving the asymptotic "correctness" of the covariance matrix).

Thus, in addition to samples larger than the customary number of cases used in survey data model tests (e.g., 200-300 cases), "managing" the total number of categories is required (more on this later).

With dummy variables, the term "estimates" (e.g., of association and significance) truly may apply. While there is no hard and fast rule, significance thresholds in theoretical model testing with dummy variables and SEM probably should be conservative (above $|t| > 2$).

The results of a categorical model estimation is typically a subset of significant dummy variables (e.g., Single and Divorced). However, an estimate of the significance of the categorical variable from which the dummy variables were formed (e.g., Marital Status) of is not available in the SEM output. Thus, the aggregate effect (e.g., the overall significance) of the dummy variables comprising Marital Status, for example, should be determined to gage hypothesis disconfirmation (an EXCEL spreadsheet to expedite this task is available below).

Finally, estimating all the dummy variables jointly does not work in popular SEM software such as LISREL, EQS, AMOS, etc., and a SEM "workaround" is taking longer than anticipated.

However, there is a “mixed SEM” estimation procedure for latent variables (LV’s) and truly categorical variables that could be used until an “all-SEM” approach is found. It produces “proper” error-dissatenuated structural coefficients just like (all) SEM does. They are Least Squares estimates rather than Maximum Likelihood estimates, however, but the approach might be preferable to omitting an important categorical variable(s), or analyzing subsets. (For more on the alternatives, see below at the “**” in the left margin.)

Using this mixed SEM approach, the steps for estimating the (total) effect of Marital Status, for example, on Y (i.e., “H1: Marital Status affects/changes/etc. Y”) in a survey-data model[1] would be to create a survey with an exhaustive list of categories for Marital Status (the number of categories might be reduced later). A large number of cases then should be obtained if possible. When the responses are available, the exhaustive list of categories for Marital Status, for example, should be reduced, if possible (i.e., categories with few cases should be dropped or combined with other categories).

Next, dummy variables for Marital Status, for example, should be created with a dummy variable for each category in Marital Status such as Single, Married, Divorced, etc. Specifically, a dummy variable such as Dummy_Single, should be created, with cases that have the value 1 if Marital Status = Single, and 0 otherwise. Dummy_Married would be similar with cases equal to 1 if Marital Status = Married, and 0 otherwise.

[1] It turns out that the suggested approach also might be appropriate for an experiment, but that is another story.

Dummy_Divorced, etc. also would be similar. There should be 1 dummy variable for each category in the (truly) categorical variable Marital Status, and the sum of the cases that are coded 1 in each categorical variable should equal the number of cases. (For example, the sum of the cases that are coded 1 across the dummy variables for Marital Status should equal the number of cases, the sum of the cases that are coded 1 across the dummy variables for any next categorical variable should equal the number of cases, etc.) In total, there would be 1 different dummy variable for each category in each of the (truly categorical) variables.

Then, a least squares regression version of the structural equation containing (all) the dummy variables should be estimated to roughly gage the strength of any ordinal effects. For example, in

1) $Y = b_1X + b_2Z + b_3\text{Dummy\_Single} + b_4\text{Dummy\_Married} + b_5\text{Dummy\_Divorced} + b_6\text{Dummy\_Separated} + b_7\text{Dummy\_Widowed}$,

the latent variables Y, X and Z would be "specified" using summed, preferably averaged, indicators, and the regression should use the "no origin" option (i.e., regression through the origin). If the regression coefficient for each of the dummy variables is non significant, it is unlikely that Marital Status is significant.

However, assuming at least 1 dummy variable was significant, the reliability, validity and internal consistency of the LV's in the hypothesized model should be gaged. (The dummy variables will have to be assumed to be reliable and valid, and they are trivially internally consistent). In particular, the single-construct measurement model (MM) for each LV should fit the data. (This step is required later for unbiased estimation.)

Next, a full MM that omits the dummy variables should be estimated to gage external consistency. Assuming this "no dummies" MM fits the data, full measurement models that omit the dummy variables one at a time should be estimated to further gage external consistency. For example, Dummy_Single should be omitted from the (full) MM for Equation 1. Then, a second full MM containing Dummy_Single but omitting Dummy_Married should be estimated. This should be repeated with each dummy variable. (Experience suggests that in real-world data the parameter estimates in these MM will vary trivially, which suggests that the dummy variables do not materially effect external consistency.)

Then, assuming the LV's are reliable, valid and consistent, the LV's should be averaged and their error-attenuated covariance matrix (CM) should be obtained using SPSS, SAS, etc. Next, this matrix is adjusted for measurement errors using a procedure suggested by Ping (1996b) (and the "Latent Variable Regression" EXCEL spreadsheet that is available on this web site). For consistent LV's, the resulting Error-Adjusted (Err-Adj) CM then is used to estimate Equation 1 without omitting dummy variables. Specifically, the error-attenuated/error unadjusted (err-unadj) CM for all the variables in Equation 1 is adjusted for measurement error using the measurement model loadings and measurement error variances from the "no dummies" MM for Equation 1. The resulting Err-Adj CM then is used as input to least squares regression. This procedure was judged to be unbiased and consistent in the Ping (1996b) article, and while it is not as elegant as SEM, it does produce "proper" unbiased and consistent structural coefficients in a model containing LV's and (truly) categorical variables just like SEM should (but so far doesn't).

Specifically, the parameter estimates from the "no dummies" MM are input to the "Latent Variable Regression" EXCEL spreadsheet that produces the Err-Adj CM matrix using calculations such as

$$\mathrm{Var}(\xi_X) = (\mathrm{Var}(X) - \theta_X)/\Lambda_X^2$$

and

$$\mathrm{Cov}(\xi_X,\xi_Z) = \mathrm{Cov}(X,Z)/\Lambda_X\Lambda_Z ,$$

where $\mathrm{Var}(\xi_X)$ is the desired error-adjusted variance of X (that is input to regression), Var(X) is the error *attenuated* variance of X (from SAS, SPSS, etc.), $\Lambda_X = \mathrm{avg}(\lambda_{X1} + \lambda_{X2} + ... + \lambda_{Xn})$, avg = average, and $\mathrm{avg}(\theta_X = \mathrm{Var}(\varepsilon_{X1}) + \mathrm{Var}(\varepsilon_{X2}) + ... + \mathrm{Var}(\varepsilon_{Xn}))$, ($\lambda$'s and ε_X's are the measurement model loadings and measurement error variances from the "no dummies" MM--1 and 0 respectively for the dummy variables--and n = the number of indicators of the latent variable X), $\mathrm{Cov}(\xi_X,\xi_Z)$ is the desired error-adjusted covariance of X and Z, and Cov(X,Z) is the error *attenuated* covariance of X and Z from SPSS, SAS, etc.[2]

The resulting Err-Adj CM (on the EXCEL spreadsheet) is then input to regression, with the "regression-through-the-origin" option (the no-origin option) (see the spreadsheet for details).

Because the coefficient standard errors (SE's) (i.e., the SE's of b_1, b_2, ..., and b_7 in Equation 1) produced by the Err-Adj CM are incorrect (they assume variables that are measured without error--e.g., Warren, White and Fuller 1974; see Myers 1986 for additional citations), they must also be corrected for measurement error. A common correction is to adjust the SE from regression using the err-unadj CM by changes in the standard error, RMSE (= $[\Sigma[y_i - \mathit{y}_i]^2]^2$, where y_i and y_i. are observed and estimated y's respectively) from using the Err-Adj CM (see Hanushek and Jackson 1977). Thus the correct SE's for the Err-Adj CM structural coefficients would involve the SE from regression using the err-unadj CM, and a ratio of the standard error from err-unadj CM regression and the standard error from Err-Adj CM regression, or

$$SE_A = SE_U{*}RMSE_U/RMSE_A ,$$

where SE_A is the Err-Adj CM regression standard error, SE_U is the SE produced by err-unadj CM regression, $RMSE_U$ is the standard error produced by err-unadj CM regression, and $RMSE_A$ is the standard error produced by err-unadj CM regression.[3]

Then, the structural coefficients of the dummies for each of the categorical variables could be aggregated to adequately test any hypotheses (e.g., "H1: Marital Status affects/changes/etc. Y") (click here for an EXCEL spread sheet to expedite this process). Specifically, if there are multiple categorical variables, the effects of each dummy variable would be aggregated (e.g., the effects of Marital Status, for example, would be aggregated, then all of the effects of categorical variable 2 would be aggregated ignoring the dummy variable effects of Marital Status, etc.).

DISCUSSION

---

[2] These equations make the classical factor analysis assumptions that the measurement errors are independent of each other, and the x_i's are independent of the measurement errors. The indicators for X and Z must be consistent in the Anderson and Gerbing (1988) sense.

[3] The correction was judged to be unbiased and consistent in Ping (2001).

* Several comments may be of interest. There is a categorical variable that could be estimated directly in SEM: a single dichotomous variable (e.g., gender). In this case the model could be estimated using SEM in the usual way (i.e., ignoring the suggested procedure) with the categorical variable specified using a loading of 1 and a measurement error variance of zero. For emphasis ML estimates, ML Robust estimates if possible, should be obtained. In addition because the difficulties introduced by a dichotomous, non-normal variable, the significance threshold for the dichotomous variable probably should be conservative (e.g., t-values probably should be at least 10% higher than usual).

** Because the above procedure is a departure from the usual LV model estimation, it may be useful to discuss the justification for the departure in detail. First, in theoretical model testing, theory is always more important than methodology. Stated differently, theory ought not be altered to suit methodology. Thus, plausible categorical variables should not be ignored simply because they cannot be tested with LISREL, AMOS, etc. Method is important, however—it must adequately test the proposed theory, and thus the proposed methodological approach should be shown to be adequate.

The options besides the suggested procedure are to analyze subsets of data for each dummy variable,[4] or to omit the categorical variable(s). Not only does omitting a plausible categorical variable to enable analysis with LISREl, AMOS, etc. in the "usual" manner force theory to accommodate method, omitting a focal categorical can reduce the "interestingness" of the model. And, omitting a plausible antecedent of Y in for example in Equation 1, courts biased model estimation results because of the "missing variable problem" (see James 1980).[5] [6]

Analyzing subsets of data for each dummy variable, however, is almost always an inadequate test. First, sequentially testing the dummy variables invites one or more spurious significances—significances that occur by chance because so many tests are performed.

In addition, splitting the data into subsets reduces statistical power, increasing the likelihood of falsely disconfirming one or more dummy variables, and thus the hypothesized categorical variable. Further, analyzing subsets provides no means to aggregate the dummy variable estimates, and thus it cannot adequately test a categorical hypothesis such as "Marital Status affects Y."

---

[4] For example, the data set would be split into a subset of respondents who were single, and another subset of those who were not. Then, the model would be estimated using these subsets, and structural coefficients would be compared between the subsets for significant differences. This would be repeated for respondents who are married, etc.

[5] Omitting an important antecedent, in this case a categorical variable, that may be correlated with other model antecedents creates the "missing variables problem." This can bias structural coefficients because model antecedents are now correlated with the structural error term(s), a violation of assumptions (structural errors contain the variance of omitted variables).

[6] A reviewer also might question the "importance" of the categorical variable(s) with logic such as: "Because it usually does not explain much variance in Y, any 'missing (categorical) variable problem' is likely comparatively small, so why not just omit the categorical variable and use SEM?" This logic misses the point of theory testing: What are the theoretically justified antecedents of Y, no matter how "important" they are? Besides, the categorical variable could be more "important" in the next study.

However, while disqualifying all the alternatives may make the proposed procedure attractive, it does not make it adequate. A formal investigation of any bias and inefficiency (e.g., using artificial data sets) would suggest adequacy. Absent that, comparisons with comparatively “trustworthy” estimates should provide at least hints of adequacy. Thus, to suggest the structural coefficient estimates for the LV’s are adequate, the proposed procedure results could be compared to several SEM models with 1 dummy missing. These model should be ML and LS, and the results should be argued to be “interpretationally equivalent.”[7] Because the dummy results will not be equivalent, the step c) (err-unadj CM) regression results should be compared to Err-Adj CM and argued to be “interpretationally equivalent.”

In addition to expecting SEM, reviewers expect ML. However, comparing “1 dummy missing” estimates using (ML) SEM and (LS) Err-Adj CM should hint that the (LS) Err-Adj CM estimates for all the dummies might be interpretationally equivalent to SEM estimates were they available.[8]

Parenthetically, it may not be necessary to burden a first draft of a mixed categorical model with these arguments—reviewers may already be aware of the estimation difficulties. I would simply mention that SEM doesn’t work for the usual approach for categoricals (dummy variables), and the procedure used in the paper was judged to have fewer drawbacks than the alternatives: omitting the categorical variable, or estimating multiple subsets of cases.

Because the above procedure has not been formally investigated for any bias or inefficiency, the estimation results of the aggregated effects of each categorical variable probably should be interpreted using a stricter criterion for significance (e.g., t-values probably should be at least 10% higher than usual). (Err-Adj CM regression is known to be unbiased and efficient.)

"Managing" the number of categories is typically required in order to minimize the reduction of the asymptotic correctness of the Err-Adj CM caused by the addition of the dummy variables to the model. Specifically, categories with a few cases should be dropped or combined with other categories if possible. For example, if there were few Divorced, Separated and Widowed respondents, those dummies could be replaced by Dummy_Not_Married (i.e., Dummy_Not_Married = 1 if the respondent was Divorced, Separated or Widowed, zero otherwise). The drawback to this combining categories is that the Divorced, Separated and Widowed categories probably should be combined in subsequent studies.

Similarly, categories might be combined or ignored to suit an hypothesis. For example, if the study were interested only in married respondents, the Divorced, Separated and Widowed dummies could be replaced by Dummy_Not_Married (i.e.,

---

[7] Interpretationally equivalent estimates have the same algebraic sign, and are both are either significant or both are not significant.

[8] This appears to beg the question, why insist on ML? ML estimates are preferred in survey-data theory tests because they are more likely to be observed in future studies.

Dummy_Not_Married = 1 if the respondent was Divorced, Separated or Widowed, zero otherwise). Afterward, Dummy_Married could be estimated directly using SEM (see "*" in the left margin above).

Aggregation is recommended even if it does not seem to be required (e.g., the study may be interest only in single respondents). In fact, it functions as an overall "F-like' test of the dummies. Specifically, if the aggregation is not significant, any significant dummies probably should be ignored.

For emphasis, nearly every SEM study in the Social Sciences has categorical variables in the "Demographics" section of the study questionnaire. Anecdotally, applied researchers routinely analyze this data "post-hoc" (after the study is completed) for "finer grained" views of study results by gender, title, marital status, VALS psychographic category, etc. The suggested procedure above would enable such analyses in theory tests using SEM. Such post-hoc "probing" is within the logic of science as long as the results are clearly presented as having been found after the study was completed (and thus provisional, and in need of disconfirmation in a future study). For example, Marital Status was actually found to be a predictor of exiting propensity in a reanalysis of a study's data. After an argument supporting a Marital Status hypothesis was created, an (as yet unpublished) new study was conducted to disconfirm this. and several other, new hypotheses. Stated differently, the demographic analysis triggered a new line of thought that resulted in new theory (several new hypotheses).

Interested readers are encouraged to read the papers, "But what about Categorical (Nominal) Variables in Latent Variable Models?" "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients" and "A Suggested Standard Error for Interaction Coefficients in Latent Variable Regression" on this web site for more details about estimating categoricals in SEM, and Err-Adj CM regression and its adjusted standard error respectively.

The additional steps for estimating categoricals with several endogenous variables are available by e-mail.

SUMMARY

In summary, the steps for estimating an hypothesized effect of Marital Status, for example, on Y (i.e., "H1: Marital Status affects/changes/etc. Y") in a survey-data model would be to:

a) create a survey in the usual way. However, a large number of cases should be obtained, if possible, to permit the addition of the dummy variables. When the responses are available, the number of categories should be reduced, if possible to improve the asymptotic correctness of the model parameter estimates.

b) Create dummy variables with 1 dummy variable for each category in the (truly) categorical variable. (The sum of the cases that are coded 1 in each categorical variable should equal the number of cases.)

c) Average the LV's and estimate a least squares regression version of the structural equation containing (all) the dummy variables, using regression through the origin, to roughly gage the strength of any ordinal effects. (If no dummy variables are significant, it is unlikely that the hypothesized categorical variable is significant.)

d) If at least 1 dummy variable was significant, the reliability, validity and internal consistency of the LV's in the hypothesized model should be gaged as usual.

e) Estimate a full MM that omits the dummy variables should be estimated to gage the external consistency of the LV's. If this "no dummies" MM fits the data, full measurement models that omit the dummy variables one at a time should be estimated to further gage external consistency, and to determine if the dummy variables materially effect external consistency.

f) Average the LV's and their error-unadjusted covariance matrix (err-unadj CM) should be obtained using SPSS, SAS, etc.

g) Input the parameter estimates from the "no dummies" MM to the "Latent Variable Regression" EXCEL spreadsheet (on this web site) to produce an Error-Adjusted covariance matrix (Err-Adj CM).

h) Input this Err-Adj CM to regression, with the "regression-through-the-origin" option (the no-origin option).

i) Correct the coefficient standard errors (SE's) from the Err-Adj CM regression using the EXCEL spreadsheet.

j) Aggregate the structural coefficients of the dummies for each of the categorical variables to adequately test any hypotheses (e.g., "H1: Marital Status affects/changes/etc. Y").

k) Interpret dummy and aggregation results using a stricter criterion for significance (e.g., $|t| > 2.2$).

REFERENCES

Bollen, Kenneth A. (1995), "Structural Equation Models that are Nonlinear in Latent Variables: A Least Squares Estimator," *Sociological Methodology*, 25, 223-251.

Cho, C. P., P. M. Bentler, and A. Satorra (1991), Scaled Test Statistics and Robust Standard Errors for Non-Normal Data in Covariance Structure Analysis: A Monte Carlo Study," *British Journal of Mathematical and Statistical Psychology*, 44, 347-357.

Hanushek, Eric A. and John E. Jackson (1977), *Statistical Methods for Social Scientists*, New York: Academic Press.

James, Lawrence R. (1980), "The Unmeasured Variables Problem in Path Analysis," *Journal of Applied Psychology*, 65 (4), 415-421.

Myers, Raymond H. (1986), *Classical and Modern Regression with Applications*, Boston: Duxbury Press, 211.

Ping, R. A. (1995), "A Parsimonious Estimating Technique for Interaction and Quadratic Latent Variables," *The Journal of Marketing Research*, 32 (August), 336-347.

Ping, R. A. (1996a), "Latent Variable Interaction and Quadratic Effect Estimation: A Two-Step Technique Using Structural Equation Analysis," *Psychological Bulletin*, 119 (January), 166-175.

Ping, R. A. (1996b), "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients," *Multivariate Behavioral Research*, 31 (1), 95-120.

Warren, Richard D., Joan K. White and Wayne A. Fuller (1974), "An Errors-In-Variables Analysis of Managerial Role Performance," *Journal of the American Statistical Association*, 69 (328) December, 886-893.

QUESTIONS of the MOMENT...

"Why are reviewers complaining about the use of moderated multiple regression in my paper?"

(The **APA citation** for this paper is Ping, R.A. (2009). " Why are reviewers complaining about the use of moderated multiple regression in my paper?" [on-line paper]. http://www.wright.edu/~robert.ping/MR.doc)

Multiple regression, and moderated multiple regression, assumes each independent variable is measured without error (i.e., the observed score is exactly the true score). Unfortunately, it is well known that the extent and direction of all regression coefficient is biased by even a single variable that contains (known or unknown) measurement error (e.g., Aiken and West 1991, Bohrnsted and Carter 1971, Cohen and Cohen 1983, Kenny 1979).

Even though this assumption was well known, it was routinely ignored in theoretical model (hypothesis) testing until Jöreskog's proposal that, among other things, allowed modeling of measurement error (Jöreskog 1970, 1971) (i.e., structural equation analysis). As a result, reviewers may reject substantive papers that rely on regression because 1) its regression's assumption of variables with no measurement error is now believed to be violated even in demographic variables such as age and income (both are typically misreported by some respondent groups, and each is typically measured in "round numbers"). 2) reviewers are (re)aware of how regression estimates can be biased (i.e., untrustworthy) in theoretical model tests when one or more variable contains measurement error (unless they are uncorrelated with any of the other independent variables, which is unlikely in real-world data). And, 3) regression usually produces Least Squares estimates--Maximum Likelihood estimates are now preferred for theoretical model testing.

As a result, some reviewers now believe that regression is an insufficient test of a theoretical model if there is measurement error in even one model variable (i.e., all the resulting coefficients used to test the hypotheses are untrustworthy).

Many suggested procedures for moderated multiple regression (e.g., Barron and Kenny 1986) are now considered inappropriate for theory testing because for example, the analysis procedures (e.g., stepping variables in, etc.) also are insufficient tests of the hypotheses.

Alternatives to ordinary least squares regression that account for measurement error include Fuller (1991) and Ping (1996), but each has drawbacks. Fuller's proposals are inaccessible to many substantive researchers. Ping's proposal relies on measurement parameter estimates from structural equation analysis, and begs the question, why not just use structural equation analysis?

The "problems" with utilizing the now preferred structural equation analysis, appear to be several: it is not taught in all terminal degree programs. And, despite texts apparently aimed at "self teaching" it (e.g., Byrne 1990), and (powerful) graphical user interfaces now available in most structural equation analysis software packages, anecdotally, structural equation analysis still seems to be inaccessible to many substantive researchers when compared to regression. For untenured researchers who may be "on a clock," this can slow productivity. For others, this can require "finding" someone who does structural equation analysis, then "managing" their involvement in the resulting paper. Structural equation analysis also can appear to "take over" a theoretical piece, producing a perhaps unwelcome intrusion on its theoretical matters.

"Solutions" to the structural equation analysis "problems" all have drawbacks. First, if structural equation analysis is not required (e.g., for a dissertation), to conserve time don't use it. However, for the reasons stated above, this may be a temporary solution.

Next, consider allowing about a month to do three things: first, finding someone to help with learning structural equation analysis, then learning only enough structural equation analysis to "get by" reviewers. Then, consider quickly creating/revising a paper with a simple model (or a simple submodel of your current model) that uses (replaces regression with) structural equation analysis, and submitting it to a good conference. Rather than acceptance, the objective would be to learn structural equation analysis in a realistic setting. Any reviewer feedback would also suggest what/where more structural equation analysis work is needed.

Click here for more about structural equation analysis as "regression using factor scores instead of averaged items," and how to learn the basics in a reasonable amount of time.

References

Aiken, L. S. and S. G. West (1991), *Multiple Regression: Testing and Interpreting Interactions*, Newbury Park, CA: Sage.
Barron, R. M. and D. A. Kenny (1986), "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology*, 51(6): 1173-1182.
Bohrnstedt, G. W. and T. M. Carter (1971), "Robustness in regression analysis," in H.L. Costner (Ed.), *Sociological Methodology* (pp. 118-146), San Francisco: Jossey-Bass.
Byrne, B.M. (1990). *A Primer of LISREL: Basic Applications and Programming for Confirmatory Factor Analytic Models*. New York: Springer-Verlag Inc.
Cohen, Jacob and Patricia Cohen (1983), *Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences*, Hillsdale, NJ: Lawrence Erlbaum.
Fuller, Wayne A. (1991), "Regression Estimation in the Presence of Measurement Error, " in *Measurement Errors in Surveys*, B. P. Beimer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz and S. Sudman, eds., NY: Wiley.
Jöreskog, Karl G. (1970), "A General Method for Analysis of Covariance Structures," *Biometrika*, 57, 239-251.

________ (1971) "Simultaneous Factor Analysis in Several Populations," *Psychometrika*, 57, 409-426.
Kenny, David (1979), *Correlation and Causality*, New York: Wiley.
Ping, R.A. (1996c), "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients," *Multivariate Behavioral Research*, 31 (1), 95-120.

QUESTIONS of the MOMENT...

"How should PRELIS or similar "preprocessor" software be used with LISREL, EQS, AMOS, etc. to create interactions/quadratics?"

(The **APA citation** for this paper is Ping, R.A. (2007). "How should PRELIS or similar "preprocessor" software be used with LISREL, EQS, AMOS, etc. to create interactions/quadratics?" [on-line paper]. http://home.att.net/~rpingjr/PRELIS.doc)

For Interactions/Quadratics, I usually do not use PRELIS or the other excellent data entry or "preprocessor" software available with LISREL, EQS, AMOS, etc. Instead, I use EXCEL[1].

Nevertheless, for Interactions/Quadratics the procedures are the same for most of them, and similar to EXCEL, and I will illustrate zero- or mean-centering, creating Interaction/Quadratic indicator(s), etc. using the hopefully familiar Microsoft EXCEL. First, key each questionnaire (case) into its row with the response to each item in its respective column on the "spreadsheet." Most research houses also number each questionnaire and key that number as the row number in column 1 (so they can match any questionable data with its questionnaire later).

Most research houses also verify each questionnaire's data entry by typing it a second time, and then either comparing the two rows of entries, or summing each row and comparing these two sums (don't forget to delete one row). This is a bother because it doubles the data entry time, but it is sometimes a comfort to know that any problems later are not due to mis-keyed data.

An optional step is to cluster the cases using Ward's method, with squared Euclidean distance, into 3 groups. The questionnaires should cluster into 2 large clusters (e.g., the happy and the unhappy respondents), and a small cluster of any oddball cases, that may be candidates for omission from the data set. (In my opinion, if it is done before any measurement or structural models are estimated it is still "good science"--e.g., similar to dropping incomplete, echeloned, etc. questionnaires--and structural equation analysis can be difficult enough without forcing a model fit data that include odd-ball cases).

Next, for a latent variable X with indicators x_1, x_2, etc., Z with indicators z_1, z_2, etc., and the interaction XZ, the indicators x_1, x_2, etc., and z_1, z_2, etc. should be mean- or zero-centered. This step is almost always required to avoid model estimation problems later. At a minimum, each indicator in each latent variable that comprises an interaction/quadratic should be centered (e.g., X and Z in XZ). This is accomplished by creating a new column of values for each centered indicator. For the centered x_1 (xc_1) column of values, for example, the column entries are computed by first obtaining the

[1] GA's, etc. can be unfamiliar with PRELIS, etc., but usually they know EXCEL. The EXCEL data can be saved a second time as a .prn file for (raw data) input to LISREL, etc. after some clean up in WORD (e.g., to remove row titles, etc.).

mean of the column of x_1 entries ($M(x_1)$). Then, this mean is subtracted from the x_1 value in row 1, for example, and placed in row 1 of the xc_1 column (e.g.,

	x_1		xc_1	
	(a)		= (a) - $M(x_1)$	
	(b)		= (b) - $M(x_1)$	
	...		...	
	=Average((a), (b), ...) (=$M(x_1)$)			

where (a), (b), etc. are values in the x_1 column, "= Average((a), (b), ...)" is the EXCEL command for the average of the x_1 column (abbreviated by "=$M(x_1)$"), and "= (a) - $M(x_1)$," for example, is the EXCEL calculation (a) minus the average of all the x_1 column entries). After computing each row of xc_1, the process is repeated for the xc_2, etc., zc_1, zc_2, etc. columns. These centered indicator values, xc_1, xc_2, etc. will be used to itemize X in the interaction/quadratic model instead of x_1, x_2, etc., zc_1, zc_2, etc.; and zc_1, zc_2, etc. will be used to itemize Z in the interaction/quadratic model instead of z_1, z_2, etc.

Next, I will illustrate creating the values for the single indicator for the Ping (1995) interaction/quadratic specification. They are created in 3 steps. First, a new column of averaged centered X indicators (xcAvg) is created with a row 1 entry that is the average of the values of the centered indicators of X in that row. Then, for the row 2 entry of the xcAvg column, the column xc_1, xc_2, etc. values in row 2 are averaged, etc. (e.g.,

	xc_1	xc_2	etc.	xcAvg	
	(c)	(e)		= Average((c), (e), ...)	
	(d)	(f)		= Average((d), (f), ...)	
	...	...		...	

where (c) = (a) - $M(x_1)$, (d) = (b) - $M(x_1)$, etc. (as above), "= Average((c), (e), ...)" is the EXCEL row average of the xc_1, xc_2, etc. values in the first row, "= Average((d), (f), ...)" is the EXCEL row average of the xc_1, xc_2, etc. values in the second row, etc.).

Next, the same thing is done for Z creating a column of averaged centered Z indicator values (zcAvg) (e.g.,

	zc_1	zc_2	etc.	zcAvg	
	(g)	(i)		= Average((g), (i), ...)	
	(h)	(j)		= Average((h), (j), ...)	
	...	...		...	

where (g), for example, is the first row value of z_1 minus the average of the column of z_1 values, "= Average((g), (i), ...)" is the EXCEL row average of the zc_1, zc_2, etc. values in the first row, "= Average((h), (j), ...)" is the EXCEL row average of the zc_1, zc_2, etc. values in the second row, etc.).

The third column contains entries in each row that are the product of the xcAvg value and the zcAvg value in that row (xcAvg*zcAvg). Specifically, in row 1 of the xcAvg*zcAvg column is the product of the row 1 value from column xcAvg times the row 1 value from column zcAvg. Next, row 2 of the xcAvg*zcAvg column contain the row 2 value of xcAvg times the row 2 value of zcAvg, etc. (see Exhibit A).

Then, a new (replacement) (raw) data set is created that contains the original columns for the indicators of the latent variables other than X and Z. However, for the indicators of X and Z the xc_1, xc_2, etc. columns, and the zc_1, zc_2, etc. columns, are substituted for the x_1, x_2, etc. and z_1, z_2, etc. columns. The xcAvg and zcAvg columns are omitted, but the xcAvg*zcAvg column of values for the interaction indicator are included. Specifically, if y_1, y_2 and y_3 are the indicator columns for the latent variable Y, the table

Exhibit A--Rearranged Spreadsheet for Input to LISREL, SPSS, etc. (After Cleanup in WORD)

xc_1	xc_2	xc_3	zc_1	zc_2	zc_3	y_1	y_2	y_3	xcAvg*zcAvg
(c)	(e)	(k)	(g)	(i)	(m)	(o)	(q)	(s)	[(c+e+k)/3]*[(g+i+m)/3]
(d)	(f)	(l)	(h)	(j)	(n)	(p)	(r)	(t)	[(d+f+l)/3]*[(h+j+n)/3]
...	...	...	...	...	...				...

is output to a file for the measurement and structural model estimations, where (c), (e), etc. are as before., and (o) - (t) are the values for the y's.

Then, see "Questions of the Moment," "Is there an example that shows all the steps..." for the next steps.

QUESTIONS of the MOMENT...

"Why would applied researchers be interested in interactions/quadratics?"

(The **APA citation** for this paper is Ping, R.A. (2008). " Why would applied researchers be interested in interactions/quadratics?" [on-line paper]. http://www.wright.edu/~robert.ping/applied.doc)

Anecdotally, applied researchers sometimes face the same "difficulties" as theoretical model testers: predictors that should be important in "the model" when it is applied to the present data are suddenly not important. Equivalently, after "the model" is extensively calibrated, it does not predict very well in the present situation.

Failing to consider interactions or quadratics in a predictive can produce these results. For example, with a significant interaction in

$$Y = b_0 + b_1X + b_2Z + b_3XZ + b_4XX + \zeta \,, \qquad (1$$

(where ζ is estimation error) the coefficient of Z in Equation 1 factors to

$$Y = b_0 + b_1X + (b_2 + b_3X)Z + b_4XX + \zeta \,, \qquad (1a$$

rather than b_2 (see Aiken and West, 1991). Because the "factored" coefficient of Z in Equation 1a is now a variable that depends on the various levels of X in the data set, the magnitude, sign, and statistical significance of $b_2 + b_3X$ are variable, and thus very different from the coefficient of Z in an Equation 1 without XZ (i.e., b_2' in

$$Y = b_0' + b_1'X + b_2'Z + \zeta' \,). \qquad (1b$$

Specifically, b_2' could be nonsignificant, while $b_2 + b_3X$ could be significant over part(s) of the range of X in the data set. In this event, it is not the case that Z does not predict Y in the data set. The Z-Y association is simply conditional, and its significance depends on the various levels of X observed in the data.

This has important implications. Because b_2' is approximately the same as $b_2 + b_3X_{avg}$, where X_{avg} is the average of X in the data set (see Aiken and West, 1991), if X_{avg} is low (small), $b_2 + b_3X_{avg}$ may be numerically small, and thus b_2' may be nonsignificant. In different words, Z may appear to be unrelated to Y in an Equation 1b model, when in the Equation 1 model, for larger values of X in the study $b_2 + b_3X$ may be significant as anticipated. This also implies that with a significant interaction XZ in Equation 1, the Z-Y association in Equation 1b may be significant in the next data set if X_{avg} is larger (higher) in that data.

Alternatively, with a significant interaction in Equation 1, b_2' could be significant, but $b_2 + b_3X$ could be nonsignificant over part of the range of X in a data set. In this case, the customary interpretation based on the significance of b_2', that Z was associated with Y, is incorrect: there is a set of cases where changes in Z had no association with Y. Further, any "recommendations" based on the apparently significant Z-Y association in Equation 1b due to b_2' may be misleading. Again, there is a set of cases where "managing" (changes in) Z had no effect on Y.

Similarly, with a significant quadratic such as XX in Equation 1, Equation 1 can be refactored into $Y = b_0 + (b_1 + b_4X)X + b_2Z + b_3XZ + \zeta$, and the coefficient of X is given by $b_1 + b_4X$, rather than by b_1' in Equation 1b. In this case, the relationship between X and Y depends on the particular level of X at which this association is evaluated. (Interpreting quadratics and "the association between X and Y depending on the particular level of X at which this association is evaluated" is discussed later.) As a result, b_1' could be significant while $b_1 + b_4X$ could be nonsignificant, or vice versa, which creates the same interpretation and implications issues as a significant interaction.

Thus, for improved understanding, explanation and prediction of model results, interactions and quadratics should be investigated. Specifically, they may provide explanations for important but nonsignificant associations, which avoids casting a shadow on the utility of the model because it appears not to apply in all cases, and it improves the interpretation of significant associations that may be conditional.

However, given the tediousness of identifying significant interactions or quadratics among latent variables, researchers who want to probe for these variables may decide to use approaches such as ordinary least squares (OLS) regression or analyzing subsets of data (e.g., median splits). Unfortunately regression estimates of interaction or quadratic structural coefficients for latent variables are well known to be biased (see the demonstrations in Aiken and West, 1991). Similarly, subset analysis is criticized in the psychometric literature for a variety of reasons, including its reduced ability to detect interactions or quadratics (see Maxwell and Delaney, 1993 and the citations therein).

There are other estimation concerns as well, such as correlations among interactions and quadratics that can produce no significant interactions or quadratics when several interactions or quadratics are estimated jointly, and extant search techniques, such as forward selection and backward elimination, can be indeterminate in that they can produce different subsets of significant interactions and/or quadratics.

These matters are discussed in the working paper "Hypothesized Associations and Unmodeled Latent Variable Interactions/Quadratics: An F-Test, Lubinski and Humphreys Sets, and Shortcuts Using Reliability Loadings" on the previous web page.

QUESTIONS of the MOMENT...

"Why are reviewers asking about reliability and validity in my interaction, XZ, and is there any way to improve XZ reliability and validity?"

(The APA citation for this paper is Ping, R. A. (2017). "Why are reviewers asking about reliability and validity in my XZ interaction, and is there any way to improve these matters?" [on-line paper]. http://www.wright.edu/~rping/ImprovXZ_AVEa.doc) .

Reliability and validity in XZ are discuss in detail in, for example, "What is the Average Variance Extracted for a Latent Variable Interaction (or Quadratic)?" and "What is the "validity" of a Latent Variable Interaction (or Quadratic)?" in QUESTIONS OF THE MOMENT on this web site. Briefly, XZ reliability is necessary for XZ validity, and because there is little agreement on what constitutes an adequate demonstration of validity. a minimal demonstration of XZ's validity should probably include the content or face validity of its indicators (how well they tap into its conceptual definition—in this case its itemization), XZ's, and its convergent and discriminant validity (its correlations with other model LV's and acceptable AVE).

The reliability and validity of XZ are as important in my opinion as that of X or Z. At the risk of overdoing it, if XZ is not reliable it is *per se* not valid, and if this is the case, one is not performing an adequate disconfirmation test of any X or Z moderation hypothesis. Specifically, and as discussed elsewhere on this web site, lack of XZ content or face validity disables the hypothesized model's test of a moderation hypothesis (because of the "factoring" difficulty—see below). Lack of convergent validity handicaps a moderation hypothesis test because any observed moderation association is likely to be false positive or false negative due to measurement error. In different words, lack of XZ reliability and validity casts a shadow on a disconfirmation moderation test.

In a convenient sample of substantive articles that estimated an LV interaction, none specifically addressed XZ reliability or validity. (Parenthetically, in my own substantive articles--see www.wright.edu/~robert.ping/rt.htm--this matter also was not specifically addressed.) (As I recall, most of my substantive LV interactions were valid and reliable—these numbers were on the EXCEL spreadsheet on my web site that I used to calculate LV interaction/quadratic loadings and measurement errors after 2003, and I formally raised the matter on this web site beginning about 2005. However, because I could/can find no citation in the literature for calculating an interaction (or quadratic) AVE other than Ping, R.A. (2005), and an exploratory factor analysis "AVE" is approximate (see "Is there any way to improve Average Variance Extracted (AVE) in a Latent Variable (LV) X?"), my substantive articles never specifically mentioned XZ reliability and validity.)

However, reliability and validity may not apply to XZ. I recall a colleague objecting that an LV interaction was not a "construct. Further, specifying X's moderation of Z-Y as XZ also implies testing the assumption that the moderation is of the form

$$Y = ...+ aX + bZ + cXZ +...$$

$$= ...+ (a+cX)Z + bz +...$$

(The moderation also could be of the form

$$Y=...+ aX + bZ + cZ/X +...$$

$$= ...+ (a+c/X)Z + bZ +...)$$

While I have yet to resolve these objections satisfactorily (e.g., can products of observed indicators actually be themselves "observed" as, for example x_i? And, the specification of c/X as a moderator is unexplored), specifying a XZ as a construct (i.e., with indicators), evaluating the reliability and validity of XZ out of concern for false negative or false positive results due to its measurement error, and specifying a moderation hypothesis with XZ may be the best we presently can do to test a moderated hypothesis involving two LV's.

Also as pointed out elsewhere on this web site, in real world data, reliable and valid X and Z are likely to produce low reliability and validity in XZ. (The disinterested reader may want to skip the following support of this statement and skip to A) below.)

The reliability of XZ is given by

$$\rho_{XZ} = \frac{r_{XZ}^2 + \rho_X\rho_Z}{r_{XZ}^2 + 1} , \qquad (1$$

where ρ denotes reliability and r_{XZ}^2 is the correlation of X and Z. Similarly, the reliability of XX is

$$\rho_{XX} = \frac{r_{XX}^2 + \rho_X\rho_X}{r_{XX}^2 + 1}$$

$$= \frac{1 + \rho_X^2}{2} \qquad (2$$

(Busemeyer and Jones, 1983). Thus, the reliability of an interaction or quadratic is a function of the product of reliabilites, and XZ is unlikely to be reliable unless X znd Z have reliability in the 0.84 range. This also suggests that XZ reliability could be improved by improving X's and/or Z's reliability.

The AVE of XZ is derived on this website (see "What is the Average Variance Extracted for a Latent Variable Interaction (or Quadratic)?"), and is

$$AVE_{XZ} = \frac{(\Sigma[\lambda_{xi}\lambda_{zj}]^2)Var(XZ)}{(\Sigma[\lambda_{xi}\lambda_{zj}]^2)Var(XZ)+\Theta_{XZ}} \qquad (3$$

$$= \frac{\Sigma[\lambda_{xi}\lambda_{zj}]^2}{\Sigma[\lambda_{xi}\lambda_{zj}]^2 + \Theta_{XZ} / Var(XZ)} , \qquad (4$$

where xi and zj are indicators of X and Z, λ is loading, Var is variance, and Θ is measurement error. Thus, XZ AVE is improved by reducing measurement error in XZ.

And, because

$$\Theta_{XZ} = [\Sigma\lambda_{xi}^2]Var(X)\Sigma Var(\varepsilon_{zj}) + [\Sigma\lambda_{zj}^2]Var(Z)\Sigma Var(\varepsilon_{xi}) + [\Sigma Var(\varepsilon_{xi})]\Sigma Var(\varepsilon_{zj}) \qquad (5$$

$$= [\Sigma\lambda_{xi}^2]Var(Z)\Sigma Var(\varepsilon_{xi}) + [\Sigma\lambda_{xi}^2Var(X) + \Sigma Var(\varepsilon_{xi})]\Sigma Var(\varepsilon_{zj})$$

where ε_{xi} and ε_{zj} are measurement error (see "What is the Average Variance Extracted for a Latent Variable Interaction (or Quadratic)?"), suggests that XZ AVE is improved by reducing measurement error in Z. Similarly, factoring by $\Sigma[Var(\varepsilon_{xi})]$ suggests XZ AVE is improved by reducing measurement error in X. Because AVE_X and AVE_Z also have the Eqation 3 form, increasing AVE in X or Z reduces measurement error in X and Z, and increases XZ AVE. (Parenthetically, this suggests AVE is increased by increasing VAR(XZ), and since

$$Var(XZ)=Var(X)Var(Z) + Cov^2(X,Z) \qquad (6$$

increasing Var(X), Var(Z) and/or Cov(X,Z) increases AVE in XZ—see Equation 4.)

A) Nevertheless, reliability and AVE of XZ can typically be improved by improving the reliability and AVE of X and/or Z. As suggested in "Is there any way to improve Average Variance Extracted (AVE) in a Latent Variable (LV) X?" in QUESTIONS OF THE MOMENT, dropping cases may improve AVE in X and Z. However the process is tedious, and experience suggests that in real world data the improvement may be limited (about 0.02-0.05 in AVE).

An alternative is to delete XZ indicators. However as mentioned elsewhere the resulting XZ would not be Face or Content Valid because it is not itemized as

XZ = (x1+x2+...+xn) (z1+z2+...+zm)

= x1z1+x1z2+...+x1zm+x2z1+x2z2+,,,+x2zm+...+xnz1+xnz2+...xnzm).

This is important because with a weeded XZ an hypothesized X moderation of Z-Y in the model Y=...+aX+bZ+cXZ+...= ...+(a+cX)Z+bZ+... (Aiken and West 1991) (i.e., the coefficient of Z depends on the level X in (a+cX)) is not tested because XZ can no longer be factored—the Z in bZ is not the same as Z in cXZ.

As mentioned above (see Equations 4 and 6) , it also may be possible to increase XZ AVE by increasing X and Z's correlation. This might be affected by using the scaling procedure described in "Is there any way to improve Average Variance Extracted (AVE) in a Latent Variable (LV) X?" (by adding, instead of subtracting, a the scaling factor to increase, instead of decrease, correlation between X and Z, in order to increase the AVE of XZ—see Equations 4 and 6). (I have not tested this procedure with real world data for several reasons including it may be difficult to justify this to reviewers when reitemizing X and Z using the EXCEL procedure mentioned above might be more appropriate. (However, see below.) In addition rescaling may just "mask" the real problem with XZ: insufficient error free variance in X and Z.)

If XZ resists the above attempts at increasing reliability and AVE sufficiently, a Median split of the data might be used to avoid the XZ reliability and AVE matter altogether. However, this approach is discouraged in the regression literature (e.g., Aiken and West 1991, Cohen 1983) because splitting data substantially reduces statistical power, and thus increases the likelihood of a False Disconfirmation (False Negative) finding. (Nevertheless, see "Why is my hypothesized interaction or quadratic non significant?" for more on median splits, and see below.)

Experience suggests that in real world data reitemizing X or Z to increase AVE can increase AVE of XZ (see "Is there any way to speed up 'item weeding' and find a set of items that jointly fits the data, contains more than 3 items, and is valid and reliable?" in QUESTIONS OF THE MOMENT). However, the procedure is tedious, even using the EXCEL spreadsheet, and if several adequate itemizations are found, it may be difficult to judge face valid of X and/or Z without an item-judging panel.

In summary, XZ reliability and validity are (or should be) as important as reliability and validity of X and Z. Unfortunately, XZ reliability and validity are usually (much) less that that of X or Z, and the remedies discussed above each have their drawbacks. For example nearly all are tedious, some trade one validity for another, some ignore reliability and validity but sacrifice power of the test, others are untested.

Nevertheless, experience suggests reitemizing X and Z using the (tedious) EXCEL worksheet procedure works sufficiently well to qualify as a first-choice remedy for XZ reliability and validity difficulties. However, if this fails to produce adequate XZ reliability and AVE, one might consider estimating XZ to see if it is significant. If so, one could try a median split.

If that is not significant, one could use the scaling procedure mentioned above to "probe" the effect of XZ's significance in the presence of low reliability/AVE. Specifically, at what point does *increasing* XZ measurement error (or the effects of measurement error decreased by X-Z covariance) make the XZ-Y association NS? In different words, how much measurement error is required (in the present case) to render the XZ-Y association NS. If it's a lot, one might argue that unacceptable XZ reliability/AVE may not matter that much in the next model test: in the present model the significant XZ-Y association appeared to be "robust to departures from inadequate reliability/AVE."

(I have not investigated this approach, and it is plausible that there may be a comparatively low "tipping point" for lack XZ-Y association when X-Z covariance is increased. In other words, XZ may not be very "robust." (Or, there may be no change) Further, even though "adequate" reliability is 0.70 and "acceptable" AVE is 0.50, these values could be viewed as "arbitrary." Nevertheless, they are the "rules used to judge these matters," and reviewers may or may not be willing to "bend" these rules.)

(Parenthetically, this suggests an additional criterion for judging the itemization of X and Z—consider the covarance of X and Z when initially itemizing X and Z.) (Maximizing X and Z's covariance could obviously be a criterion in reitemization, but it would seem formally more appropriate to do this when first choosing items.)

If "bending" the rules is not acceptable, one could argue for in effect "suspending" XZ reliability and validity concerns. The logic would be that the model may be sufficiently interesting to not suppress its first test. In different words, the primary focus of the (any) paper is on the new theory developed, and that contributions include a "first test;" and that in this case more measurement work is simply needed on the low AVE X and Z measures. (A less desirable alternative with low AVE would be to drop the XZ hypothesis, or create a propositional paper, with the caveat that both might be received as considerably less "interesting.")

REFERENCES

Aiken and West 1991)
Busemeyer and Jones, 1983
Cohen 1983
Ping, R.A. (2005). "What is the average variance extracted for a latent variable interaction (or quadratic)?" [on-line paper]. http://www.wright.edu/~robert.ping/ave1.doc (formerly Ping, R.A. (2005). "What is the average variance extracted for a latent variable interaction (or quadratic)?" [on-line paper]. http://home.att.net/ ~rpingjr/ave1.doc.)

QUESTIONS of the MOMENT...

"When theory proposes an X-Y association and it also proposes that Z moderates this association, but theory is mute about or doesn't propose a Z-Y association, why does one still include Z in addition to X and XZ in the model to be tested?"

(The **APA citation** for this paper is Ping, R.A. (2009). "Why must one always include Z in addition to X and XZ in the model to be tested?" [on-line paper]. http://www.wright.edu/~robert.ping/MissingZ.doc)

By hypothesizing Z's moderation of the X-Y association but excluding the Z-Y association in the model to be tested, one is assuming that Z has no relationship with Y. However, reviewers will likely request support (argumentation) for this assumption, and validation of these arguments by testing (including) the Z-Y association it in the model. There are several reasons for this, including that the logic of science requires an argument (hypothesis) to support an assumption such as the Z-Y association does not exist in the population. This in turn requires an hypothesis test (i.e., including the Z-Y association in the model to be tested).

Omitting the Z association also creates the "missing variable" problem (see James, 1980), which can bias all structural coefficients, and standard errors in the model. This in turn casts a shadow on the trustworthiness of the test of the proposed model. (The Z-Y association will never be exactly zero in real-world data, and the fitted covariance matrices for the structural models with and without Z will be slightly, to very different, depending on the covariances of Z with the other model variables.) (In addition, missing variables are accounted for in the error term(s) (structural disturbances) of the dependent variable(s). Because these missing variables are almost always correlated with other antecedent variables in real-world data, when Z is missing, error tem(s) are correlated with antecedent variables in the model, which is a violation of an important structural equation analysis assumption.)

In addition, if XZ is significant, excluding Z from the model biases the factored ("true" contingent) association of Z with Y, EVEN IF A Z-Y ASSOCIATION IS HYPOTHESIZED TO NOT EXIST IN THE POPULATION. In $Y = b_1X + b_2XZ + b_3Z$, the factored (true contingent) coefficient of Z is $(b_2X + b_3)Z$. Since b_3 will never be zero in real-world data, this factored coefficient of Z and its significance will be biased by omitting Z. Further, excluding Z misses an opportunity to "explore" (discuss) the novel fact that Z, even if it is supposed to be nonsignificant, is significantly (conditionally) associated with Y.

Thus, if theory suggests that the Z-Y association is zero in the population, or the theory is mute on this matter, one should consider adding a formal argument, including an hypothesis, that Z has no association with Y to the paper. Then, one should consider including the Z-Y association in the model to be tested (to test this hypothesis).

Optionally, to gauge the effect (significances) of excluding Z from the model, one could rerun the model with Z omitted. (This practice, called "trimming," was once popular in Sociology, but it is now rarely seen in Social Science.) In addition, if XZ is significant, one could also compute the factored coefficient of Z and its significances at various levels of Z in the sample, then discuss this "discovery" in the Discussion section of the paper.

References

James, Lawrence R. (1980), "The Unmeasured Variables Problem in Path Analysis," *Journal of Applied Psychology*, 65 (4), 415-421.

QUESTIONS of the MOMENT...

"Why is my hypothesized interaction or quadratic non significant?"

(The **APA citation** for this paper is Ping, R.A. (2007). "Why is my hypothesized interaction or quadratic?" [on-line paper]. http://www.wright.edu/~robert.ping/NS.doc)

It is usually a good idea to verify that model specification and estimation are not the "problem." Specifically, consider verifying that the model is properly specified (e.g., the correlations among the exogenous variables, including the interactions, are free; the correlations between exogenous variables and endogenous variables are not free; structural disturbances are not correlated, etc.). Then, check that the interactions are properly specified (e.g., the "essential" correlations between X-XZ and Z-XZ, for example, are free, the variance of XZ is also free, and the values for the loading and measurement error variance have been properly calculated and keyed into the estimation program), and the model indicators are all mean- or zero-centered. Next, verify that the structural model fits the data, all the coefficient estimates are admissible (see Step VI, "Admissible Solutions Revisited" in the *Testing Latent Variable Models Using Survey Data* monograph on this web site), and the measurement parameters of X and Z in the structural model (i.e., the loadings, measurement error variances and the variances of X and Z) are within a few points of their measurement model values. If the measurement parameters of X and Z in the structural model are different from their measurement model values, recalculate the interaction's loading and measurement error variance using the structural model measurement parameter values.

If the structural model and its estimates check out, the possible next steps are several. However, it turns out that there are infinitely more (mathematical) forms of an interaction besides XZ. I once proposed to find several dozen interaction forms besides XZ, and eventually suggested XZ^w, where W can be any (positive or negative) number. These interaction forms include not only XZ (w = 1), they also include X/Z (see Jaccard, Turrisi and Wan 1995) (as Z increases in the study X is attenuated--w = -1). They also include XZ^2, the interaction between X and the square of Z (see Aiken and West 1991), and they include XX^w, where Z = X and X is moderated by itself (which is called a quadratic when w = 1). Thus, an hypothesized interaction may not have the form XZ, and specifying it in that form may produce NS results for the hypothesized interaction.

Unfortunately, in this case the options become limited. It may be prudent to test the hypothesized interaction in the form XX (X interacts with itself). However, if XX is NS, interaction specifications besides XZ and XX are unknown at present. Nevertheless, it may be efficacious to test a median split of the data. If this is significant, it suggests the hypothesized interaction was significant, but its form was not easily determined (XZ and XX were NS).

One other possibility remains if the median split was significant and XZ and XX were NS. The interaction may "approach significance"--its t-value is in some neighborhood of 2 using maximum likelihood estimation. This is usually the result of low XZ (or XX) reliability or insufficient sample size. Interaction reliability can be verified using the EXCEL spreadsheets on the previous web page. Insufficient sample size can be checked by calculating the sample size (N) that would have been required to produce a t-value of 2 using the equation $N = 4*n/t^2$, where n is the current sample size, "*" indicates multiplication, and t^2 is the square of the current t-value. If the reliability of XZ is 0.7 or above, and a few more cases would push the t-value above 2, it might suffice to simply state that the interaction "approaches significance," and proceed as though the interaction were significant. No statistical assumptions are violated by declaring that, for example, t greater than 1.95 in absolute value suggests significance. It is simply conventional in structural equation analysis to declare a structural coefficient twice or more the size of its standard error to be significant. In regression studies and correlational analysis there are two conventions for significance, p-value = 0.05 and p-value = 0.10. A p = 0.05 corresponds to t-value = 1.97 with 200 degrees of freedom (df), and p = 0.10 corresponds to t = 1.65 with 200 df.

Thus, depending on reviewers, it *might* suffice to state that the interaction "approaches significance" and no more. If challenged it might be useful to respond by computing the p-value for the t-value-less-than-two and the model degrees of freedom, and comparing it to the p-value of t = 2 with the model degrees of freedom. For example, if the target structural coefficient has a t-value of 1.97 with 267 degrees of freedom, the corresponding p-value is roughly 0.0499, which is close to the p-value for t = 2 with 267 df, 0.0465. A fallback position is to preface any discussion, implication, etc. involving the target coefficient with "if significant in future studies..." This preface is actually the preferred opening remark for any discussion of the model test results because it is well known that a single study "proves" nothing. It merely suggests what future studies may observe.

QUESTIONS of the MOMENT...

"Is there any way to improve Average Variance Extracted (AVE) in a Latent Variable (LV) X?"

(The APA citation for this paper is Ping, R. A. (2009). "Is there any way to improve Average Variance Extracted (AVE) in a Latent Variable (LV) X (Revised)?" [on-line paper]. http://www.wright.edu/~robert.ping/ImprovAVE2.doc)

(Click here for an earlier version of this paper, Ping, R. A. (2007). "Is there any way to improve Average Variance Extracted (AVE) in a Latent Variable (LV) X?" [on-line paper]. http://www.wright.edu/~robert.ping/LowAVE.doc.)

Average variance extracted (AVE) almost always can be improved by dropping cases, or by dropping the item with the largest measurement error variance. The result may be desirable to improve AVE, or to raise it above the square of a correlation with another latent variable (LV) (i.e., to improve discriminant validity in the Fornell and Larker (1981) sense).

One approach to dropping cases is to use a "Jacknife-like" procedure (Efron 1981). Specifically, a case is removed from the data set, and AVE is computed for the remaining cases.[1] Then, the removed case is replaced, a different case is removed, and AVE is computed for the remaining cases. This process is repeated for each of the rest of the cases to find the case that produces the largest AVE improvement.[2]

Additional AVE improvement may be obtained by repeating this process using the improved AVE data set (i.e., with the case that produces the largest AVE improvement removed), instead of the full data set. Specifically, a case is removed from the first improved AVE data set, and AVE is computed for the remaining cases. Then, the case just removed (not both cases) is replaced, a different case is removed, and AVE is computed for the remaining cases. This process is repeated for each of the rest of the cases to find the largest AVE improvement with two cases removed (but, see Footnote 2).

This process could be repeated using combinations of the above and Footnote 2) procedures, but experience suggests that dropping about three cases or approximately a .05 AVE improvement is the most AVE improvement that dropping cases will produce in real-world data.

---

[1] In a (maximum likelihood) exploratory factor analysis, the "Percent of Variance Explained" by Factor 1 can be used to gauge changes in AVE in a unidimensional set of items--it is roughly the same as AVE.

[2] Dropping the case that detracts most from AVE is arguably "not random," and this casts a shadow over sample "representativeness." An improvement would be to randomly select a case from the set of cases that detract most from AVE. Alternatively, cases could be deleted randomly and the first case that improves AVE could be dropped.

Dropping an item also will improve AVE, frequently by more that deleting cases will. However, the procedure for this is "messy," and the resulting set of higher AVE items can be less content or face valid than before items were dropped (i.e., the resulting set of items may match its conceptual and/or operational definitions less well). The results also may be less internally consistent (i.e., the single construct measurement model of the resulting items may fit the data less well).

Experience suggests that in real-world data, several consistent subsets of items from a measure usually can be found. An alternative to dropping items is to create one or more additional subsets of items and gauge the AVE of each these subsets. Replacing any deleted cases, (maximum likelihood) exploratory factor (with varimax rotation) the full measure. Then for the Factor 1 items, find highest reliability subset of these items (there are procedures in SPSS, SAS, etc. to accomplish this, usually in the "reliability" procedures). If the AVE of the resulting highest reliability subset of items is unacceptable, try dropping cases from this subset of items. (Dropping the item with the largest error term from the highest reliability subset of items usually reduces AVE.)

If model-to-data fit or content validity problems arise when items are dropped, combinations of one or more of the following procedures could be used.

Replacing any deleted cases, consider finding another consistent subset of items using Modification Indices (see Appendix A--Item Weeding in "On the Maximum of About Six Indicators..." on this web site). If this subset has unacceptable AVE, try dropping the item with the largest error, and/or drop case(s). Experience suggests that Modification Indices sometimes works better than maximizing reliability. However, AVE improvement seems to be limited to about 10 points (i.e., 0.10).

If AVE is still unacceptable, replace any deleted cases, then, using the Factor 1 items, drop the item with the largest error term first. Then, reitemize the resulting set of items using Modification Indices or by maximizing reliability. Next, drop cases, and/or drop the item with the largest error from the results.

However, if AVE of the resulting measure is within a few points of "acceptable" (0.50), this may not always be "fatal" to publishing a model test. Experience suggests that not all reviewers accept AVE as "the" measure of convergent validity, some prefer reliability. Thus, if an LV is reliable, that may be a sufficient demonstration of convergent validity for some reviewers.

In addition, the logic for possibly ignoring low AVE might be that many "interesting" theoretical model-testing studies involve a "first-time" model, and an initial model test, that together should be viewed as largely "exploratory." This "first test" usually uses new measures in a new model tested for the first time, etc., and insisting that the new measures be "perfect" may be inappropriate because new knowledge would go unpublished until a "perfect" study is attained. AVE adherents of course might reply that concluding anything from measures that are more than 50% error is ill advised, because there are so few replication studies.

In my opinion, an AVE slightly below 0.50 might be acceptable in a really "interesting" "first-time" study, 1) if it does not produce major discriminant validity problems (discussed below), 2) the diminished AVE is noted and discussed in the Limitations section of the paper, 3) any significant effects involving the low AVE LV's are held to a higher significance requirement (e.g., $|t| >= 2.2$ rather than $|t| >= 2.0$), and 4) any discussion of interpretation, and especially implications, involving the low AVE LV's are clearly labeled as "very provisional" and in need of replication.

Again, the logic would be that the model may be too interesting to suppress its first test. In different words, the focus of the paper should be on the new theory developed, and the contributions include a "first test," and that more measurement work is needed on the low AVE measures. (A less desirable alternative with low AVE would be a propositional paper, which might be considerably less "interesting.")

This "first-time study" argument also may apply when there are discriminant validity problems, and more measurement work is needed on the low discriminant validity measures. Nevertheless, it always possible to reduce the correlation between two LV's using a procedure similar to Residual Centering (see Lance 1988). The procedure involves reducing the covariation between the target LV's X and Z until the squared correlation between them is less than the AVE of both X and Z. Specifically, average the indicators for X and Z, then regress the lower AVE LV, X for example, on the higher AVE LV, Z, to produce $Z = b_0 + b_1X$. Next, subtract a percentage of $b_0 + b_1X$ (i.e.,

1) $K^*(b_0 + b_1X)$,

where K is between 0 and 1) from Z in each case to scale (reduce) the covariance (and thus the (squared) correlation) between X and Z.

However experience suggests that in real-world data, scaling simply masks a discriminant validity problem rather than remedying it. Specifically, in real-world data, experience suggests that with lower AVE and correlated X and Z, the unique error variance (i.e., error variance that is unshared in the correlation between X and Z) of one or both X and Z can be greater than 50%. This in turn increases the instability (variability) of structural coefficients involving X or Z across studies beyond that which could be expected with sampling variation. Stated differently, with declined AVE and (even moderately) correlated X and Z, their association with Y, for example, can be largely the result of measurement error, which should produce different results (i.e., instability--reduced "reproducibility" in Campbell and Fiske's (1959) terms) in subsequent studies. Increased correlation between X and Z, especially when it is greater than X or Z's AVE, increases this potential for instability.

1) As a perhaps surprising example from a real-world survey, the AVE's of two LV's were both 0.59, and their correlation was -0.59 (their covariance, the square of their correlation, was 0.33). Scaling one LV to zero correlation with the other, reduced its AVE to 0.47, and scaling the other LV to zero (after removing the previous scaling)

reduced its AVE to 0.48. Thus, the amount of unique error variance in the LV's was 53% (= 1 - 0.47) in one, and 52% (1 - 0.48) in the other. This suggests that any associations involving X or Z and a dependent variable is (slightly) more the result of error variance than it is the result of error-free variance, which should (slightly) amplify any difference in results from sampling variation in subsequent studies. Note that both LV's were discriminant valid using Fornell and Larker's (1981) AVE's-versus-squared-correlation discriminant validity criterion, and they likely would not have had their discriminant validity questioned--both LV's had "acceptable" AVE's, and their correlation was less than |0.70|. Also note that in this case the unacceptable unique err-free variances might be remedied by increasing AVE in the LV's.

As another example again using real-world data, the AVE's of two LV's were 0.58 and 0.72, while their correlation was .82 (their covariance, the square of their correlation, was 0.67). Scaling the larger AVE LV to zero correlation, reduced its AVE to 0.44. and scaling other LV (after undoing the previous scaling) reduced its AVE to 0.34. In different words, the amount of unique error variance in the larger AVE LV was 56% (= 1 - 0.44), and the amount of unique error variance in the smaller AVE LV was 66% (= 1 - 0.34). This suggests that their associations with another LV,Y for example, are more the result of error variance than error-free variance, which should produce more instability (different results) in subsequent studies than if it were lower.

Finally, the AVE's of two other LV's were 0.72 and 0.87, while their correlation was -.71 (their covariance, the square of their correlation, was 0.50). Scaling the larger AVE LV to zero correlation, reduced its AVE to 0.77, and scaling other LV (after undoing the previous scaling) reduced its AVE to 0.55. In different words, the amount of unique error variance in the larger AVE LV was 23% (= 1 - 0.77), and the amount of unique error variance in the smaller AVE LV was 45% (= 1 - 0.55). This suggests that their associations with another LV,Y for example, are more result of error-free variance than error variance, which should produce (comparatively) less instability (differing results) in subsequent studies than if error variance were higher. Note that both LV's had "acceptable" AVE's, but their correlation was slightly greater than |0.70|.

These examples suggest that Fornell and Larker's AVE's-versus-squared-correlation (discriminant validity) test may or may not signal a problem with unique error variance, and thus Fornell and Larker's discriminant validity test may or may not signal declined "reproducibility," Campbell and Fiske's stated objective of validity.

Several comments may be of interest. As the first example suggests, low AVE's should be investigated for low unique error-free variance. There probably can be no firm rule, but Fornell and Larker's "AVE at least 0.50" may be insufficient. Experience suggests that all correlations above 0.7 should be investigated (see Example 1 above), especially when the AVE's of the LV's involved are less than 0.6.

Any low unique error-free variance problems should be discussed in the Limitations section of the study's paper, and any discussion of the implications of the associations involved should be prefaced with a caveat that these associations are mostly error and may be an artifact of the study.

Z and its t-value is unchanged by scaling (i.e., subtracting K*(b_0 + b_1X) from Z in each case). However, scaling reduces the variance of Z, and thus it reduces any standardized structural coefficient (beta) involving Z. The range of Z is also reduced by scaling.

For an LV, X, that fails Fornell and Larker's AVE's-versus-squared-correlation (discriminant validity) test with Z, it is easy to show that all of X's error free variance is not contained in the covariance of X and Z. (X's AVE less than its correlation with Z might mean that all of X's error-free variance, its AVE, is contained in the covariance.)

For example when two LV's, with AVE's of 0.49 and 0.72, and a squared correlation of 0.65 (i.e., their covariance (squared correlation) was larger than one LV's error-free variance (AVE), a failure of Fornell and Larker's discriminant validity test), had the smaller AVE LV's variance scaled to zero correlation between them (i.e., K = 1 in Equation 1), in a measurement model of the resulting smaller AVE LV, it had 21% error free variance (i.e., a 0.21 AVE). In different words, all the covariation between the LV's was removed by scaling, yet there still was error-free variance (AVE) in both LV's (i.e., 21% error-free variance in the smaller AVE LV and 72% in the larger).

This could be interpreted as suggesting that the LV's were operationally distinct (the customary meaning of discriminant validity[3]). (In real-world data, only if the variance of an LV is equal to its covariance with another LV is there complete operational indistinctness--this matter is further discussed below).

In real-world data, experience suggests that improving an LV's AVE does not materially change correlations with that LV.

Substantive authors have used other single-sample discriminant validity tests besides Fornell and Larker's AVE's-versus-squared-correlation (discriminant validity) test, and these tests may be attractive when there are problems with discriminant validity. These tests include testing the correlation confidence interval (see Anderson and Gerbing 1988) or a single degree of freedom test (see Bagozzi and Phillips 1982). However, it is easy to show that these tests are likely to produce untrustworthy results in theory tests with survey data. Specifically, in theory tests with real-world survey data, testing the correlation confidence interval for two LV's to see if it contains 1, which would suggest that the two LV's are empirically (operationally) indistinct (i.e., they are "discriminant invalid" in the popular sense--see Footnote 3), almost always suggests empirical distinctness. Typical sample sizes (hundreds of cases) and the internal consistency requirement in survey data theory tests typically combine to produce small correlation

---

[3] A careful reading of Campbell's writings suggests that their notion of discriminant validity is evidenced by low correlations with other variables, rather than the popular requirement of a lack of population correlations of 1 (i.e., empirical or operational distinctness).

standard errors that in turn produce confidence intervals are usually too small to include a correlation value of 1.

For example, two LV's that were known to be theoretically indistinct (their items contained only slight variations in item wording) and that had a correlation of 0.9988, produced a 95% correlation confidence interval of [.9984, .9993] in a sample of 200 surveys. In different words, the 95% confidence interval for these LV's suggested they were operationally distinct (i.e., "discriminant valid") even though they were theoretically indistinct and had a correlation of .9988.

A single degree of freedom test can be applied to a measurement model containing the two target LV's, or it can be applied to the full measurement model containing the target LV's, to compare the model to one where the two LV's correlation is constrained to 1. In a two-LV measurement model, a single degree of freedom test frequently produces untrustworthy test results (i.e., two highly correlated LV's that have their correlation constrained to 1 will usually fit the data significantly worse that when their correlation is free). For example, in a two-LV measurement model, two LV's that may or may not have been theoretically distinct (they were two factors of the same LV) had a correlation of .9295. In a single degree of freedom test, their correlation could not be constrained to 1 in LISREL (the fitted covariance matrix was not positive definite). However, constraining the correlation to 0.9795 instead of 1 produced a chi square difference (chi square = 388 for the constrained correlation model, chi square = 207 for the unconstrained correlation model) with 1 degree of freedom (degrees of freedom = 54 for the constrained correlation model, degrees of freedom = 53 for the unconstrained correlation model) that was significant (chi square difference = 388 - 207 = 181, which has 1 - α of 1.0000 with 1 degree of freedom), suggesting they were (very) operationally/empirically distinct (and thus "discriminant valid").

Then, two sets of items, that could be argued to be conceptually the same (both were from Factor 1 of the same LV), with a correlation of 0.9969, were tested in a two-LV measurement model. Again in a single degree of freedom test, their correlation could not be constrained to 1 in LISREL. However, constraining the correlation to 0.9998 instead of 1 produced a chi square difference (218 = 249 for the constrained correlation model, minus 31 for the unconstrained model) with 1 degree of freedom (9 for the constrained correlation model, degrees of freedom = 8 for the unconstrained model) that was significant (1 - α = 1.0000 with 1 degree of freedom), which suggested they were (very) operationally distinct, and thus "discriminant valid."

However, the chi square statistic is sensitive to sample size. So, the sample size was reduced in steps until the parameter estimates became unstable, comparing chi square differences at each step. Nevertheless, the chi square difference tests continued to be significant, suggesting that sample size did not affect these results.

A full measurement model produced similar results. Two LV's that may or may not have been conceptually the same (they were two factors of the same LV) had a correlation of .9274. While their correlation could not be constrained to 1 in a larger measurement model containing them, constraining them to a correlation of .9476 produced a chi square difference test with a significance of .9990. This suggested they were operationally distinct (i.e., they were "discriminant valid").

Then, two sets of items, that could be argued to be conceptually indistinct (both were from Factor 1 of the same LV), with a correlation of 0.9998, were tested in a two LV measurement model. In this case their correlation could be constrained to 1, and the chi square difference (= 12 = (5626 for the constrained correlation model, minus 5626 for the unconstrained model)) with 1 degree of freedom (1836 for the constrained correlation model, 1835 for the unconstrained model) was significant (1 - α = .9999 with 1 degree of freedom), which suggested they were (very) operationally distinct, and thus "discriminant valid."

Again, because the chi square statistic is sensitive to sample size, the sample size was reduced in steps until the parameter estimates became unstable, comparing chi square differences at each step. Again, the chi square difference tests continued to be significant, suggesting that sample size did not affect these results.

In summary, experience suggests that in real-world data, alternative "discriminant validity" tests, such as correlation confidence intervals or single degree of freedom tests, are untrustworthy, usually suggesting "discriminant validity" even for nearly collinear LV's.

Discriminant validity in the (original) Campbell and Fiske (1959) sense (low correlations with conceptually distinct LV's) could be viewed in terms of the amount of unique err-free variance in correlated LV's after scaling, and the potential for instability of structural coefficient estimates. Stated differently, do AVE's and the covariance between two LV's combine to reduce the unique error-free variance of either (or both) (after scaling) to less than 50%, and thus increase the instability potential of these LV's structural coefficients? In two of the three examples above, AVE's and correlations combined to reduce unique error-free variance after scaling to questionable levels for "reproducibility" in Campbell and Fiske's (1959) terms.

For emphasis,

o with or without obvious discriminant validity problems in the Fornell and Larker sense (i.e., failure(s) of Fornell and Larker's AVE's-versus-squared-correlation (discriminant validity) test), lower AVE's can produce discriminant validity/reproducibility problems (i.e., AVE is insufficient to avoid an increased potential for structural coefficient instability) (see Example 1 above).

o Lower AVE LV's should be investigated for the possibility that their unique error-free variances are less than 50%. There probably can be no firm rule, but Fornell and Larker's suggestion that AVE should be above 0.50 may be insufficient with independent variable correlations above 0.30. Experience suggests that unique error-free variances should be investigated in all correlations larger than 0.7 (again see Example 1 above), especially if an AVE of the LV's involved is less than 0.60.

o Discriminant validity problems should be addressed by raising AVE, not by scaling.

o And finally, any unremedied low unique error-free variance problem should be discussed in the Limitations section of the study's paper, and any discussion of the implications of a significant unremedied low unique error-free variance LV should be prefaced with a caveat that these results are based on more than 50% unique error variance, and thus may be an artifact of the study.

References

Anderson, James C. and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 103 (May), 411-23.

Bagozzi, Richard P. and Lynn W. Phillips (1982), "Representing and Testing Organizational Theories: A Holistic Construal," *Administrative Science Quarterly*, 27 (September), 459-489.

Campbell, D. T. and D. W. Fiske (1959), Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," Psychological Bulletin, 56, 81-105.

Efron, B. (1981), Nonparametric Estimates of Standard Error: The Jackknife, the Bootstrap, and other Resampling Methods. *Biometrika*, 68, 589-599.

Fornell, Claes and David F. Larker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.

Lance, Charles E. (1988), "Residual Centering, Exploratory and Confirmatory Moderator Analysis, and Decomposition of Effects in Path Models Containing Interactions," *Applied Psychological Measurement*, 12 (2) (June), 163-175.

(End)

QUESTIONS of the MOMENT...

"Is there any way to speed up "item weeding" and find a set of items that jointly fits the data, contains more than 3 items and is valid and reliable?"

(The APA citation for this paper is Ping, R. A. (2017). "Is there any way to speed up 'item weeding' and find a set of items that jointly fits the data, contains more than 3 items and is valid and reliable?" [on-line paper]. http://www.wright.edu/~rping/ItemWeed.doc) .

(Note: This material supercedes the prior discussion of the "weeding" EXCEL spreadsheet that appeared on earlier versions of this web site—click here for that discussion.)

New measures will almost never "fit the data" using a single construct measurement model without dropping items to attain model-to-data fit. In addition, most well established measures developed before covariant structure analysis (LISREL, AMOS, etc.) became popular also will not fit the data without item weeding.

It turns out that with real world data, measures used with covariant structure analysis are usually limited to about six items (see discussions in Anderson and Gerbing 1984, Gerbing and Anderson 1993, Bagozzi and Heatherton 1994, and Ping 2008). One explanation is that correlated measurement errors, ubiquitous in survey data but customarily not specified in covariant structure analysis, eventually overwhelm model-to-data fit in single-construct and full measurement models as indicators are added to the specification of a construct. And, that usually happens with about six items per construct.

There are ways around item weeding, such as various item aggregation techniques (see Bagozzi and Heatherton 1994), but many reviewers in the Social Sciences do not like these approaches. Unfortunately, reviewers also may not like dropping items from measures because of concerns over face- or content validity (how well the items "tap" the conceptual and operational definitions of their target construct).

Anecdotally, several "ad hoc" consistency approaches are in use. A subset of consistent items usually can be found using a SPSS, SAS, etc. reliability "search." The reliability of all the Factor 1 items is computed using SAS, SPSS, etc., the item that contributes least to reliability is deleted, and the reliability of the remaining items is computed. This process is continued until deleting any item reduces reliability. However, with real world data the remaining items may or may not fit the data in a single construct measurement model, and these items may not be Convergent and Discriminant valid.

(Note: The following is adapted from Ping, R.A. (2004), *Testing latent variable models with survey data*, 2nd edition, [on-line monograph], http://www.wright.edu/~robert.ping/lv1/v.doc .)

Documented procedures for obtaining consistent measures include using ordered similarity coefficients (see Anderson and Gerbing (1982:454) and Gerbing and Anderson 1988). Ordered similarity coefficients may help identify inconsistent items. Alternatively, consistency can be attained using a procedure that involves internal and external consistency (e.g., Jöreskog 1993). The procedure involves estimating a single construct measurement model (i.e., one that specifies a single construct and its items) for each construct, then measurement models with pairs of constructs, etc., through estimating a full measurement model containing all the constructs. Items are omitted as required at each step to obtain adequate measurement model fit (and thus consistency because the process begins with single construct measurement models) while maintaining content or face validity. Standardized residuals, or "specification searches" (e.g., involving modification indices in LISREL/SIMPLIS or LMTEST in EQS) can also be used to suggest items to be omitted at each step to improve model-to-data fit.

Partial derivatives of the likelihood function (of the Structural Equation Model being estimated) with respect to the measurement errors of the model indicators (termed "FIRST ORDER DERIVATIVES" in LISREL) also can be used to suggest inconsistent items for deletion (see Ping, 1998, 2004). This approach involves the examination of the matrix of these derivatives from a single construct measurement model (i.e., one that specifies a measure with all its items). Then, the item with the largest summed first derivatives without regard to sign that preserves the content or face validity of the measure is omitted. The matrix of first derivatives is estimated without the omitted item, and the process is repeated until the single construct measurement model fits the data (see the Appendix below for an example).

Items with similarly sized summed first derivatives suggest that there are at least two consistent subsets of items, and my experience with this procedure and real-world data has been that it produces several comparatively large (i.e., about six) internally consistent subsets of item. The approach was inspired by Saris, de Pijper and Zegwaart's (1987) and Sörbom's (1975) proposals to improve model-to-data fit using partial derivatives of the likelihood function with respect to fixed parameters (i.e., to suggest paths that could be freed--e.g., modification indices in LISREL), and modification indices can be used in addition to partial derivatives.

(End of adapted material)

There is a “weeding” EXCEL template on this web site that helps find at least two subsets of measure items that fit the data in real world data. The spreadsheet can be used as follows: First, exploratory (common) factor analyze the target measure using Maximum Likelihood estimation and varimax rotation.

Next, estimate a single construct (confirmatory) measurement model using the Factor 1 items. If the first measurement model fits the data item omission is not required. If this measurement model does not fit the data, find the "First Order Derivatives" in the output. (I will assume LISREL 8, which requires "all" on the OU line to produce First Order Derivatives. As far as I know, other estimation packages produce statistics equivalent to First Order Derivatives. For example in SIMPLIS “First Order Derivatives” are available by adding the line “LISREL Output: FD.”). Paste the lower triangle of First Order Derivatives for "THETA-EPS" into the template making sure to retain the item names in order to figure out which item to drop (see the example on the template). Then find the largest value in the "Column Sum" column--it will be the same as the "Max =" value in the lower right corner of the EXCEL spreadsheet.

Now, reestimate the measurement model with the item having the largest "Overall Sum" omitted (call this Reestimation 1). Record the Chi Square and RMSEA values on the spreadsheet for reference.

There is no agreement on acceptable single construct measurement model fit. I use either a Chi Square that is slightly nonzero (e.g., 1E-07, not 0), or an RMSEA that is .08 or slightly below, but many authors would suggest much stronger fit criteria for single construct measurement models.[i]

If the unomitted items in Reestimation 1 do not fit the data, find the "First Order Derivatives" for "Theta-Eps" in the Reestimation 1 output. Paste these into the second matrix in the template, and record the Chi Square and RMSEA values.

Repeating this process, Chi Square eventually will approach zero, and RMSEA will decline to 0.08 or less (the recommended minimum for fit in full measurement and structural models--see Brown and Cudeck 1993, Jöreskog 1993). This should happen with about 7 or 8, down to about 5, remaining items. If acceptable fit does not happen by about 4 items, an error has probably been made, usually by omitting the wrong item.

Each subset obtained after Chi Square nears zero or RMSEA is below .08 is a candidate subset for "best," but because items are being dropped in each step, these smaller subsets are usually less face valid, and thus the first acceptable subset is usually the preferred one.

To search for another subset of consistent items, repeat the above process using "Modification Indices" for "Theta Epsilon" instead of First Derivatives. (The SIMPLIS command line is “LISREL Output: MI.”) The theory behind Modification Indices is different from First Derivatives, and a different subset usually results.

Returning to the full measure, if it is multidimensional, there may be several more consistent subsets that can be found by repeating the above procedures using the full measure's items instead of the Factor 1 items. Experience suggests these subsets are smaller, but they frequently include items from Factor 2, etc. and thus they may be judged more face valid. This process also can be used on any combinations of Factor 2 and Factor 3, etc. items.

Further, there may be more consistent subsets that can be found by omitting the next largest "Overall Sum" item instead of the "Max =" item. Specifically, the second largest item in Reestimation 1 could be omitted in place of the largest. Then, continuing as before omitting the largest "Overall Sum" items, The result is frequently a different subset of items that fits the data. Another subset can usually be found using this "Second Largest" approach using modification indices instead of first derivatives. Others can be found omitting the second largest overall sum item in Reestimation 2, instead of Reestimation 1, etc., with or without deleting the second largest in Reestimation 1. This "Second Largest" strategy can also be used on the full set of items.

Perhaps surprisingly, experience suggests that there are about N-things-taken-6-at-a-time combinations of items with real world data that will fit the data, where N is the number of items in the full measure. (There may be more, if 5, 4 and 3 item subsets are counted). For example in the Appendix, the original measure had 8 items, and there may have been about 8!/(5!(8-5)!) = 56 5-tem subsets of items that might fit the data.

(Parenthetically, in the Appendix example I stopped looking after finding 17 more adequately consistent subsets. I began by excluding item 1 from the full set of items with which to start the above procedure; then I used the above procedure to delete more items. Next I started over by replacing item 1 and excluding item 2, etc., etc. for 7 examinations (skipping the exclusion of item 4 which had in effect already been done for the Appendix example). Then, I repeated what I had just done using Modification Indices (for 8 more examinations). Finally, I examined 2 more 8 item sets retaining, and never deleting, the “gold item”—the item judged to be the “most” Content Valid—at every step using Partial Derivatives, then Modification Indices. None of the 17 additional consistent subsets was larger than 5 items, and a “panel” of judges was used to help select the “most” content valid subsets. The itemization finally selected was the one with the “gold item” and the highest AVE.)

Experience suggests the above First Derivative/Modification Indices approach usually identifies at least two “attractive” subsets of items that are comparatively large, and appear to adequately tap the target construct. However, as the Appendix example suggests, the above First Derivative/Modification Indices procedure (beginning with all items) may not always produce the highest reliability/AVE subsets of items. For this reason, key concepts should probably be subjected to the “17 more” examinations mentioned parenthetically above.

REFERENCES

Anderson, James C. and David W. Gerbing (1984), "The Effect of Sampling Error on Convergence, Improper Solutions, and Goodness of Fit Indices for Maximum Likelihood Confirmatory Factor Analysis," *Psychometrika*, 49, 155-73.

Bagozzi, Richard P. and Todd F. Heatherton (1994), "A General Approach to Representing Multifaceted Personality Constructs: Application to Self Esteem," *Structural Equation Modeling*, 1 (1), 35-67.

Browne, Michael W. and Robert Cudeck (1993), "Alternative Ways of Assessing Model Fit," in *Testing Structural Equation Models*, K. A. Bollen et al. eds., Newbury Park CA: SAGE Publications.

Gerbing, David W. and James C. Anderson (1993), "Monte Carlo Evaluations of Goodness-of-Fit Indices for Structural Equation Models," in *Testing Structural Equation Models*, K. A. Bollen and J. S. Long, eds., Newbury Park, CA: SAGE Publications.

Jöreskog, Karl G. (1993), "Testing Structural Equation Models," in *Testing Structural Equation Models*, Kenneth A. Bollen and J. Scott Long eds., Newbury Park, CA: SAGE.

_______ (2004), "On Assuring Valid Measures for Theoretical Models Using Survey Data," *Journal of Business Research*, 57 (2), 125-41.

Ping, R. (1998), “Some Suggestions for Validating Measures Involving Unobserved Variables and Survey Data,” *1998 Winter American Marketing Association Educators= Conference*, Chicago: American Marketing Association.

_______ (2004) "On Assuring Valid Measures for Theoretical Models Using Survey Data," *Journal of Business Research*, 57 (2) (February), 125-141.

_______ (2008) “On the Maximum of About Six Indicators per Latent Variable with Real-World Data." *2008 Winter American Marketing Association Educators= Conference*, Chicago: American Marketing Association.

Saris, W. E., W. M. de Pijper and P. Zegwaart (1978), "Detection of Specification Errors in Linear Structural Equation Models," in *Sociological Methodology*, K. E. Schuesster, ed., San Francisco: Jossey-Bass.

Sörbom, D. (1975), "Detection of Correlated Errors in Longitudinal Data," *British Journal of Mathematical and Statistical Psychology*, 28, 138-51.

Appendix- Consistency Improvement using First Derivatives

A measure of the latent variable X had eight items in a Marketing survey that produced more than 200 usable responses. The first derivatives with respect to the items' measurement error terms, and their sum without regard to sign for each item, x_i, from a single construct measurement model of X are shown in Table 1 (which is not in the EXCEL template format—Table 1 also shows the diagonal and the (symmetric) upper triangle). The item with the largest Table E column Sum (x_4) was deleted, and the measurement model was re-estimated to produce Table 2. This process was repeated until RMSEA was .08 or less (see Table 4). An investigation of all other measurement models with of five items (not shown) produced combinations of items that were less consistent (i.e., they had worse model fit statistics), suggesting the Table 4 items were maximally consistent.

However, maximizing consistency does not necessarily maximize reliability or Average Variance Extracted (AVE). The items with maximum reliability and AVE were x_4, x_5, x_6, x_7, and x_8 (Reliability = .884 and AVE = .606, but χ^2 = 25, df = 5, p-value = .0001, RMSEA = .135).

There is no guidance for trading off reliability and consistency in cases where they diverge. In the present case the reliabilities of both itemizations would likely be judged acceptable. However AVE for the Table 4 itemization is only slightly above the suggested cutoff (i.e., .5), and x_4 through x_8 are "just" consistent. In cases where reliability and consistency diverge, I suggest using the itemization with the higher Face or Content Validity.

Table 1- First Derivatives for the Eight Item Measure

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0.000	-0.439	-0.025	-0.086	0.047	0.006	0.010	0.371
x_2	-0.439	0.000	-0.272	0.287	0.217	0.042	-0.200	0.143
x_3	-0.025	-0.272	0.000	-0.527	0.184	0.364	0.422	-0.207
x_4	-0.086	0.287	-0.527	0.000	-0.943	0.505	0.534	0.144
x_5	0.047	0.217	0.184	-0.943	0.000	0.222	0.359	0.019
x_6	0.006	0.042	0.364	0.505	0.222	0.000	-0.929	-0.187
x_7	0.010	-0.200	0.422	0.534	0.359	-0.929	0.000	-0.113
x_8	0.371	0.143	-0.207	0.144	0.019	-0.187	-0.113	0.000
Sum[a]	0.983	1.600	2.000	3.027	1.991	2.254	2.565	1.184

χ^2 = 86 df = 20 p-value = 0 RMSEA = .123 Reliability = .860 AVE = .442

Table 2- First Derivatives with x_4 Deleted

	x_1	x_2	x_3	x_5	x_6	x_7	x_8
x_1	0.000	-0.442	-0.064	-0.057	0.037	0.044	0.354
x_2	-0.442	0.000	-0.287	0.129	0.214	-0.067	0.195
x_3	-0.064	-0.287	0.000	-0.172	0.319	0.382	-0.313
x_5	-0.057	0.129	-0.172	0.000	0.090	0.231	-0.252
x_6	0.037	0.214	0.319	0.090	0.000	-0.544	0.012
x_7	0.044	-0.067	0.382	0.231	-0.544	0.000	0.112
x_8	0.354	0.195	-0.313	-0.252	0.012	0.112	0.000
Sum[a]	0.998	1.334	1.537	0.933	1.217	1.381	1.239

χ^2 = 56 df = 14 p-value = .44D-6 RMSEA = .117 Reliability = .828 AVE = .416

Table 3- First Derivatives with x_3 and x_4 Deleted

	x_1	x_2	x_5	x_6	x_7	x_8
x_1	0.000	-0.445	-0.086	0.045	0.054	0.304
x_2	-0.445	0.000	0.036	0.190	-0.103	0.107
x_5	-0.086	0.036	0.000	0.114	0.270	-0.383
x_6	0.045	0.190	0.114	0.000	-0.252	-0.013
x_7	0.054	-0.103	0.270	-0.252	0.000	0.096
x_8	0.304	0.107	-0.383	-0.013	0.096	0.000
Sum[a]	0.937	0.883	0.891	0.616	0.776	0.904

χ^2 = 36 df = 9 p-value = .38D-4 RMSEA = .116 Reliability = .814 AVE = .433

(Continued)

Table 4- First Derivatives with x_1, x_3 and x_4 Deleted

	x_2	x_5	x_6	x_7	x_8
x_2	0.000	-0.026	0.110	-0.180	0.079
x_5	-0.026	0.000	0.104	0.252	-0.352
x_6	0.110	0.104	0.000	-0.233	0.064
x_7	-0.180	0.252	-0.233	0.000	0.173
x_8	0.079	-0.352	0.064	0.173	0.000
Sum[a]	0.395	0.734	0.511	0.838	0.668

$\chi^2 = 5.89$ df = 5 p-value = .136 RMSEA = .028 Reliability = .835 AVE = .509

[a] Without regard to sign

ENDNOTES

[i] In my opinion, some authors go too far in real world data with single construct measurement model fit, resulting in unnecessarily small submeasures. There are several issues here, including model fit versus face or content validity, and experience suggests that with real-world data, "barely fits" in single construct measurement models is almost always sufficient to attain full measurement model fit. Thus, in real world data, subsets of items that each produce a comparatively small but nonzero Chi Square or an RMSEA that is just below .08 are usually "consistent enough" to later produce a full measurement model that fits the data. I prefer the RMSEA criterion because it seems to produce fewer problems later. Again, however, many authors would not agree with this strategy. Later, if it turns out that the full measurement model does not adequately fit the data, I simply estimate the next item weeding single construct measurement model and drop the next largest "Overall Sum" items to improve full measurement model fit.

QUESTIONS of the MOMENT...

"How should mixed interactions involving manifest, observed, continuous, single-indicator, etc. variables be estimated?"

(The **APA citation** for this paper is Ping, R.A. (2005). "How are mixed interactions involving manifest, observed, continuous, single-indicator, etc. variables estimated?" [on-line paper]. http://www.wright.edu/~robert.ping/mixed.doc)

One view of a "mixed interaction" XZ, where X is a manifest, observed, continuous, single-indicator, etc. variable and Z is a latent variable, is that XZ still involves 2 latent variables, but X has only 1 indicator. The loading of x, the indicator of X, is 1 and the measurement error of x is 0, and specification and estimation is the same as for 2 latent variables. Don't forget to zero- or mean center X and Z, and free the X-XZ and Z-XZ correlations, and you might want to use the EXCEL spreadsheets to calculate the loading, measurement error and starting values for XZ.

The only difficulty is that it is well known that X seldom has zero measurement error (see Nunnally 1993). As a result, the reliabilities of X and XZ are overstated because the reliability of x is rarely 1, and the structural coefficients in their associations with Y, for example, are typically biased in real-world data. To account for this, one approach would be to relax the assumption of zero measurement error in x. However, X with a single indicator is underdetermined, and thus an input value for the loading and the measurement error in x must be provided. There has been some confusion over the next steps because the loading and measurement error variance are not independent of each other. The relationship between them involves p_x, the assumed reliability of x, and the familiar $e_x = Var_x*(1 - p_x)$, where e_x is the measurement error variance of x, and Var_x is the error attenuated variance of x (e.g., from SAS, SPSS, etc.). In addition, a well-known estimate of p_x is the square of the loading of x on X (see for example Bollen 1989). This estimate is exact for standardized Var_x ($Var_x = 1$). Thus, the loading of x and the corresponding measurement error variance of x vary together, and they depend on the choice of the assumed reliability, which usually ranges from 0.7 to 1 (e.g., for an assumed reliability of 0.7 for x, the corresponding loading of x on X is the square root of 0.7, 0.837, and the measurement error variance is computed using $Var_x(1 - 0.7)$ or 0.3 if Var_x is standardized).

To make things simple, consider testing just the "theory-testing extremes" of reliability, $p_x = 0.7$ and $p_x = 1$, to see if the structural coefficients of the X-Y and XZ-Y associations become non significant (NS) with either of these choices. In particular, mean-center X and Z, and standardize X. Then if you haven't already done so, to estimate the X-Y and XZ-Y structural coefficients at $p_x = 1$, specify X's loading (λ_x) with the square root of $p_x = 1$ ($\lambda_x = SQRT(1) = 1$), and specify X's measurement error variance (e_x) with $Var_x*(1 - p_x) = 1*(1 - 1) = 0$ ("*" denotes multiplication) in the structural model. Next, use these $\lambda_x = 1$ and $e_x = 0$ values with the EXCEL spreadsheet values to re-compute the loading and measurement error variance of XZ, specify XZ with the resulting loading and measurement error variance of XZ, and estimate the structural model. Then, repeat this process using $p_x = 0.7$ ($\lambda_x = SQRT(0.7)$ and $e_x = Var_x*0.3$). If neither structural coefficient of X-Y or XZ-Y is NS at these "extreme" p_x's, the safest approach is probably to be conservative and interpret the smaller of the two associations (with the caveat and limitations suggested below).

However, if either of the structural coefficients of X-Y or XZ-Y become NS with these "extreme" p_x's, there are several possibilities. If a structural coefficient is NS at reliabilities of 0.7

and 1 it probably should be judged to be zero in the population. If it is NS at p_x = 0.7 and significant at p_x = 1, the conservative approach would be to judge the NS association(s) to be very likely to be zero in the population. This is because the reliability of X might actually be less than 0.7. If it is NS at 1 and significant at 0.7, the conservative approach again would be to judge the NS association(s) to be somewhat likely to be zero in the population. This is because there is some chance the reliability of X might actually be high.

Thus, if either of the structural coefficients of X-Y or XZ-Y is significant for one of the reliability "extremes" and NS for the other, that association should probably be judged to be zero in the population. However, depending on the model, this may not be a fatal blow. The lack of significance is likely due to a small standardized structural coefficient for the NS association, and thus this association would not be comparatively "important" to helping explain variance in Y. To practitioners this result could actually be as or more important than a "confirmed" association.

Unfortunately, however, there is more. There is a risk that the reliability of X is less than 0.7 (see below). This may be why theory testers prefer to avoid manifest variables if they can. The possibility that the reliability of X could be less than 0.7 should obviously be stated as a study limitation, and it should be a ***caveat*** to any interpretations or implications involving the X-Y or XZ-Y associations, even if they were significant at both "extremes" of reliability. In addition, because the reliability of X is actually unknown in the study, in the strictest sense this suggests that the X-Y and XZ-Y associations observed in the study should not be trusted. Thus, the study limitation and caveats that attend manifest variables are (or should be) very serious.

What are the options if the study is complete and X is a focal variable? In general there are several alternatives. These include ignoring X's reliability problem and hoping that reviewers will too (this is not recommended, however, because readers might notice it after publication and dismiss the study), performing a reliability study, and "argumentation." Unfortunately, the term "reliability study" has several meanings. Reliability studies for manifest variables in theoretical model tests *should* involve estimates of intra- and inter-subject rating or measurement of the manifest variables. However, I have not found anything that might be useful in theoretical model tests yet. In the meantime, consider reading the material on "Scenario Analysis" in the *Testing Latent Variable Models Using Survey Data* monograph on this web site. It may be possible to use Scenarios with students to provide multiple inter-subject or inter-subject estimates of X, or a surrogate for X, from which its reliability could be roughly estimated.

A plausible argument might be used to limit the possibilities for the amount of error in X, and provide a rough estimate of its reliability. For example, in several social science literatures the length of the relationship (LENG) is negatively associated with relationship exiting. However, LENG is usually measured in years, which obviously contains measurement error. Nevertheless, the "true" value of LENG for each respondent, informant, or subject (case) is unlikely to be more than about 10 years different from the "observed" or reported LENG in each case. Thus, one more "observation" of LENG could be computed in the data set by adding a (uniform) random number from -10 to 10 to LENG in each case to create LENG_T, an artificial "true" value for LENG. The coefficient alpha of the resulting (artificial) "measure" with the items LENG and LENG_T *might* be successfully argued to be a plausible estimate of the reliability of LENG. "Might" of course would depend on the reviewers.

As you might suspect the "reliability" depends on several attributes of the distribution of LENG, for example. The coefficient alpha of LENG and LENG_T in a real-world data set of committed relationships (mean = 13.46 years, maximum = 76) was 0.9560. In the same data set the "reliability" of the reported number of employees, EMPL (mean = 7.78, maximum=167), that could be argued to be "off" by 10, 20 or 50 employees, was 0.9612, 0.8693, and 0.5299 respectively. A "better" "reliability" estimate might involve averaging the reliabilities produced by a 100 replications of this procedure.

A slightly different approach might involve estimating a range of values such as those for EMPL, and picking the most conservative, likely, etc. However, this could be labeled "not good science," because "the argument/hypothesis" should always come first in theory testing. Because there are additional difficulties with this range approach, consider resisting it and develop a plausible argument for one "different from" number instead. I would chose 10 for EMPL because experience suggests that in the real world most people know these things and the mean for EMPL was 7.78.

REFERENCES

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.
Nunnally, Jum C. (1993), *Psychometric Theory*, 3rd Edition, New York, NY: McGraw-Hill.

QUESTIONS of the MOMENT...

"Why is my hypothesized interaction significant using a "median split" but non significant when specified in my model?"

(The **APA citation** for this paper is Ping, R. A. (2005). "Why is my hypothesized interaction significant using a "median split" but non significant when specified in my model?" [on-line paper]. http://www.wright.edu/~robert.ping/splits.doc)

One of the problems in theory testing with estimating the model using median splits of the data, or subgroup analysis, is it does not actually test the hypothesized latent variable model. The interaction is missing from the final test of the hypothesized model and the effects of its correlations with other predictors on the model's structural coefficients are not accounted for--the "missing variable" problem--see James 1980.

Nevertheless, while subgroup analysis can produce a false positive interaction result, this is unlikely (Ping 1996 observed 8% false positives with subgroup analysis, see Chapter V, "Subgroup Analysis," of the *Latent Variable Interactions and Quadratics*... monograph on this web site). It is more likely there is something amiss in the structural model or its estimation. Consider verifying that the model is properly specified (e.g., the correlations among the exogenous variables are free, the correlations between exogenous variables and endogenous variables are not free, structural disturbances are not correlated, etc.). Then, check that the interaction XZ is properly specified (e.g., the "essential" correlations between X-XZ and Z-XZ are free, the variance of XZ is also free, and the values for the loading and measurement error variance have been properly calculated and keyed into the estimation program), and the model indicators are all mean- or zero-centered. Next, verify that the structural model fits the data, all the coefficient estimates are admissible (see Step VI, "Admissible Solutions Revisited" in the *Testing Latent Variable Models Using Survey Data* monograph on the previous web page), and the measurement parameters of X and Z in the structural model (i.e., the loadings, measurement error variances and the variances of X and Z) are within a few points of their measurement model values. If the measurement parameters of X and Z in the structural model are different from their measurement model values, recalculate the interaction's loading and measurement error variance using the structural model measurement parameter values (see the EXCEL template on the previous web page for this purpose). A population interaction is also likely to be significant in OLS regression (it produced 19% false negatives in Ping 1996). However, if everything checks out in the structural model, and it plus OLS regression suggest XZ's association with Y is non significant (NS), the subgroup analysis results are probably spurious (the problems with regression are several and include that the coefficients are biased and inefficient).

If everything checks out in the structural model, and OLS regression suggests the interaction is significant, a final check would be to rerun the structural model using GLS to approximate regression's ordinary least squares estimation. Sometimes an interaction is significant using GLS/OLS but not significant using maximum likelihood. If the GLS estimation is NS, the OLS regression results might also be spurious (3% false negatives in Ping 1996).

If the GLS estimation is significant, experience with real-world data suggests the interaction is likely "borderline significant"--its t-value is very nearly 2 using maximum likelihood estimation. This is usually the result of low XZ reliability or insufficient sample size. Interaction reliability can be verified using the EXCEL spreadsheets on the previous web page. Insufficient sample size can be checked by calculating the sample size (N) that would have been required to produce a t-value of 2 using the equation $N = 4*n/t^2$, where n is the current sample size, "*" indicates multiplication, and t^2 is the square of the current t-value. If the reliability of XZ is 0.7 or above, and a few more cases would push the t-value above 2, it might suffice to simply state that the interaction "approaches significance," and proceed as though the interaction were significant. No statistical assumptions are violated by declaring that, for example, t greater than 1.95 in absolute value suggests significance. It is simply conventional in structural equation analysis to declare a structural coefficient twice or more the size of its standard error to be significant. In regression studies and correlational analysis there are two conventions for significance, p-value = 0.05 and p-value = 0.10. A p = 0.05 corresponds to t-value = 1.97 with 200 degrees of freedom (df), and p = 0.10 corresponds to t = 1.65 with 200 df. Thus, depending on reviewers, it *might* suffice to state that the interaction "approaches significance" and no more. If challenged it might be useful to respond by computing the p-value for the t-value-less-than-two and the model degrees of freedom, and comparing it to the p-value of t = 2 with the model degrees of freedom. For example, if the target structural coefficient has a t-value of 1.97 with 267 degrees of freedom, the corresponding p-value is roughly 0.0499, which is close to the p-value for t = 2 with 267 df, 0.0465. A fallback position is to preface any discussion, implication, etc. involving the target coefficient with "if significant in future studies..." This preface is actually the preferred opening remark for any discussion of the model test results because it is well known that a single study "proves" nothing. It merely suggests what future studies may observe.

However, there are other estimators and estimation approaches that may produce a t-value of 2 for the XZ-Y structural coefficient. EQS for example provides a ROBUST ML estimator that is less affected by non-normality in the data. Other estimators include pseudo maximum likelihood estimation. Other estimation approaches include bootstrapping the interaction's structural coefficient (i.e., averaging the resulting coefficients and standard errors--see "Bootstrapping" in the *Testing Latent Variable Models Using Survey Data* monograph on this web site), or removing influential case(s) (outliers that contribute most to "flattening" the XZ-Y regression line) using Cook's distance in regression, or a scatterplot of the interaction, then re-estimating the structural model. However, all of these estimators and approaches have their drawbacks, most telling of which is typically reviewer resistance to anything that is not simple, straightforward or familiar (a variation on parsimony or Occam's razor, see Charlesworth 1956). It may be easier to quickly conduct another study using Scenario Analysis (see the *Testing Latent Variable Models Using Survey Data* monograph on this web site). A significant interaction in the second study would lend weight to an insufficient sample-size argument.

There is one other possibility when a median or similar splits of the data suggests the presence of an interaction that is non significant when it is specified as XZ in a structural model. It turns out that there are infinitely more (mathematical) forms of an interaction between X and Z (or the moderation of the X-Y effect by Z). I once proposed to find two dozen interaction forms besides XZ, and wound up suggesting XZ^w, where W can be any positive or negative number. This form includes XZ (w = 1), and it includes X/Z (see Jaccard, Turrisi and Wan 1995) (as Z increases in the study X is attenuated--w = -1). It also includes XZ^2, the interaction between X and the square of Z (see Aiken and West 1991). (Parenthetically, this "form" also includes XX^w, where Z = W and X is moderated by itself, which is called a quadratic when w = 1.) Thus, a significant median split may be detecting an interaction that is not of the form XZ (i.e., specifying the significant interaction as having the form XZ is incorrect). In this case there is little more that can be done besides respecifying the interaction as the quadratic XX and testing that form of an "interaction" (i.e., X interacts with itself). Interaction specifications besides XZ and XX are unknown at present. It may be sufficient to state in the Discussion section of any paper that documents the test of the proposed model that the hypothesized interaction was not disconfirmed (the median split was significant), but the mathematical form of this interaction cannot be easily determined (if XZ and XX were both NS).

REFERENCES

Charlesworth, M. J. (1956), "Aristotle's Razor," *Philosophical Studies* (Ireland), 6: 105-112.

James, Lawrence R. (1980), "The Unmeasured Variables Problem in Path Analysis," *Journal of Applied Psychology*, 65 (4), 415-421.

Ping, R. A. (1996), "Improving the Detection of Interactions in Selling and Sales Management Research," *J. of Personal Selling and Sales Management*, 16 (Winter), 53-64.

QUESTIONS of the MOMENT...

"Why are most of my hypothesized interactions non significant?"

(The **APA citation** for this paper is Ping, R.A. (2005). "Why are most of my hypothesized interactions non significant?" [on-line paper]. http://www.wright.edu/~robert.ping/mult.doc)

Estimating multiple interactions is discussed in detail in Chapters VIII and IX of the *Latent Variable Interactions and Quadratics*... monograph on this web site. In summary, just as adding Z and W, for example, to a model with X and Y can change the significance of the X-Y structural coefficient, adding XW and ZW, for example, can attenuate (or amplify) other structural coefficients, including those for XZ-Y. This is especially true for adding XW, for example, because XZ and XW share a common constituent variable X.

However, it is usually a good idea to verify that model specification and estimation are not contributing to the "problem." Specifically, consider verifying that the model is properly specified (e.g., the correlations among the exogenous variables, including the interactions, are free; the correlations between exogenous variables and endogenous variables are not free; structural disturbances are not correlated, etc.). Then, check that the interactions are properly specified (e.g., the "essential" correlations between X-XZ and Z-XZ, for example, are free, the variance of XZ is also free, and the values for the loading and measurement error variance have been properly calculated and keyed into the estimation program), and the model indicators are all mean- or zero-centered. Next, verify that the structural model fits the data, all the coefficient estimates are admissible (see Step VI, "Admissible Solutions Revisited" in the *Testing Latent Variable Models Using Survey Data* monograph on this web site), and the measurement parameters of X and Z in the structural model (i.e., the loadings, measurement error variances and the variances of X and Z) are within a few points of their measurement model values. If the measurement parameters of X and Z in the structural model are different from their measurement model values, recalculate the interaction's loading and measurement error variance using the structural model measurement parameter values.

If the structural model and its estimates check out, the possible next steps are several. However, most would be labeled "not good science" in theory (hypothesis) testing because they amount to searching for significant interactions. For example, in applied regression studies such as epidemiology it is common to "step" variables in or out of an equation based on their significance to find the "best" model. There is an equivalent technique in structural equation analysis. However, the results obviously capitalize on chance and are thus probably inappropriate for a theory test.

An approach that *might* be defensible includes "trimming" or removing the non significant (NS) interactions. "Might" of course would depend on the reviewers. Trimming NS associations was a common practice in theory tests years ago, especially in studies that might have an intervention component (e.g., Sociology). However, based on the research behind *Testing Latent Variable Models Using Survey Data* on this web site, trimming has declined for several reasons, including that it is usually not theory-driven. Nevertheless, one could delete the interaction with the smallest NS structural coefficient t-value, then the next smallest, etc. until there are no more NS interactions. However, because this is "backward elimination" which was criticized above, an additional study is desirable to investigate the element of chance introduced by trimming. A Scenario Analysis using student subjects (see the *Testing Latent Variable Models Using Survey Data* on this web site) might provide a comparatively easily executed second study of the trimmed model using the existing questionnaire. Specifically, the hypotheses involving the NS interactions could be trimmed for the second study and the study would investigate the model without the trimmed interactions. The result could become a paper with two studies. "Multiple study" papers are common in social science disciplines such as Social Psychology and Consumer Behavior, and it might be instructive to examine a few of them to determine how best to present two-study results (see recent issues of *The J. of Consumer Research*, for example).

Although this is not helpful after the study is completed, the best approach would be to limit the number of interactions hypothesized in a theoretical model to the one or two that are theoretically most important. It is important to understand that interactions generally will explain little additional variance, and their primary purpose in theory testing is to help explain why attitudes, intentions, etc. in one subgroup in the study are or should be different from another subgroup. For example, changes in alternative attractiveness should have less effect on exiting for satisfied subjects than dissatisfied subjects (i.e., satisfaction should suppress the alternatives-exiting association).

QUESTIONS of the MOMENT...

"What is the Average Variance Extracted for a Latent Variable Interaction (or Quadratic)?"

(The **APA citation** for this paper is Ping, R.A. (2005). "What is the average variance extracted for a latent variable interaction (or quadratic)?" [on-line paper]. http://www.wright.edu/~robert.ping/ave1.doc)

Average Variance Extracted was proposed by Fornell and Larker (1981) as a measure of the shared or common variance in a Latent Variable (LV), the amount of variance that is captured by the LV in relation to the amount of variance due to its measurement error (Dillon and Goldstein 1984). In different terms, AVE is a measure of the error-free variance of a set of items.

AVE is used as measure of convergent validity. Authors in the Social Sciences disagree on what constitutes an adequate demonstration of validity. Nevertheless, a minimal demonstration of the validity of any LV should probably include the content or face validity of its indicators (how well they tap into the conceptual definition of the construct), the LV's construct validity, and its convergent and discriminant validity (e.g., Bollen, 1989; DeVellis, 1991; Nunnally, 1993). The "validity" of this LV would then be qualitatively assessed considering its reliability and its performance over this minimal set of validity criteria.

Construct validity is concerned in part with an LV's correspondence or correlation with other LV's. The other LV's in the study should be valid and reliable, then their correlations with the target LV (e.g., significance, direction and magnitude) should be theoretically sound. Convergent and discriminant validity are Campbell and Fiske's (1959) proposals involving the measurement of multiple constructs with multiple methods, and they are frequently considered to be additional facets of construct validity. Convergent measures are highly correspondent (e.g., correlated) across different methods. Discriminant measures are internally convergent. However, convergent and discriminant validity are frequently not assessed in substantive articles as Campbell and Fiske (1959) intended (i.e., using multiple traits and multiple methods). Perhaps because constructs are frequently measured with a single method (i.e., the study at hand), reliability is frequently substituted for convergent validity, and LV correlational distinctness (e.g., the target LV's correlations with other measures are less than about 0.7) is substituted for discriminant validity.

However, LV reliability is a measure of the correspondence between the items and their LV, the correlation between an LV and its items, and "correlations less than 0.7" ignores measurement error. Fornell and Larker (1981) suggested that adequately convergent LV's should have measures that contain more than 50% explained or common variance in the factor analytic sense (less than 50% error variance, also see Dillon and Goldstein 1984). Their Average Variance Extracted (AVE) for X with indicators x_1, x_2, ... , x_n is

$$AVE = \frac{\Sigma[\lambda_i^2]Var(X)}{\Sigma[\lambda_i^2]Var(X)+\Sigma[Var(\varepsilon_i)]}\ , \qquad (1$$

where λ_i is the loading of x_i on X, Var denotes variance, ε_i is the measurement error of x_i, and Σ denotes a sum (Fornell & Larker, 1981).

Unfortunately, acceptably reliable LV's can have less than 50% explained variance (AVE). Nunnally raised his suggested minimum acceptable reliability from 0.7 (Nunnally 1978) to 0.8 (Nunnally 1993) perhaps in response to this. Thus, a compelling demonstration of convergent validity would be an AVE of .5 or above.

Although there is no firm rule for discriminant validity, correlations with other LV's less than |.7| are frequently accepted as evidence of discriminant validity. A larger correlation can be tested by examining its confidence interval to see if it includes 1 (see Anderson and Gerbing, 1988). It can also be tested by using a single-degree-of-freedom test that compares two measurement models, one with the target correlation fixed at 1, and a second with this correlation free (see Bagozzi and Phillips, 1982). If the difference in resulting chi-squares is significant, this suggests the correlation is not 1, and this suggests the LV's are correlationally distinct, thus suggesting discriminant validity.

AVE can also be used to gauge discriminant validity (Fornell and Larker 1981). If the squared (error-disattenuated or structural equation model) correlation between two LV's is less than either of their individual AVE's, this suggests the LV's each have more internal (extracted) variance than variance shared between the LV's. If this is true for the target LV and all the other LV's, this suggests the discriminant validity of the target LV.

As far as I know, AVE for a LV Interaction (or a LV Quadratic) has not been derived. However, for LV's X and Z with indicators $x_1, x_2, \ldots, x_m$, and $z_1, z_2, \ldots, z_n$ with the usual assumptions (the indicators are multivariate normal with mean zero, and the measurement errors are uncorrelated and not correlated with indicators), using the Equation 1 formula for AVE and substituting the variance of the Latent Variable interaction XZ

$$\mathrm{Var}(XZ) = \mathrm{Var}(X)^{*}\mathrm{Var}(Z) + \mathrm{Cov}^2(X,Z) ,$$

where Cov denotes covariance and Var and Cov are error dissattenuated or structural equation model estimates (Kendall and Stewart 1958), the AVE of a LV Interaction is

$$\mathrm{AVE}_{XZ} = \frac{(\Sigma[\lambda_{xi}\lambda_{zj}]^2)\mathrm{Var}(XZ)}{(\Sigma[\lambda_{xi}\lambda_{zj}]^2)\mathrm{Var}(XZ)+\Theta_{XZ}} , \qquad (2$$

where $[\lambda_{xi}\lambda_{zj}]^2$ is the square of $\lambda_{xi}\lambda_{zj}$ for all x_i's and z_j's (i.e., $\Sigma[\lambda_{xi}\lambda_{zj}]^2$ is the sum of squares of the products of λ_{x1} with $\lambda_{z1}, \lambda_{z2}, \ldots, \lambda_{zn}$, λ_{x2} with $\lambda_{z1}, \lambda_{z2}, \ldots, \lambda_{zn}, \ldots$, and λ_{xm} with $\lambda_{z1}, \lambda_{z2}, \ldots, \lambda_{zn}$), and Θ_{XZ} is

$$\Theta_{XZ} = \Sigma\lambda_{xi}^2\mathrm{Var}(X)\Sigma[\mathrm{Var}(\varepsilon_{zj})] + \Sigma\lambda_{zj}^2\mathrm{Var}(Z)\Sigma[\mathrm{Var}(\varepsilon_{xi})] + (\Sigma[\mathrm{Var}(\varepsilon_{xi})])(\Sigma[\mathrm{Var}(\varepsilon_{zj})]) , \qquad (3$$

where $\Sigma\lambda_{xi}^2$ is the sum of the squares of the loadings on X, $\Sigma\lambda_{zj}^2$ is the sum of the squares of the loadings on Z, $\Sigma[\mathrm{Var}(\varepsilon_{xi})]$ is the sum of the measurement error variances of the indicators of X, and $\Sigma[\mathrm{Var}(\varepsilon_{zj})]$ is the sum of the measurement error variances of the indicators of Z. Equation 3 obtains using expectation algebra and by summing the error variances of the indicators x_iz_j

$$\begin{aligned}\mathrm{Var}(x_iz_j) &= \mathrm{Var}([\lambda_{xi}X + \varepsilon_{xi}][\lambda_{zj}Z + \varepsilon_{zj}]) \\ &= \mathrm{Var}(\lambda_{xi}\lambda_{zj}XZ + \lambda_{xi}X\varepsilon_{zj} + \lambda_{zj}Z\varepsilon_{zj} + \varepsilon_{xi}\varepsilon_{zj}) \\ &= (\lambda_{xi}\lambda_{zj})^2XZ + \lambda_{xi}^2X\mathrm{Var}(\varepsilon_{zj}) + \lambda_{zj}^2Z\mathrm{Var}(\varepsilon_{xi}) + \mathrm{Var}(\varepsilon_{xi})\mathrm{Var}(\varepsilon_{zj}) ,\end{aligned}$$

where $\lambda_{xi}^2 X Var(\varepsilon_{zj}) + \lambda_{zj}^2 Z Var(\varepsilon_{zj}) + Var(\varepsilon_{xi})Var(\varepsilon_{zj})$ is the error variance of $x_i z_j$. For emphasis, Var in Equation 2 are error dissattenuated or structural equation model estimates, and λ_{xi}, λ_{zj}, $Var(\varepsilon_{xi})$ and $Var(\varepsilon_{zj})$ in Equation 3 are unaveraged values.

Similarly, for a LV Quadratic, XX, substituting the variance of a LV quadratic

$$Var(XX) = 2Var^2(X)$$

(Kendall and Stewart 1958) into Equation 1

$$AVE_{XX} = \frac{(\Sigma[\lambda_{xi}\lambda_{xj}]^2)Var(XX)}{(\Sigma[\lambda_{xi}\lambda_{xj}]^2)Var(XZ)+\Theta_{XX}} \ , \qquad (4$$

where $[\lambda_{xi}\lambda_{xj}]^2$ is the square of the unique products of the loading of the x_i's and x_j's, $\lambda_{xi}\lambda_{xj}$, for all x_i's and x_j's (i.e., $\Sigma[\lambda_{xi}\lambda_{xj}^2]$ is the sum of squares of the products of λ_{x1} with $\lambda_{x1}, \lambda_{x2}, \ldots , \lambda_{xm}$, λ_{x2} with $\lambda_{x2}, \lambda_{x3}, \ldots , \lambda_{xm}, \ldots$, and λ_{xm} with λ_{xm}), and Θ_{XX} is

$$\Theta_{XX} = 2(\Sigma[\lambda_{xi}]^2)Var(X)\Sigma[Var(\varepsilon_{xi})] + (\Sigma[Var(\varepsilon_{xi})])^2 \ , \qquad (5$$

where $\Sigma[\lambda_{xi}]^2$ is the square of the sum of the loadings on X, and $\Sigma[Var(\varepsilon_{xi})]$ is the sum of the measurement error variances of the indicators of X. Equation 5 obtains using expectation algebra and by summing the error variances of the indicators $x_i x_j$

$$\begin{aligned} Var(x_i x_j) &= Var([\lambda_{xi}X + \varepsilon_{xi}][\lambda_{xj}X + \varepsilon_{xj}]) \\ &= Var(\lambda_{xi}\lambda_{xj}XX + \lambda_{xi}X\varepsilon_{xj} + \lambda_{xj}X\varepsilon_{xi} + \varepsilon_{xi}\varepsilon_{xj}) \\ &= (\lambda_{xi}\lambda_{xj})^2 XX + \lambda_{xi}^2 X Var(\varepsilon_{xj}) + \lambda_{xj}^2 X Var(\varepsilon_{xi}) + Var(\varepsilon_{xi})Var(\varepsilon_{xj}) \ , \end{aligned}$$

where $\lambda_{xi}^2 X Var(\varepsilon_{xj}) + \lambda_{xj}^2 X Var(\varepsilon_{xi}) + Var(\varepsilon_{xi})Var(\varepsilon_{xj})$ is the error variance of $x_i x_j$. Again for emphasis, Var in Equation 4 are error dissattenuated or structural equation model estimates, and λ_{xi} and $Var(\varepsilon_{xi})$ in Equation 5 are unaveraged values.

To compute an AVE for an LV Interaction or Quadratic, first compute the squared (unique) products of loadings from the structural model, then sum them. Then, sum the loadings and measurement errors, and sum the sum of squares of the loadings, and compute the variance of the LV Interaction or Quadratic. Next, substitute these values into Equations 3 or 5. Then, substitute the results into Equations 2 or 4.

(The EXCEL Spreadsheet "For specifying a single indicator LV Interaction or Quadratic..." on this web site calculates AVE.)

Unfortunately, experience suggests that AVE in LV Interactions and Quadratics is typically low, frequently less than 50%. For example, while they are not below 50% see the low LV Interaction and Quadratic AVE's in the EXCEL Spreadsheet "For a Single Indicator LV...," on this web site, that result from the comparatively high reliabilities of X and Z. Thus, to judge the validity of an LV Interaction or Quadratic, first it must be acceptably reliable (validity assumes reliability). Content or face validity is usually assumed unless fewer than all the indicators of the constituent variables are used to itemize the LV Interaction or Quadratic. Construct or correlational validity is usually difficult to judge, and it might be ignored. Convergent validity (AVE) should be 0.50 or above (the LV Interaction or Quadratic should be composed of 50% or less error) and it should be discriminant valid with the other model LV's, except perhaps its constituent variables (X or Z) (i.e., it is empirically distinct from the other model LV's--its AVE is larger than the squared correlations of the other LV's). In summary, while there are no hard and fast rules, reliability, and content, convergent and discriminant validity are probably sufficient to suggest the validity of an LV Interaction or Quadratic. Reliability, and content and convergent validity would be necessary, and construct (correlational) validity is usually ignored. With an AVE near 0.50 an LV Interaction or Quadratic *might* be argued to be empirically indistinct from 5-10% of the other model LV's by chance (depending on reviewers). More than that would suggest the LV Interaction or Quadratic is discriminant invalid, and its validity is impugned.

Experience suggests the substantive effect of the typically low AVE's in LV Interactions and Quadratics is their structural coefficients and their significances vary widely across replications. Specifically, with an AVE near 0.50 an hypothesized interaction or quadratic can be significant in one study but nonsignificant in a replication or near-replication. As a result, replication of a model test with hypothesized interactions or quadratics becomes comparatively more important. Specifically, an hypothesized interaction or quadratic that is NS in a model test could be significant in a replication, or vice versa. A Scenario Analysis using student subjects (see the *Testing Latent Variable Models Using Survey Data* on this web site) might provide a comparatively easily executed second study of the hypothesized interaction or quadratic using the existing questionnaire. The result could become a paper with two studies. "Multiple study" papers are common in social science disciplines such as Social Psychology and Consumer Behavior, and it might be instructive to examine a few of them to determine how best to present two-study results (see recent issues of *The J. of Consumer Research*, for example).

For an LV Interaction or Quadratic with an AVE below 0.50, the alternatives besides ignoring AVE and hoping reviewers do likewise (see "Is there any way to improve Average Variance Extracted (AVE) in a Latent Variable (LV) X?" for the logic of ignoring *slightly* unacceptable AVE) are discussed in several “QUESTIONS OF THE MOMENT” on the “Latent Variable Interaction and Quadratic Research” web page. In summary, one can drop cases for a slight improvements, drop XZ or XX indicators, or improve the AVE of XZ or XX constituent LV’s. (See “Is there any way to improve Average Variance Extracted (AVE) in a Latent Variable (LV) X” also under “QUESTIONS OF THE MOMENT” for more.) However, dropping cases is tedious and experience suggests that in real world data more than about a 3 point improvement in AVE may be unattainable. Dropping XZ or XX indicators is also tedious, and among other things the Face or Content validity of XZ or XX is questionable.

REFERENCES

Bagozzi, Richard P. and Lynn W. Phillips (1982), "Representing and Testing Organizational Theories: A Holistic Construal," *Administrative Science Quarterly*, 27 (September), 459-489.

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.

Campbell, Donald T. and Donald W. Fiske (1959), "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin*, 56, 81-105.

DeVellis, Robert F. (1991), *Scale Development: Theory and Applications*, Newbury Park, CA: SAGE Publications.

Dillon, William R. and Matthew Goldstein (1984), *Multivariate Analysis: Methods and Applications*, New York: Wiley.

Fornell, Claes and David F. Larker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.

Kendall, M.G. and A. Stuart (1958), *The Advanced Theory of Statistics*, Vol. 1, London: Charles Griffith.

Nunnally, Jum C. (1978), *Psychometric Theory*, 2nd Edition, New York, NY: McGraw-Hill.

_____ (1993), *Psychometric Theory*, 3rd Edition, New York, NY: McGraw-Hill.

QUESTIONS of the MOMENT...

"What is the "validity" of a Latent Variable Interaction (or Quadratic)?"

(The **APA citation** for this paper is Ping, R.A. (2005). "What is the "validity" of a latent variable interaction (or quadratic)?" [on-line paper]. http://www.wright.edu/~robert.ping/validity.doc)

Anecdotally, not all authors agree that reliability and validity apply to latent variable (LV) interactions or quadratics. Nevertheless, because adding an LV with unknown reliability and validity, and thus unknown levels of measurement error, to a theoretical model in order to test hypotheses seems unwise. The resulting structural coefficients could be an artifact of measurement error.

Authors in the Social Sciences disagree on what constitutes an adequate demonstration of validity. Nevertheless, a minimal demonstration of the validity of any LV should probably include the content or face validity of its indicators (how well they tap into the conceptual definition of the second-order construct), the LV's construct validity, and its convergent and discriminant validity (e.g., Bollen, 1989; DeVellis, 1991; Nunnally, 1993). The "validity" of this LV would then be qualitatively assessed considering its reliability and its performance over this minimal set of validity criteria.

Construct validity is concerned in part with an LV's correspondence or correlation with other LV's. The other LV's in the study should be valid and reliable, then their correlations with the target LV (e.g., significance, direction and magnitude) should be theoretically sound. Convergent and discriminant validity are Campbell and Fiske's (1959) proposals involving the measurement of multiple constructs with multiple methods, and they are frequently considered to be additional facets of construct validity. Convergent measures are highly correspondent (e.g., correlated) across different methods. Discriminant measures are internally convergent. However, convergent and discriminant validity are frequently not assessed in substantive articles as Campbell and Fiske (1959) intended (i.e., using multiple traits and multiple methods). Perhaps because constructs are frequently measured with a single method (i.e., the study at hand), reliability is frequently substituted for convergent validity, and LV correlational distinctness (e.g., the target LV's correlations with other measures are less than about 0.7) is substituted for discriminant validity.

However, LV reliability is a measure of the correspondence between the items and their LV, the correlation between an LV and its items, and "correlations less than 0.7" ignores measurement error. Fornell and Larker (1981) suggested that adequately convergent LV's should have measures that contain more than 50% explained or common variance in the factor analytic sense (less than 50% error variance, also see Dillon and Goldstein 1984), and they proposed a statistic they termed Average Variance Extracted (AVE) as measure of convergent validity. AVE is a measure of the shared or common variance in an LV, the amount of variance that is captured by the LV in relation to the amount of variance due to its measurement error (Dillon and Goldstein 1984). In different terms, AVE is a measure of the error-free variance of a set of items (AVE and its computation are discussed in detail elsewhere on this web site).

AVE can also be used to gauge discriminant validity (Fornell and Larker 1981). If the squared (error-disattenuated or structural equation model) correlation between two LV's is less than either of their individual AVE's, this suggests the LV's each have more internal (extracted) variance than variance shared between the LV's. If this is true for the target LV and all the other LV's, this suggests the discriminant validity of the target LV.

Unfortunately, experience suggests that AVE in LV Interactions and Quadratics is typically low, frequently less than 50%. For example, while they are not below 50% see the low LV Interaction and Quadratic AVE's in the EXCEL Spreadsheet "For a Single Indicator LV...," on this web site, that result from the comparatively high reliabilities of X and Z. Thus, to judge the validity of an LV Interaction or Quadratic, first it must be acceptably reliable (validity assumes reliability). Content or face validity is usually assumed unless fewer than all the indicators of the constituent variables are used to itemize the LV Interaction or Quadratic. Construct or correlational validity is usually difficult to judge, and it might be ignored. Convergent validity (AVE) should be 0.50 or above (the LV Interaction or Quadratic should be composed of 50% or less error) and it should be discriminant valid with the other model LV's, except perhaps its constituent variables (X or Z) (i.e., it is empirically distinct from the other model LV's--its AVE is larger than the squared correlations of the other LV's). In summary, while there are no hard and fast rules, reliability, and content, convergent and discriminant validity are probably sufficient to suggest the validity of an LV Interaction or Quadratic. Reliability, and content and convergent validity would be necessary, and construct (correlational) validity is usually ignored. With an AVE near 0.50 an LV Interaction or Quadratic *might* be argued to be empirically indistinct from 5-10% of the other model LV's by chance (depending on reviewers). More than that would suggest the LV Interaction or Quadratic is discriminant invalid, and its validity is impugned.

Experience suggests the substantive effect of the typically low AVE's in LV Interactions and Quadratics is their structural coefficients and their significances vary widely across replications. Specifically, with an AVE near 0.50 an hypothesized interaction or quadratic can be significant in one study but nonsignificant in a replication or near-replication. As a result, replication of a model test with hypothesized interactions or quadratics becomes comparatively more important. Specifically, an hypothesized interaction or quadratic that is NS in a model test could be significant in a replication, or vice versa. A Scenario Analysis using student subjects (see the *Testing Latent Variable Models Using Survey Data* on this web site) might provide a comparatively easily executed second study of the hypothesized interaction or quadratic using the existing questionnaire. The result could become a paper with two studies. "Multiple study" papers are common in social science disciplines such as Social Psychology and Consumer Behavior, and it might be instructive to examine a few of them to determine how best to present two-study results (see recent issues of *The J. of Consumer Research*, for example).

For an LV Interaction or Quadratic with an AVE below 0.50, the alternatives besides ignoring AVE and hoping reviewers do likewise are to improve AVE in the LV Interaction or Quadratic. Low AVE in XZ is caused by low correlation between X and Z and/or comparatively large measurement errors in the items of X and or Z (i.e., low X and/or Z reliability). I have a few strategies besides rerunning the study with a measurement study to improve the reliability of X and Z that would take to long to explain here. If you need to improve AVE *a few points* you might consider e-mailing me for details of these strategies--more than about a 3 point improvement in AVE, however, may be unattainable without rerunning the study.

REFERENCES

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.

Campbell, Donald T. and Donald W. Fiske (1959), "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin*, 56, 81-105.

DeVellis, Robert F. (1991), *Scale Development: Theory and Applications*, Newbury Park, CA: SAGE Publications.
Dillon, William R. and Matthew Goldstein (1984), *Multivariate Analysis: Methods and Applications*, New York: Wiley.
Fornell, Claes and David F. Larker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.
Nunnally, Jum C. (1993), *Psychometric Theory*, 3rd Edition, New York, NY: McGraw-Hill.

QUESTIONS of the MOMENT...

"How does one remedy a 'not Positive Definite' message?"

(The **APA citation** for this paper is Ping, R.A. (2012). "How does one remedy a 'not Positive Definite' message?" [on-line paper]. http:// www.wright.edu/~robert.ping/ NotPD1.doc .)

(An earlier version of this paper, Ping, R.A. (2009). "How does one remedy a 'not Positive Definite' message?" [on-line paper]. http:// www.wright.edu/~robert.ping/NotPD.doc . is available here.)

Few things are as frustrating, after gathering and entering survey data for a new model, and creating and debugging the estimation software program for the model, as the first bug-free software run producing a "not Positive Definite" message ("Ill Conditioned" in exploratory factor analysis). The definition of this problem provides little help, and there is little guidance for remedying matters, besides "check the data and the data correlation matrix," "delete items," or "use Ridge Option estimates" (the last of which produces biased parameter estimates, standard errors, and fit indices).

Besides data entry errors,[1] [2] experience suggests that in real-world survey data, causes for a Not Positive Definite (NPD) message usually are 1) collinearity among the items, 2) measure(s) inconsistency, 3) inadequate starting values, and 4) model misspecification.

(1) Collinearity among the items is easiest to investigate. Specifically, SAS, SPSS, etc. item correlations of 0.9 or above should be investigated by removing the item(s) involved to see if it remedies NPD. In this case, dropping or combining the highly correlated items may remove the NPD message. [3]

Occasionally in real world data, NPD can be the result of two or more parallel measures (measures of the same construct). With NPD and parallel measures, the measure(s) with the lesser psychometrics (reliability/validity) probably should be abandoned, and NPD should be reassessed.

(2) Measure Inconsistency: In real world data NPD can accompany lack of consistency in the Anderson and Gerbing (1988) sense (lack of model-to-data fit). Unfortunately, the procedure for investigating this possibility is tedious.

---

[1] Experience suggests that random data entry errors seldom produce a "not Positive Definite" message. However, because they may create other problems later, it always prudent to examine the data for data entry errors.

[2] There are other plausible conditions--see for example http://www2.gsu.edu/~mkteer/npdmatri.html.

[3] Deleting an item should be done with concern for the content or face validity of the resulting measure. Combining items may be less desirable than removing them because it could be argued that the resulting combined "item" is no longer an observed variable.

The process begins with maximum likelihood exploratory (common) factor analysis (FA) of each measure. Specifically, each measure should be FA'd. Then, pairs of measures should be FA'd, then triplets, etc. (Note that one or more measures may be multidimensional—experience suggests that usually will not produce NPD.)

Typically, NPD occurs when adding a measure to a group of m (m < n, where n is the number of measures) measures that was Positive Definite (PD). When a measure is found to create NPD, its items should be "weeded" using reliability maximization.[4] In particular, after an item is dropped, NPD should be evaluated in a FA with all n of the measures. After that, deleted items should be added back to measures, beginning with the item contributing most to content validity of the most "important" measure, then proceeding to the next most "important" item, etc. With each added item, NPD should be checked using FA and all n measures.

Occasionally, the above approach does not remedy NPD (it does not produce a set of n measures that is PD in a FA with all n measures) without excessive item weeding (too many items are weeded out), or without weeding an item(s) that are judged to be essential for face or content validity. In this case, experience suggests that weeding using Modification Indices instead of reliability might remedy NPD. (See the EXCEL template "For 'weeding' a multi-item measure so it 'fits the data'..." on the preceding web page).

Specifically, the measure to be weeded should be estimated in a single LV measurement model (MM) and:

a) it should be estimated using covariances and Maximum Likelihood estimation.

b) The single LV MM should have one (unstandardized) loading constrained to equal 1, and all the other (unstandardized) loadings should be free (unconstrained) and their estimates should be between 0 and 1.[5]

c) In the single LV MM, the unstandardized LV variance should be free, and the measurement model LV variance estimate should be positive and larger than its error-attenuated (i.e., SAS, SPSS, etc.) estimate.

d) All measurement error variances should be free and uncorrelated,[6] and their estimates each should be zero or positive.

---

[4] SAS, SPSS, etc. have procedures that assess the reliability of the remaining items if an item is dropped from the group.

[5] If one or more loadings in a measure is greater than one, the largest loading should be fixed at 1 and the other measure loadings should be freed, including the loading previously fixed at 1.

[6] Uncorrelated measurement errors is a classical factor analysis assumption. If this assumption is violated (e.g., to obtain model to data fit with an item that must not be deleted), experience suggests that the above procedure may still work. If it does not, the interested reader is encouraged to e-mail me for suggestions for their specific situation.

e) The measurement model should fit the data very well using a sensitive fit index such as RMSEA (i.e., RMSEA should be at least 0.08--see Brown and Cudeck 1993, Jöreskog 1993).

In the unusual case that MI weeding does not remedy NPD, experience suggests that during MI weeding two items in a measure may have had practically the same MI. In this case, the removal of either item will improve model to data fit, and, because the deletion of the first item did not remove NPD, the second item should be deleted instead (rarely, both items should be deleted). Again, deleted items should be added back selectively to their respective measures until NPD reoccurs.

(The case where NPD remains unremedied is discussed below.)

For emphasis, the objective should be to find the item(s) responsible for NPD. Stated differently, each measure should retain as many items as possible.

(3) Inadequate Starting Values: It is possible in real-world survey data that is actually Positive Definite (PD), to obtain a fitted or reproduced covariance matrix that is NPD. Experience suggests this usually is the result of structural model misspecification, which is discussed below, but it also can be the result of inadequate (parameter) starting values.

Inadequate starting values usually occur in the structural model, and while they can be produced by the software (LISREL, EQS, AMOS, etc.), more often they are user supplied. Fortunately, adequate starting values for LV variances and covariances can be obtained from SAS, SPSS, etc. using averaged items. Starting estimates for loadings can be obtained from maximum likelihood exploratory (common) factor analysis,[7] and regression estimates (with averaged indicators for each measure) can be used for adequate structural coefficient starting values. For emphasis, the parameters with these starting values should always be unfixed (i.e., free) (except for each measure's item with a loading of 1).

(4) Model Misspecification: It is also possible with PD input to obtain a fitted, or reproduced, structural model covariance matrix[8] that is NPD because of structural model misspecification. Remedying this is very tedious.

The procedure uses three steps beginning with verifying that the structural model paths reflect the hypotheses exactly. If they do, check that the full measurement model (FMM) is PD. Next, verify that the structural model is specified exactly as the FMM, except that

---

[7] Each measure should be factored individually, and in each individual factor analysis the resulting standardized loadings should be divided by the largest loading in that factor analysis for SEM starting values.

[8] NPD parameter matrices (e.g., PSI or THETA-EPS) also can occur, usually with under identified LV's having fixed parameters or correlated measurement errors. The interested reader is encouraged to e-mail me for suggestions for their specific situation.

the hypothesized structural paths have replaced several MM correlations. Then, remove any structural model misspecification.

Specifically, estimate a full measurement model with all the measures present using steps a-e above with “single,” “each,” etc. replaced with “full.”

If the FMM is NPD and remedies 1-3 above have been tried, an approach that uses Anderson and Gerbing's 1988 suggestions for establishing internal and external consistency should be used. It begins with estimating each LV in its own measurement model (with no other measures present), then estimating pairs of LV's in two-measure measurement models (with only the two measures present). Next, triplets of LV's are estimated in three-measure measurement models, then quadruplets of LV's and so on until the full measurement model is estimated.

Specifically, each of the single LV measurement models should be estimated as described in steps a) through e) above to establish "baseline" parameter estimates for each measure for later use. (If these parameter estimates are already available, this step can be skipped.)

Then, the LV's should be estimated in pairs--for example with 4 LV's, 6 measurement models, each containing two LV's are estimated. These measurement models also should be estimated using steps a) through e) above.

In addition,

f) Each LV should be specified exactly as it was in its own single LV measurement model, no indicator of any LV should load onto a different LV, and no measurement error variance of one LV should be allowed to correlate with any measurement error variance of any other LV.

g) The covariances between the LV's should be free, and in real world data the resulting estimated variances and covariances of each LV, their loadings and measurement error variances should be nearly identical to those from the LV's own single LV measurement model.

h) Each of these measurement models should fit the data very well using a sensitive fit index such as RMSEA.

Next, the LV's should be estimated in triplets--for 4 LV's this will produce 3 measurement models with 3 LV's in each. Each of these measurement models should be estimated using steps a) through h) above.

Then, this process of estimating larger and larger combinations of LV's using steps a) through h) should be repeated until the full measurement model with all the LV's present would be estimated using steps a) through h).

At this point at lease one measure should have been found to be problematic. Then, please contact me for the next steps.

If the FMM is PD and possibilities 1-3 above have been checked, reverify that the structural model reflects the hypotheses exactly, and that the structural model is specified exactly as the FMM, except that the hypothesized structural paths have replaced several MM correlations. Then, try dropping an LV from the structural model. If the structural model is still NPD, try dropping a different LV. If repeating this process remedies NPD, please email me for the next steps. If repeating this process, dropping each measure one-at-a-time does not remedy NPD, please email me for different next steps.

REFERENCES

Anderson, James C. and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 103 (May), 411-23.

Browne, Michael W. and Robert Cudeck (1993), "Alternative Ways of Assessing Model Fit," in *Testing Structural Equation Models*, K. A. Bollen et al. eds., Newbury Park CA: SAGE Publications.

Jöreskog, Karl G. (1993), "Testing Structural Equation Models," in *Testing Structural Equation Models*, Kenneth A. Bollen and J. Scott Long eds., Newbury Park, CA: SAGE.

QUESTIONS of the MOMENT...

"Why are reviewers complaining about the use of PLS in my paper?"

(The **APA citation** for this paper is Ping, R.A. (2009). "Why are reviewers complaining about the use of PLS in my paper?" [on-line paper]. http://www.wright.edu/~robert.ping/PLS.doc)

Theory-test papers propose theory that implies a path model. Then, they report a first, hopefully adequate, disconfirmation test[1] of the model (and by implication the theory) that involves a data gathering protocol and a model estimation protocol. Reviewers usually have little difficulty evaluating the proposed theory and the data gathering protocol, but they may have difficulty evaluating the adequacy of a test that relies on a model estimation protocol involving PLS. PLS is not widely used in the social sciences, and some reviewers may be unfamiliar with PLS. These reviewers may reject the paper because they are unable to judge the adequacy of PLS as estimation software for the theory test (see Footnote 1). For the same reason, other reviewers may want to see SEM results, and absent those results, they also may reject the paper.

Reviewers who are familiar with PLS may judge PLS to be inadequate for theory testing. Anecdotally, some object to its use of least squares estimation that maximizes variance explained rather than model-to-data fit of the covariances (as in SEM). Others may object to PLS's reliance on bootstrap standard errors (SE), and that the newer PLS software implementations appear to produce inconsistent estimates.

BACKGROUND

PLS was proposed about the same time as LISREL (see Wold 1975 for PLS, and Jöreskog 1973 for LISREL). However, the differences between PLS and LISREL are considerable. For example, PLS assumes formative[2] latent variables (LV's), instead of

---

[1] The logic of science dictates that an adequate test should falsify the proposed theory--it should show that it is false. If the test fails to falsify the theory, the test may be inadequate. Only after the test is (independently) judged to be adequate despite its failure to disconfirm, should the test results be viewed as suggesting "confirmation" (i.e., confirmation in this one case--confirmation of the theory is an inductive process requiring many disconfirmation tests that fail to disconfirm, and thus building confidence in the theory.)

[2] Blalock (1964) proposed that an LV can be formative or reflective. Reflective items are affected by (diagrammatically "pointed to" by) the same underlying concept or construct (i.e., the reflective LV). LISREL, EQS, AMOS, etc. assume reflective LV's.

Formative indicators are measures that affect an LV. Diagrammatically, formative indicators point to the LV. A classic example of a formative LV is socio-economic status (SES), which is defined by items such as occupational prestige, income and education. That the indicators "cause" or point to SES, rather than vice versa, is suggested by the reflective LV's as in SEM (e.g., LISREL, EQS, AMOS, etc.). PLS factors are estimated as linear combinations (composites) of their indicators, a form of principal component analysis. In addition, PLS maximizes the ability of factors (X's) to explain variance in responses (Y's).

PLS's positives include that it estimates nominal variables, and it estimates collinear LV's without resorting to Ridge estimation. Its maximization of explained variance improves forecasting, and, as a result, PLS has a large following outside of theory testing. In addition, PLS can estimate reflective LV's. As a result, mixed models with reflective and formative LV's are possible.[3]

PLS's negatives include that, as previously mentioned, it is not widely seen in theory testing articles within the social sciences. Anecdotally, it is unknown to some theory testers. Its path coefficient estimates are not maximum likelihood (ML), which is preferred in theory testing. PLS's path coefficients also are not covariances, and thus they may be difficult to interpret. Also, as previously mentioned, PLS assumes formative LV's, the need for which may not be well understood in theory testing.

Anecdotally, some reviewers view PLS as a way to avoid dealing with (reflective) measures that have poor psychometric properties (e.g., are unreliable, have low Average Variance Extracted, are discriminant invalid, etc.). In addition, PLS's ability to specify reflective LV's with weights that are proportional to their measurement model loadings may be a minus in theory tests. Since real world models also are likely to have reflective LV's, substantive researchers who want to estimate mixed models with formative and reflective LV's, may have to learn both PLS and SEM software (however, see "How are Formative Latent Variables estimated with LISREL, EQS, AMOS, etc.?" on this web site.).

PLS's negatives also include issues that appear to be less widely known or appreciated outside of statistical circles, such as its reliance on bootstrap (resampling based) Standard Errors (SE's). These statistics are biased without correction. (Efron, who popularized bootstrapping, apparently spent many years trying to resolve this problem--see Efron and Tibshirani 1993, 1997. In an informal review of popular PLS software documentation I could find no indication of bootstrap estimates that were corrected for bias and inconsistency.) Finally, software implementations of Wold's proposals appear to produce inconsistent estimates (e.g., Temme, Kreis and Hildebrandt 2006).

---

likelihood that increased occupational prestige would increase SES, rather than increased SES necessarily would increase occupational prestige. (That being said, judging formative and reflective LV's, including SES, can become messy--see "How are Formative Latent Variables estimated with LISREL, EQS, AMOS, etc.?" on this web site.)

[3] PLS factors with indicator weights that are proportional to their SEM loadings should produce factors that are similar to their SEM counterparts (e.g., Schneeweiss 1993). However, I have yet to produce such results.

In addition, most of PLS's strengths--nominal, formative, and collinear LV's; handling LV's with poor psychometrics, and forecasting--are all plausibly "covered" by SEM. For example, (truly) categorical (nominal) variables can be estimated in SEM (see "How does one estimate categorical variables..." on this web site).

Formative LV's and LV's with poor psychometrics also can be estimated in SEM (see "How are Formative Latent Variables estimated with LISREL...?" on this web site). While PLS may have an advantage in estimating collinear LV's--its SE's for collinear LV's may be less biased than SEM's Ridge estimates--collinear LV's are usually not discriminant valid in real-world theory tests, so they seldom appear in real world survey data tests (see "What is the "validity" of a Latent Variable Interaction (or Quadratic)?" on this web site).

PLS's forecasting capability may be neither a plus nor a minus in theory testing. Prediction-versus-explanation is a contentious area in the philosophy of science. Some authors argue that explanation is a better test of theory than prediction (e.g., Brush 1989), while others argue the reverse (e.g., Maher 1988). Nevertheless, it would be interesting to compare the consistency of a model's interpretations across multiple samples between SEM (i.e., explanation) and PLS (i.e., prediction).

That being said, SEM eventually may have no advantage over PLS in theory testing. SEM's interpretations may be no more consistent across samples than PLS's. And, PLS's unfamiliarity to reviewers, and its unadjusted SE's and inconsistent software should be remedied over time.

However, at present, a substantive paper that relies solely on PLS may be difficult to publish in the social sciences. It is likely that many reviewers will reject PLS because they are unfamiliar with it. A few reviewers may reject PLS because they disagree with its assumptions. Still fewer reviewers may reject PLS because of its software implementation's apparent "inadequacies."

While strong arguments for PLS might be provided in a paper, it may be necessary to report PLS and SEM results.[4] Specifically, if the model contains nominal LV's, the SEM results could be compared to those of PLS. If LV collinearity is a problem, Ridge and PLS estimates could be compared.[5] Finally, formative LV's and LV's with poor

[4] PLS results may or may not approximate SEM results (see McDonald 1996). However, it is plausible that generally consistent interpretations for PLS versus those of SEM across a holdout sample might support the efficacy of one estimation technique over another in the study at hand.

[5] However, Ridge SE's are believed to be biased.

psychometric properties[6] could be compared between SEM and PLS on their performance versus the hypotheses.

REFERENCES

Blalock, H.M. (1964) *Causal Inferences in Nonexperimental Research*, Chapel Hill, NC: University of North Carolina Press.

Brush, S.G. (1989), "Prediction and Theory Evaluation: The Case of Light Bending," *Science*, New Series (246, 4937) (Dec), 1124-1129.

Efron, B. and Tibshirani, R.J. (1993), *An Introduction to the Bootstrap*, New York: Chapman and Hall.

Efron, B. and Tibshirani R.J. (1997), "Improvements on Cross-Validation: The e.632+ Bootstrap Method," *Journal of American Statistical Association* 92, 548-560.

Jöreskog, K. (1973), "A General Method for Estimating a Linear Structural Equation System," in A.S. Goldberger and O.D. Duncan eds., *Structural Equation Models in the Social Sciences* (85-112), NY: Seminar.

Maher, P. (1988): "Prediction, Accommodation and the Logic of Discovery," *PSA* (1), 273-285.

McDonald, R. P. (1996), "Path Analysis with Composite Variables," *Multivariate Behavioral Research* (31), 239-270.

Schneeweiss, H. (1993), "Consistency at Large in Models with Latent Variables," in K. Haagen, D. J. Bartholomew and M. Deistler eds., *Statistical Modelling and Latent Variables*. Amsterdam: Elsevier, 288-320.

Temme, D., K. Henning and L. Hildebrandt (2006), "PLS Path Modeling – A Software Review," [on-line paper], http://ideas.repec.org/p/hum/wpaper/sfb649dp2006-084.html#provider. (Last accessed Nov 30, 2009.) (Paper provided by Sonderforschungsbereich 649, Humboldt University, Berlin, Germany in its series SFB 649 Discussion Papers with number SFB649DP2006-084.)

---

[6] A formative specification might enable estimation of older "well established" (i.e., before SEM) measures that require extensive weeding when they are used in SEM. LV's with poor psychometrics (e.g., LV's with low reliability or Average Variance Extracted, discriminant invalidity, low model-to-data fit, etc.) may include second order LV's (see "Second-Order Latent Variable Interactions... " and "How are Formative Latent Variables estimated with LISREL...?" on this web site).

Wold, H. (1975), "Path Models with Latent Variables: The NIPALS Approach," in *Quantitative Sociology: International Perspectives on Mathematical and Statistical Modeling*, H. M. Blalock, A. Aganbegian, F. M. Borodkin, R. Boudon, and V. Cappecchi eds., Academic Press, New York, 307-357.

QUESTIONS of the MOMENT...

"How are Formative Latent Variables estimated with LISREL, EQS, AMOS, etc.?"

(The **APA citation** for this paper is Ping, R.A. (2010). "How are Formative Latent Variables estimated with LISREL, EQS, AMOS, etc.?" [on-line paper]. http://www.wright.edu/~ robert.ping/Forma.doc)

(An earlier version of this paper, Ping, R.A. (2009). "How are Formative Latent Variables estimated with LISREL, EQS, AMOS, etc.?" [on-line paper]. http://www.wright.edu/~robert.ping/Form.doc, is available here.)

In theory testing, internal inconsistency (the single construct measurement model does not fit the data) usually occurs in a new measure that is developed using many items. In fact, authors have noted that about six items seems to be the maximum number of items in an internally consistent measure (Anderson and Gerbing 1984, Bagozzi and Heatherton 1994, Gerbing and Anderson 1993). Parenthetically, this may explain the disappearance of older well-established measures, developed before structural equation analysis became popular for theory testing, and that typically contained more than six items, (e.g., Comer, Machleit and Lagace 1989).

Anecdotally, inconsistency is remedied by deleting items, trading off face validity for consistency, until adequate (or trivial—three items) consistency is attained. Occasionally, authors have summed a (presumably inconsistent) measure (e.g., Williams and Hazer 1986) to maintain its face validity. The resulting single-indicator latent variable is specified with a loading of unity and a measurement error of zero. In different words, the items are assumed to define the latent variable. This was the idea behind Blalock's (1964) proposal of "formative" latent variables.

LATENT VARIABLES

Blalock (1964) noted that unobserved or latent variables can produce changes in observed variables, or vice versa. A "reflective" latent variable is specified using observed variables (indicators) that are affected primarily by (diagrammatically "pointed to" by) one underlying concept (i.e., its latent variable). Reflective latent variables are ubiquitous in theory tests and structural equation analysis.

However, "formative" indicators produce changes in their latent variable. Diagrammatically, formative indicators "point to" their latent variable. Formative latent variables are rare in theory tests in the social sciences, when compared to reflective latent variables. Although they are not new (e.g., Blalock 1964), formative latent variables may not be well understood by substantive researchers.

A classic formative latent variable is Socio-Economic Status (SES), that that has the indicators Occupational Prestige, Income and Education. The "direction" of these indicators is suggested by it being more likely that increased Occupational Prestige will increase SES, than increased SES will increase Occupational Prestige (i.e., Occupational Prestige "causes" or points to SES).

However when SES is examined more closely, things become less clear. An increase in SES plausibly could increase income. Thus, the relationship between SES and income could be bi-directional (i.e., with arrows between them in both directions) instead of just from income to SES.

The definition of a formative latent variable involves the notion that the indicators “create” or adequately define the latent variable (Blalock 1964). The notion of an “adequate” operational definition is important, because SES, for example, could be judged to have additional indicators—Physical Attractiveness for example.

Authors also have noted that formative latent variables have indicators that may not be highly correlated (e.g., Bollen and Lennox 1991). Nevertheless, the indicators of SES are obviously correlated. Increased education is well known to be correlated with increased income. Increased occupational prestige is associated with increased income. And, increased education is associated with increased occupational prestige.

In summary, SES appears to have attributes of a formative and a reflective latent variable. Experience suggests that such indicator ambiguity in formative latent variables may not be uncommon. For example, Ping (2007) specified a latent variable, Goal Congruity, reflectively, when it had previously been specified as formative (see Anderson 1988). Stated differently, formative latent variables may be more likely than their reported specification would suggest (see Cohen, Cohen, Teresi, Marchi and Velez 1990).

Returning to weeding versus face validity, Ping (2007) reported a well-established measure, Organizational Commitment (OC), that had low Average Variance Extracted (AVE), and, as a result, it exhibited unacceptable discriminant validity using Fornell and Larker’s 1981 “squared correlation versus AVE” criterion. As a result, this measure was "quarantined" in the paper.[1] However, as with SES, the measure might have been specified formatively—OC's indicators might have been argued to cause OC, and to adequately define it; and OC then might have been a candidate for the structural equation analysis model.

As another example, Ping (1997) specified the Cost of Exiting a socio-economic exchange relationship (COE) as a reflective second-order latent variable. It had the first-order (indicator) latent variables Unattractive Alternative relationships (ALTU), relationship Investment (INV) and Switching Costs (SC) (presumably because as COE changed one or more of ALTU, INV or SC were likely to change). However, as with SES, if Unattractive Alternatives, Investments or Switching Costs increased, the Cost of Exiting also should increase. Thus, Cost of Exiting might have been specified as a

[1] Because OC was a "well established" measure, a model containing OC was estimated. However, OC contained more than 50% error variance and it was discriminant invalid, so the OC model results received limited discussion.

formative latent variable instead of reflective latent variable. Specifically, ALTU, INV and SC might plausibly have been argued to adequately define COE.

Nevertheless, COE was not estimated formatively because it was psychometrically adequate when specified reflectively (i.e., it exhibited adequate reliability, validity and consistency).

However, experience suggests this is not always the case with second-order latent variables. They can be “cobbled together” post hoc in a study to simplify a large model by combining related first-order latent variables (e.g., Dwyer and Oh 1987; Gerbing, Anderson Freeman 1994). As a result, second-order latent variables can be minimally-determined, and they can exhibit inadequate reliability or validity. In addition, they can exhibit inadequate external consistency: Their full measurement model (containing all the latent variables) can exhibit low model-to-data fit because their first-order "indicator" latent variables are correlated with other latent variables in the model (the first-order "indicator" latent variables of the second-order latent variable should have only one underlying construct, the second-order latent variable). Thus, a formative second-order specification may be an alternative to "breaking up" a reflective second-order latent variable into its component ("indicator") first-order latent variables because of inadequate model-to-data fit.

FORMATIVES

As the reader may have guessed, this discussion aims at the admittedly difficult task of arguing for a formative (re)specification of a “difficult” measure (e.g., one that is in danger of being weeded down to face invalidity in order to attain adequate consistency, or one that is in danger of being dropped from the model because of inadequate psychometrics (e.g., inadequate reliability, AVE, etc.).

However, one might object to using a formative specification for a measure that was "created" (itemized) as a reflective measure. In addition, some reflective measures may not be able to be re-specified formatively. (Specifically, it may be difficult to argue that a measure's items adequately define the target latent variable. Or, some items may not cause their construct.) Finally, a formative specification for a mixed reflective-formative structural equation analysis model is currently unknown.

We shall address each of these. The first matter, considering a formative re-specification for a reflective measure, would require a compelling reason for "saving" a "difficult" construct, or “rescuing” items, by resorting to formative re-specification. Such reasons might include that the “difficult” latent variable is a focal construct, or it should explain nontrivial variance in an endogenous (dependent) variable(s) (i.e., its omission would contribute to structural coefficient bias caused by the missing variable problem—see James 1980). Or, the items to be rescued are important items, and the operational definition of the latent variable would be impaired if they are dropped (i.e., without them the remaining items would no longer adequately tap the conceptual definition).

It also might be argued that such a change in itemization is not good science: Items that were "created" as reflective can not be re-specified formatively. Nevertheless, this issue may be more apparent than real. A reflective item pool should adequately itemize its construct (i.e., the item pool should contain items measuring the all the salient observed aspects of the target construct). Experience suggests that some of these items also can be viewed as formative, and these "bidirectional" items might be judged to adequately define their construct (suggesting that their construct might be re-specified formatively). (For example, in Cost of Exit (COE) above, the “indicators” Unattractive Alternative relationships (ALTU), relationship Investment (INV) and Switching Costs (SC) operationally could be reflective or formative).

Philosophically, that "reflective items" should not be re-specified formatively is beyond the scope of this discussion. It is tempting to point out, however, that some female reef fish change their gender to males when there are too few males present (e.g., Ross 1990). Thus, the end (in the present case model disconfirmation) may justify the means (re-specification). In addition, it may be instructive to note that method should accommodate theory, and not vice versa. It could be argued that discarding an important construct, or dropping items that are important to a measure’s face validity, in order to accommodate an estimation technique (I.e., structural equation analysis), is theory forced to accommodate method.

The second matter, that a measure's items should adequately define its latent variable, obviously is a matter of judgment. SES, for example, routinely is judged to be defined (adequately itemized) by its indicators Occupational Prestige, Income and Education. However, SES might be argued to be inadequately defined (itemized) because Physical Attraction was not included in its itemization. However, if an indicator in question is important to the latent variable, but it does not “cause” its latent variable, the measure may not be able to be re-specified formatively.

Items that clearly do not “cause” their construct should be dropped before the remaining items that define their latent variable is argued to be adequate. However, experience suggests that in real-world data, the paths between a latent variable and its indicators are seldom clearly "one way," (e.g., from the indicator to the latent variable, or vice versa). Stated differently, experience suggests that comparatively few indicators that were “created” reflectively are clearly reflective in real-world survey data.

The third matter, specifying a mixed reflective-formative structural equation analysis, may be a matter of perspective. PLS, a software package for specifying formative latent variables, creates “composites”—linear combinations—of indicators, which could be argued to be a form of principal component analysis. A linear combination (weighted sum) of items from principal component analysis could be specified in structural equation analysis with a loading of unity and a measurement error variance of zero. A sum of items has been suggested as an alternative specification in structural equation analysis (Bagozzi and Heatherton 1994), and a sum of items has been used in structural equation analysis (e.g., Williams and Hazer 1986). Thus, principal component factor scores might be used in structural equation analysis to adequately specify a formative latent variable.

Specifically, items that have been argued to adequately define their latent variable, could be combined using their factor scores (weighted sums) from a principal components factoring (not a common-factor analysis), and then they could be specified as a single item latent variable in the structural equation analysis model with a loading of unity and a measurement error variance of zero.[2]

Nevertheless, there are drawbacks (to both specifications). Reflective measures may be weeded down to a few items so they fit the data with a reflective specification. As a result, they may be in danger of being judged bloated specific (i.e., they no longer adequately itemize the construct—see Cattell 1973). Or, a reflective measure can be discarded as psychometrically inadequate, and a potentially important construct is thus missing from the model. This in turn removes an hypothesis(es), and can raise the specter of the estimated model being an inadequate test of the proposed theory (because of the "missing variable" problem—see James 1980).

If a latent variable is re-specified as formative, it may receive a less-than-warm reception from reviewers, for the reasons mentioned above. In addition, in a re-specification from reflective to formative the items appear to no longer have any measurement error.[3] In addition, a measure re-specified as formative may be forever viewed as a formative measure.

However, there also are advantages with either specification. A reflective specification is typically well-understood by reviewers. A formative re-specification can retain items that are important to their construct's face or content validity. In addition, experience suggests that older, well established (even multidimensional), measures developed before the advent of structural equation analysis (e.g., Walker, Churchill and Ford 1977) might be re-specified as formative (see Williams and Hazer 1986), instead of being weeded to the point of being a "shadow of their former self" (e.g., Comer, Machleit and Lagace 1989). Formative factor scores more nearly approximate structural equation analysis' continuous data assumption, which is routinely ignored in the use of reflective ordinal indicators. Finally, many theory tests propose interesting new theory and a first test of that theory. A formative re-specification to "save" a measure in order to provide a first test of of new theory may be reasonable because insisting that measures be "perfect" actually may restrict the flow of knowledge—new knowledge might go unpublished until a "perfect" study is attained.

RE-SPECIFICATION

---

[2] Other specification approaches are plausible (see for example Jarvis, MacKenzie, and Podsakoff 2003).

[3] However, measurement error is present in both specifications. A reflective estimation of a measure produces explicit measurement errors. However, despite the fact that a formative specification of the same measure produces no explicit measurement error, the formative specification's items do contain measurement error, and that can be gauged using coefficient alpha (i.e., measurement error is independent of specification).

Re-specifying a measure generated to be specified reflectively involves three steps: identifying items that could be argued to be “bi-directional,” determining if these items adequately define the target construct, then creating principal components factor scores for the formative re-specification.

Specifically, the process should begin with item judging each item in the full candidate measure (not a weeded subset) for their formative directionality potential (e.g., will an increase in the item likely cause an increase in the target latent variable?). Next, the items that survive this item judging should be judged again to determine if they adequately define the target construct. Ideally these judgings would be done using a panel of experts who are familiar with the target construct and the construct's conceptual definition (i.e., using formal item judging). Then, the item-to-total correlations of the items that survived these item judgings should be examined, and any negatively correlated items should be reflected (recoded) to force all the surviving items to have positive item-to-total correlations. Next, any surviving items judged to be conceptually similar should be combined (averaged) to reduce overweighting (e.g., combine similar affect items, similar intention items, similar action items, similar attribute items, etc.). (Ideally, this also would be accomplished with formal item judging.) Alternatively, the conceptually similar item(s) with lesser item-to-total correlation could be dropped.

Next, using principal components (PC) (not common-factor) analysis, the surviving/combined items should be factored, and the resulting factors should be examined. Because important surviving/items to an adequate formative latent variable definition may be in Factor 2, etc., the Factor 1 items should be re-judged for their face validity (i.e., will dropping the Factor 2, etc. items materially degrade face validity?). Any Factor 2, etc. items that are judged to be critical to face validity then should be included in a forced single factor.

Then, PC factor scores should be obtained for the surviving/combined (perhaps forced) items (i.e., the surviving/combined items should be combined into a single (formative) indicator).

Finally, assuming the formative latent variable has adequate reliability and validity[4] [5] (see below), it could be specified with a single factor-score indicator using a loading of 1 and a measurement error of zero.

---

[4] Even though formative latent variables appear to be “error free,” error still exists in their items (see Footnote 3). Thus, because (model) “reproducibility” in the Campbell and Fiske (1959) sense (error is sufficiently low in the model that the observed structural effects (associations) are likely to be adequately reproduced in subsequent studies) is important in theory tests, the reliability and validity of any formative latent variable added to a theory-testing model should be evaluated. Stated differently, the reliability and validity of all the model latent variables, formative or otherwise, should be judged to be adequate, in order to improve the likelihood that the study results are reproducible.

A procedure for specifying a second-order latent variable formatively would be to item judge each first-order "indicator" latent variable for its formative potential (e.g., will an increase in the first-order "indicator" latent variable plausibly cause an increase in the target second-order latent variable?). Next, determine if the surviving first-order "indicator" latent variables adequately define (cause) the target construct. Ideally these judgings would be done using formal item judging (discussed earlier). Then, an argument to support the first-order "indicator" latent variables being used to adequately define the second-order latent variable should be created. (Parenthetically, this can be challenging—in SES, it was comparatively easy to find an unmeasured (missing) item that also could be used to define SES. For emphasis however, the issue is an adequate definition—judging the latent variable to be adequately defined—not necessarily an exhaustive definition.)

Next, psychometrically adequate (reflective) first-order latent variables should be obtained (i.e., they should be reliable and valid, each should fit its single construct measurement model, etc.). Then, (confirmatory) factor scores for each first-order latent variable should be produced using a measurement model for each first-order latent variable (i.e., linearly combine each first-order latent variable into its (confirmatory) single factor indicator).[6] Then, the resulting (confirmatory) single factor indicators should be factored again using principal components (PC) (not common-factor analysis) to obtain PC factor scores for the second-order latent variable. (I.e., the first-order latent variables should be turned into single indicators using their measurement model factor scores, then these single indicators should be combined using PC factor scores into a single indicator for the second-order latent variable.)

Finally, assuming adequate reliability and validity (see below), the second-order latent variable should be specified with this single (formative) indicator using a loading of 1 and a measurement error of zero.

It also may be important to construct a convincing argument for replacing the target latent variable's reflective specification with a formative re-specification. It probably is insufficient to argue for a formative specification simply because it is possible. A strong argument for "rescuing" the target latent variable probably would be more persuasive. (E.g., no psychometrically adequate reflective specification could be found, and the

---

[5] Social Scientists disagree on what constitutes adequate validity. Nevertheless, a minimal demonstration of validity probably should include content or face validity (how well items tap into the conceptual definition of their latent variable), construct validity (having plausible correlations with the other latent variables in the model), convergent validity (e.g., having adequate AVE) and discriminant validity (having small correlations with the other latent variables). Validity could then be qualitatively assessed considering reliability and the latent variable's performance over this minimal set of validity criteria.

[6] If one or more of the first-order "indicator" latent variables is "troubled," it could use principal components factor scores instead of common factor scores.

construct was in danger of being omitted from the model, which would require hypothesis omission and potentially increase the missing variable problem. Alternatively, one could argue that the formative latent variable was more face valid than the “final” psychometrically adequate reflective latent variable.

RELIABILITY AND VALIDITY

This topic may require a preface. Footnotes 3 and 4 mentioned error in formative latent variables, and the importance to model “reproducibility” (Campbell and Fiske 1959) of gauging formative latent variables' reliability and validity in theory testing. Although it was proposed for, and is used exclusively with, reflective latent variables, Coefficient Alpha (Cronbach 1951) should provide an adequate gauge of formative latent variable reliability: It gauges the ratio of “True Score” (measurement-error free) variance to total variance in a set of items, regardless of the intended use of the items.

While Average Variance Extracted (AVE) was proposed by Fornell and Larker (1981) using logic from Canonical Correlation, rather than a reflective specification, AVE is calculated using structural equation analysis (SEM) software that assumes a reflective specification—i.e., LISREL, EQS, AMOS, etc. However, it could be argued that a reflective specification is simply a device to estimate the common variance in a set of items. In different words, a set of items’ AVE is exists independently of the method used to estimate it.

However in real-world data, items that are re-specified using the suggested procedure are typically inconsistent, and their SEM AVE estimate may not be trustworthy. Nevertheless, experience suggests that for these items an estimate of their (formative) AVE is provided by the percent of variance (POV) statistic in (common) factor analysis.[7] Specifically, because it could be interpreted as the percent item (common) variance explained by the set of re-specified formative latent variable items, POV plausibly might be used to help gauge validity in a formative re-specification. In particular, after a formative re-specification is judged for its face, and construct validity (see Footnote 5), (common factor) POV could be used to judge convergent and discriminant validity. Specifically, (common factor) POV greater than 50% would suggest that items have at least 50% common variance, which is the Fornell and Larker (1981) criterion for adequate convergent validity. The discriminant validity of two latent variables would be suggested by AVE/POV’s that are greater than the squared correlation between the two latent variables, which is the Fornell and Larker (1981) criterion for adequate discriminant validity.

DISCUSSION

Authors have noted that in artificial data, a reflective specification of a latent variable that was first specified formatively produces biased structural coefficient(s) (i.e., an upward bias for the latent variable’s exogenous path, and a downward bias for its endogenous

[7] (Common factor analysis) percent of variance is the sum of each squared (common factor) loading (communalities) in Factor 1 divided by the sum of the (standardized) item variances, and it has the same algebraic form as AVE (see Fornell and Larker 1981).

path) (see Jarvis, MacKenzie, and Podsakoff 2003; MacKenzie, Podsakoff and Jarvis 2005). However, it is difficult to generalize these results to the suggested procedure because the items in the suggested formative re-specification of a "difficult" measure typically will be different from the items in its (consistent) reflective specification.

Nevertheless, it is possible that a formative re-specification of indicators that first were generated to be reflective might somehow bias paths to and from the re-specified latent variable. Thus, because the suggested formative re-specification approach has not been formally investigated for possible bias and inconsistency, a conservative significance criterion probably should be used with it (e.g., |t-values| for associations involving the re-specified latent variable(s) probably should be greater than 2.20 to suggest significance).

Perhaps curiously, experience suggests that the suggested formative re-specification procedure can improve a measure's reliability; or its face validity; or convergent or discriminant validity.[8] Organizational Commitment, for example, had an AVE of 0.49 in Ping 1997. Re-specified as formative using the suggested procedure, it had a POV of 65%.

Forcing Factor 2, etc. items into Factor 1 typically reduces the POV statistic, and it can produce a formative re-specification with a POV below 50%. This situation suggests an additional negative for the suggested formative re-specification procedure: item weeding still may be required in a formative re-specification.

However, POV might be improved by dropping cases from the data set (dropping cases for Organizational Commitment, for example, re-specified as formative using the suggested procedure, increased POV to 70%). A "Jackknife-like" (Efron 1981) procedure could be used to remove a case from the data set, and POV could be determined for the remaining cases. Then, the removed case is replaced, a different case is removed, and POV is determined for the remaining cases. This process is repeated for each of the rest of the cases to find the case that produces the largest POV improvement.[9]

---

[8] This appears to beg the question, how can the formative re-specification of an invalid reflective measure make it valid (in real-world data)? Obviously, the invalid and valid sets of items are likely to be different in size and composition. The formative set of items may be more reliable because it typically contains more items. The face validity of the formative items may be improved because important face validity items can be purposefully retained. Experience suggests that the suggested process of selecting items based on their apparent contribution to face validity may improve their convergent validity. Discriminant validity may be improved if convergent validity is improved. Perhaps curiously, in real-world data, experience suggests that construct (correlational) validity changes comparatively little in the suggested re-specification procedure.

[9] Dropping the case that detracts most from POV arguably is not a random process, and it casts a shadow over any sample "representativeness." An improved procedure would be to randomly select a case from a set of cases that detract most from POV. Alternatively, cases could be deleted randomly and the first case that improves POV could be dropped.

Additional POV improvement may be obtained by repeating this process using the data set with the case that produced the largest POV improvement removed, instead of the full data set. Specifically, a case is removed from the first improved POV data set, and POV is determined for the remaining cases. Then, the case just removed (not both cases) is replaced, a different case is removed, and POV is determined for the remaining cases. This process is repeated for each of the rest of the cases to find the largest POV improvement with two cases removed (also see Footnote 10).

SUMMARY

The document highlighted the distinction between a latent variable and its specification. It also suggested that a formative re-specification of a reflective latent variable might be efficacious in "difficult" measure situations,

1) as alternative to omitting items that are important to the measure's face or content validity in order to attain adequate internal consistency, or acceptable reliability or Average Variance Extracted (AVE);
2) when faced with omitting a latent variable from the model (and an hypothesis(es)) because of inadequate psychometrics (e.g., inadequate reliability or AVE);
3) as alternative to "breaking up" a second-order latent variable into its component ("indicator") first-order latent variables because of inadequate psychometrics.

Because reliability and validity are important in theory tests to promote "reproducibility" in the Campbell and Fiske (1959) sense (error is sufficiently low that the observed structural effects (associations) are likely to be adequately reproduced in subsequent studies), the discussion also suggested reliability and validity statistics for a reflective latent variable re-specified as formative using the suggested procedure.

REFERENCES

Anderson, James C. and David W. Gerbing (1984), "The Effect of Sampling Error on Convergence, Improper Solutions, and Goodness of Fit Indices for Maximum Likelihood Confirmatory Factor Analysis," *Psychometrika*, 49, 155-73.

Anderson, Erin (1988), "Transaction Costs as Determinants of Opportunism in Integrated and Independent Sales Forces," *Journal of Economic Behavior and Organization*, 9 (April), 247-64.

Bagozzi, Richard P. and Todd F. Heatherton (1994), "A General Approach to Representing Multifaceted Personality Constructs: Application to Self Esteem," *Structural Equation Modeling*, 1 (1), 35-67.

Blalock, H.M. (1964) *Causal Inferences in Nonexperimental Research*, Chapel Hill, NC: University of North Carolina Press.

Bollen, K. A. and R. Lennox (1991), "Conventional Wisdom on Measurement: A Structural Equation Perspective," *Psychological Bulletin* (110:2), 305-314.

Campbell, D. T. and D. W. Fiske (1959), Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," Psychological Bulletin, 56, 81-105.

Cattell, R. B. (1973), *Personality and Mood by Questionnaire*, San Francisco: Jossey-Bass.

Cohen, P., J. Cohen, J. Teresi, M. Marchi and C.N. Velez (1990), "Problems in the Measurement of Latent Variables in Structural Equations Causal Models," *Applied Psychological Measurement* (14:2), 183-196.

Comer, J. M., Machleit, K. A. and Lagace, R. R. (1989), "Psychometric Assessment of a Reduced Version of INDSALES, *Journal of Business Research*, 18, 291-302.

Cronbach, L. J. (1951), Coefficient Alpha and the Internal Structure of the Tests, Psychometrika, 16, 297-334.

Dwyer, F. Robert and Sejo Oh 1987, "Output Sector Munificence Effects on the Internal Political Economy of Marketing Channels, *Journal of Marketing Research*, XXIV (Nov.) 347-358.

Efron, B. (1981), Nonparametric Estimates of Standard Error: The Jackknife, the Bootstrap, and other Resampling Methods. *Biometrika*, 68, 589-599.

Fornell, Claes and David F. Larker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.

Gerbing, David W. and James C. Anderson (1993), "Monte Carlo Evaluations of Goodness-of-Fit Indices for Structural Equation Models," in *Testing Structural Equation Models*, K. A. Bollen and J. S. Long, eds., Newbury Park, CA: SAGE Publications.

Gerbing, David W., Janet G. Hamilton and Elizabeth B. Freeman (1994), "A Large-scale Second-order Structural Equation Model of the Influence of Management Participation on Organizational Planning Benefits," *Journal of Management*, 20, 859-85.

James, Lawrence R. (1980), "The Unmeasured Variables Problem in Path Analysis," *Journal of Applied Psychology*, 65 (4), 415-421.

Jarvis C., S. MacKenzie and P.A. Podsakoff (2003), "Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research," *Journal of Consumer Research*, 30 (2), 199-218.

MacKenzie S., P. Podsakoff and C. Jarvis (2005), "The Problem of Measurement Model Misspecification in Behavioral and Organizational Research and some Recommended Solutions," *Journal of Applied Psychology*, 90 (4): 710-730.

Ping, R.A. (1997), "Voice in Business-to-Business Relationships: Cost-of-Exit and Demographic Antecedents," *Journal of Retailing*, 73 (Summer), 261-281.

Ping, R.A. (2007), "Salesperson-Employer Relationships: Salesperson Responses to Relationship Problems and their Antecedents," the *Journal of Personal Selling and Sales Management*, 27 (1) (Winter), 39-57.

Ross, R.M. 1990, "The Evolution of Sex-Change Mechanisms in Fishes," *Environmental Biology of Fishes* 29: 81-93.

Walker, O.C., G.A. Churchill and N.M. Ford (1977), "Motivation and Performance in Industrial Selling: Present Knowledge and Needed Research," *Journal of Marketing Research*, 14 (May), 156-68.

Werts, C.E., R.L. Linn and K.G. Jöreskog (1974), "Intraclass Reliability Estimates: Testing Structural Assumptions," *Educational and Psychological Measurement*, 34, 25-33.

Williams, L. J. and Hazer, J. T. (1986), Antecedents and Consequences of Satisfaction and Commitment in Turnover Models: A Reanalysis using Latent Variable Structural Equation Methods, *Journal of Applied Psychology*, 71 (May), 219-231.

FREQUENTLY ASKED QUESTIONS (FAQ's) About Interactions and Quadratics...

The following are e-mailed questions, in no particular order of importance, that have been asked enough times that the responses appear below.

FAQ (D), how does one test hypothesized interaction(s) and/or quadratic(s)?, has become a "cook-book" that may be of some use to PhD students, substantive researchers, and educators who are interested in successfully estimating their first interaction(s) or quadratic(s). It contains a "fast start" feature for those who just want essentials.

The APA citation for this paper is Ping, R.A. (2006). "Frequently Asked Questions (FAQ's) About Interactions and Quadratics." [on-line paper]. http://www.wright.edu/~robert.ping/Faq.doc .

You may want to use "Find" to go to the answers to the frequently asked questions shown below. To do this, launch "Find" by clicking on Edit on the Word toolbar above, then click on Find. Next, copy and paste part of the desired question text shown below into the "Find what:" box. Then, click on "Find Next."

FREQUENTLY ASKED QUESTIONS:

A. What are the available latent variable interaction and quadratic estimation techniques?

B. What are the differences among them?

C. Which ones should be used for model testing?

D. How does one test hypothesized interactions or quadratics?

E. What about the assumptions behind these techniques, and violations of these assumptions in real-world data?

F. What if one or more measures have a natural zero point and mean or zero centering is inappropriate?

G. How does one investigate the possibility that a significant but unmodeled interaction or quadratic might be responsible for a nonsignificant hypothesized association?

H. How does one interpret a significant interaction or quadratic?

I. Can these interaction and quadratic estimation techniques be used with all of the popular structural equation modeling software packages?

J. How should reviewer comments regarding interactions and/or quadratics be handled?

K. How are Latent Variable Cubics estimated?

L. How is a "Second-Order" interaction estimated?

RESPONSES:

A. What are the available latent variable interaction and quadratic estimation techniques?

The accessible latent variable interaction and quadratic estimation techniques include,

1) Kenny and Judd (1984), which specifies an interaction *XZ* using indicators that are the unique cross products of *X* and *Z*. E.g., for *X* and *Z* with indicators x_1, x_2, ... , x_n and z_1, z_2, ... , z_m , *XZ* is specified with n times m product indicators, x_1z_1, x_1z_2, ... , x_1z_m, x_2z_1, x_2z_2, ... , x_2z_m, ... , x_nz_1, x_nz_2, ... , x_nz_m.

2) Bollen (1995)--*XZ* is specified with all Kenny and Judd (1984) product indicators, and 2-stage least squares estimation is used.

3) Jöreskog and Yang (1996), *XZ* is specified with from one to all Kenny and Judd (1984) product indicators, and it assumes intercepts for *X* and *Z*.

4) Ping (1995)--*XZ* is specified with a single indicator $x{:}z = (x_1 + x_2 + ... + x_n)(z_1 + z_2 + ... + z_m)$. x:z can be specified with a either a free, but constrained, loading and error term (direct estimation), or a previously calculated and fixed loading and error term (2-step estimation).

5) Ping (1996a)--*XZ* is specified with all Kenny and Judd product indicators. Coefficients are estimated using 2-step estimation.

6) Ping (1996c), which uses an adjusted covariance matrix and OLS or ML regression to estimate the coefficient(s) of interactions.

7) Jaccard and Wan (1995)--*XZ* is specified with a 4-indicator subset of the Kenny and Judd (1984) product indicators.

8) Jöreskog (2000)--*XZ* is specified with a single indicator that is the product of the "latent variable" scores (factor scores) of *X* and *Z*.

9) Wall and Amemiya (2001)-- *XZ* is specified with subsets of the Kenny and Judd (1984) product indicators, and the covariances of XZ with *X* and *Z* are freed.

10) Mathieu, Tannenbaum and Salas (1992)-- *XZ* is specified with a single indicator that is the sum of the indicators of *X* times the sum of the indicators of Z, and it uses a reliability

loading for that indicator with a reliability-based measurement error variance, Var_{XZ}*(1-reliability$_{XZ}$).

11) Algina and Moulder (2001)--proposed modifications to the Jöreskog and Yang (1996) approach to address convergence problems.

12) Marsh, Wen and Hau (2004)--*XZ* was specified with subsets of the Kenny and Judd (1984) product indicators, the loadings and measurement error variances of which were allowed to be free for estimation rather than fixed or constrained.

13) Klein and Moosbrugger (2000); Schermelleh-Engle, Kein and Moosbrugger (1998); Klein and Muthén (2002)--XZ is not explicitly specified with product indicators.

14) Moulder and Algina (2002)--*XZ* is specified with a 4-indicator subset of the Kenny and Judd (1984) product indicators, using the Jöreskog and Yang (1996) approach, and coefficients are estimated using 2-step estimation.

Other techniques include Hayduk (1987) and Wong and Long (1987) that require dummy variables to estimate a latent variable interaction.

B. What are the differences among them?

Techniques (1), (2), (3), and (5) use all the Kenny and Judd (1984) product indicators and thus require n times m of these indicators. Technique (3), (9), (11), (12) and possibly (13) and (14) can also be used with all the Kenny and Judd (1984) product indicators. Technique (7) uses a subset of the Kenny and Judd product indicators, and techniques (9), (12), (13) and (14) can use a subset of the Kenny and Judd product indicators. Techniques (4) and (10) use 1 indicator, x:z. Techniques (3), (11) and possibly (14) can use a single Kenny and Judd product indicator, usually the product indicator with a loading of 1. Techniques (1), (3), (7), (9), (11) and the direct estimation version of (4) require the nonlinear constraint equations (e.g., available in LISREL 8 and SAS's 'Proc Calis,' but not available in EQS or AMOS). Technique (13) requires the proprietary software QML. Techniques (2), (5), (6), (9), (10), (12), possibly (14) and the 2-step version of (4) can be used with EQS and AMOS, as well as LISREL 8 and Calis.

Technique (2) does not assume x_1, x_2, ... , x_n and z_1, z_2, ... , z_m are multivariate normal; the rest do make this assumption. Thus, technique (2) requires the use of the 2 Stage Least Squares estimator (the customary Maximum Likelihood estimator can not be used). Technique (6) can be used with OLS or ML Regression.

Techniques (3), (11) and (14) do not require zero or mean centered indicators of *X* and *Z*, and assumes intercept(s) for the structural equation(s) (a zero or mean centered indicator has a mean of zero, and is created by subtracting the indicator's mean from the indicator in each case in the data set). The other techniques implicitly assume the indicators of all the latent variables in the model, including the dependent variable(s), are zero or mean centered.

Technique (6) was proposed for interactions only, and was proposed with no standard error term (however, Ping 2001 has proposed a standard error term).

C. Which ones should be used for model testing?

There is little agreement on the "best" estimation technique. Formal comparisons of several of these techniques using artificial data (e.g., Li, Harmer, Duncan, Duncan, Acock and Boles 1998; Marsh, Wen and Hau 2004; Molder and Algina 2002) suggest some do not perform well, especially with nonnormal data. I have been unable to duplicate some of these results, and several lower-performing techniques appear to be the result of assumptions that are unrealistic in real-world data. Thus, experience with estimating an *XZ* structural coefficient, for example, in theory (hypothesis) tests of "interesting" models (i.e., models with more than 3 exogenous constructs, not including *XZ*) and over-determined *X* and *Z* (i.e., *X* and *Z* with 4 to 6 or more indicators), with real world survey data, suggests nearly all of these techniques produce interpretationally equivalent results. That is, standardized coefficients and t-values produce the same interpretations of the theory test. (Unstandardized coefficients will vary among these techniques because not all use Maximum Likelihood estimation are used, and techniques (3), (6), (11) and (14) produce intercepts which will change the unstandardized coefficients.)

They are all tedious to use. Technique (3) can produce convergence problems (however (11) is a modified procedure which reduces convergence problems in (3)). Some of these techniques are simply proposals, and have not been formally evaluated for any bias and efficiency (e.g., (8), (9) and (10)).

I personally am drawn to Kenny and Judd (1984) product indicator approaches that explicitly use all the Kenny and Judd product indicators. The use of all the Kenny and Judd product indicators is mathematically elegant and intuitively appealing. However, in "interesting" models (i.e., models with more than 3 exogenous constructs, not including *XZ*) with over-identified *X* and *Z* (i.e., with 4 to 6 or more indicators each), *XZ* specified with the full set of Kenny and Judd product indicators will almost always be inconsistent with real-world data (i.e., *XZ* in a single construct measurement model will not fit the data, and the full measurement and the structural model containing *XZ* will usually exhibit unacceptable model-to-data fit). As a result, I have found that using all the Kenny and Judd product indicators (and thus techniques (1), (2), (3), and (5), and (9), (11), (12) and possibly (13) and (14) with all Kenny and Judd (1984) product indicators), despite its appeal, is not particularly useful for theory testing with "interesting" structural models and over-identified *X* and *Z* in real-world data.

However, most of the techniques that use the full set of Kenny and Judd product indicators can be made to fit the data by using a subset of the Kenny and Judd product indicators. However, I am usually slow to use technique (7), and techniques (3), (9), (11), (12), (13) and (14) for theory testing with a subset of the Kenny and Judd product indicators for several reasons. In artificial data "weeding" the Kenny and Judd product indicators curiously does not bias the asymptotic results of these techniques. However, it is easy to show with real-world data that various choices of subsets of the Kenny and Judd product indicators can produce various structural coefficient values and significances, and thus interpretations. In addition, without all the Kenny and Judd product indicators present, *XZ*, for example, could be judged to be no longer content- or face valid, Further, the reliability of *XZ* with a weeded subset of the Kenny and Judd product indicators is unknown (the formula for the reliability of *XZ* is a function of *X* and *Z*, and thus it assumes *XZ* is operationally *X* times *Z* --see Bohrnstedt and Marwell 1978). Finally, a detailed interpretation of a significant "weeded" *XZ* that uses factored coefficients (see FAQ H) is problematic because *XZ* is no longer factorable (the *X* in *XZ* is no longer operationally the same as the latent variable *X* because they are itemized differently).

For hypothesis testing using survey data this usually leaves technique (4), and technique (6) (with the standard error term suggested in Ping 2001). However, technique (6) is limited to a single endogenous variable, and it is generally unknown outside of the methods literature. The direct estimation version of technique (4) requires tedious (manual) coding of constraint equations, and it is difficult to use with more than one interaction or quadratic.

If *X* or *Z* cannot be mean- or zero centered, neither technique (4) nor (6) is appropriate. I have seen proposals to use an input *correlation* matrix to avoid zero or mean centering, but unfortunately correlational structural analysis can alter the model structure, and it will change model-to-data fit and produce incorrect standard errors (see Cudeck 1989 and Jöreskog and Sörbom 1996). Techniques (3), (11) or (14) with a subset of the Kenny and Judd product indicators may be the only alternatives in this case.

It is widely believed that Maximum Likelihood (ML) coefficient estimates are robust to "reasonable" departures from normality that occur, for example, in survey data, but coefficient standard errors may not be. If the data is badly non normal or raising the t-value cutoff for the significance of *XZ*, for example, to more than 2 in absolute value, techniques (1), (3), (5), (9), (10), (12) and (14), and in particular the 2-step version of technique (4), can be used with EQS's 'ROBUST' ML estimator to obtain better estimates of the coefficient standard errors.

D. How does one test hypothesized interactions or quadratics?

Unfortunately the answer is, "with considerable effort" when latent variables are involved. To understand why, some background is desirable. (As a less desirable alternative, one could skip down to the "**In summary**..." paragraph below.)

The largest barrier to latent variable interaction/quadratic estimation in my opinion is the amount of work involved for even a single interaction or quadratic. This is changing, however. For example, technique (8) has been implemented in LISREL's PRELIS (although it has received no formal evaluation, and preliminary results suggest it produces unusual standard errors--see Schumacker 2002). This web site also has EXCEL templates to expedite the computation of interaction and quadratic loadings and measurement error variances.

The next largest barrier to latent variable interaction/quadratic estimation with "interesting" models (i.e., models with more than 3 exogenous constructs, not including *XZ*) and over-determined *X* and *Z* (i.e., *X* and *Z* with 4 to 6 or more indicators) in real world data seems to be model-to-data fit. A sufficient condition for unidimensionality of *X*, for example, is that its single construct measurement model (i.e., one involving only *X* and its indicators) fit the data (e.g., the p-value of chi-square is at least slightly non zero). Thus, for *X*, *Z*, and *XZ*, and the dependent latent variable *Y*, for example, the single construct measurement model for each of these variables should fit the data well.

Without good single-construct measurement model fits, the structural model fit will be degraded, and adding interactions/quadratics frequently makes things worse. Adding an indicator for an interaction that is the product of other indicators does not improve model to data fit. In fact, with more than about 6 Kenny and Judd product indicators, model fit becomes unacceptable in "interesting" models with over-determined *X* and *Z* in real world data. As Bagozzi and Heatherton (1994) and Gerbing and Anderson (1993) point out, the same "about 6 items" limit also seems to apply to the items of *X*, *Z*, etc. (especially in "interesting" models and over-determined *X* and *Z*). This usually means that a single construct measurement model of an interaction or quadratic specified with more than about 6 product indicators will not fit the data without somehow reducing the number of product indicators. (see "On the Maximum of About Six Indicators per Latent Variable with Real-World Data" on this web site for more on this "puzzle of about six indicators").

Although probably not initially intended as an interaction estimation technique, Jaccard and Wan (1995) addressed these model fit problems by "weeding" or deleting Kenny and Judd product indicators as one does to attain an internally consistent *X* or *Z* (i.e., to attain acceptable single construct model-to-data fit). Possibly as a result, most of the more recent interaction techniques mentioned in FAQ A use subsets of the Kenny and Judd product indicators. However, experience suggests this can produce different structural coefficients and t-values, and thus interpretations in a single real world data set, depending on the subset of product indicators used. (Curiously the Jaccard and Wan 1995 results, for example, were acceptably asymptotically unbiased in artificial data by using "weeded" subsets of Kenny and Judd product indicators.) It could also be argued that *XZ* is no longer content- or face valid when its Kenny and Judd product indicators are omitted. In addition, the reliability of *XZ* is undefined, and detailed interpretation of *XZ* is problematic (see FAQ C).

Correlating measurement errors in the indicators of *X*, *Z* or *XZ* simply to improve model fit is not allowed: None of the techniques discussed in FAQ C are valid if this is done (however, see *Latent Variable Interactions...* Chapter VIII: Intercorrelations on this web site for corrections to the specifications with correlated measurement errors). Correlating structural disturbances (i.e., the estimation error(s) for the dependent or endogenous variable(s)) to improve model fit may or may not be a good idea, depending on the substantive theory behind the model. Data transformations to improve model fit may make coefficient interpretation difficult.

Model fit may also be degraded by interaction or quadratic specification. For example, the variance of the interaction *XZ* is frequently constrained to equal the Kenny and Judd (1984) value Var[*X*]*Var[*Z*]+Cov[*X*,*Z*]**2, where ** indicates "raised to the power." However, this can produce unacceptable model fit or convergence problems in real world data, and experience suggests this may bias the structural coefficient of *XZ* in real-world data. This can occur because constraining the variance of *XZ* to Var[*X*]*Var[*Z*]+Cov[*X*,*Z*]**2 assumes the data is multivariate normal, which is seldom true in survey data. Thus, the variance of *XZ* should be free rather than constrained to the Kenny and Judd value Var[*X*]*Var[*Z*]+Cov[*X*,*Z*]**2 in survey data.

Another specification in survey data that degrades model fit is not freeing the correlations among *XZ*, *X* and *Z*. Although the correlation of *XZ* with *X* or *Z*, *should be* zero in multivariate normal data (see Kenny and Judd 1984), in real-world data *X* and *Z* are seldom sufficiently multivariate normal to not be correlated with *XZ*, and *XZ* should be free to correlate with and *X* and *Z* in their structural model to avoid model fit problems, and possible bias in the structural coefficient of *XZ*.

Returning to the barriers, estimation convergence can be a problem (i.e., no admissible estimates are produced because the iteration limit is exceeded) and improper solutions may obtain even with convergence. To avoid these problems, a structural model with an interaction(s) and/or quadratic(s) will usually need input starting values for the interaction/quadratic parameters (i.e., manually calculated loadings, measurement error variances, and interaction/ quadratic variances/covariances). It also may need starting values for all the structural coefficients, and the structural disturbance(s). While this is annoying, starting values for structural coefficients and structural disturbance(s) can be obtained using OLS regression (the structural disturbance, e, for *Y* in $Y = b_1X + b_2Z + ... + b_nW + e$ is estimated by Var(Y)(1-R^2), where Var(Y) is the SPSS, SAS, etc. variance of Y, and R^2 is from the OLS regression of Y on its independent variables). A starting value for the variance of *XZ*, Var(*XZ*), is approximately the SPSS, SAS, etc. variance of XZ, Var(XZ). Starting values for the loadings and measurement error variances of *XZ* can be computed using the EXCEL spreadsheets on this web site.

In addition, the covariance matrix used to estimate the model should contain variances and covariances that are about equal to 1 to avoid computational problems during estimation. Numerically large variances in the covariance matrix to be estimated can produce a large determinant of the covariance matrix, and the reciprocal of this determinant is used to estimate the model. If this determinant is large, its reciprocal is a number that is near zero, and the model may be empirically not identified and it may not converge. Thus, if convergence is a problem after providing good starting values for every estimated parameter in the model, consider scaling down any unusually large indicator variances. The variances of indicators of a latent variable should be approximately the same if they are congeneric. If not, check for input errors (e.g., "1" keyed as "10," etc.). Next, scale the large (first order, e.g., *X*, *Z*, etc.) indicator variances (interaction and quadratic variance are scaled indirectly as will be explained later). Although there is little guidance for scaling in this situation, it useful to think of scaling as re-coding a variable from cents to dollars; from a Likert scale of 1, 2, 3, 4, 5 to a scale of .2, .4, .6, .8, 1, for example, using a scaling factor of 5. The effect of a scaling factor is squared in the resulting variance, so if variance should be reduced by a factor of 5, plan to divide each case value by the square root of 5. Further, interactions and quadratics should not be scaled directly. Plan to scale their constituent variables instead--scaling X and Z, for example will automatically scale XZ by the product of the squares of the scaling factors for X and Z. Finally, verify that all the indicators of a construct have about the same scaled variance, and start by scaling the largest variance in the input covariance matrix (the entire matrix does not have to be scaled). In addition to changing variances, scaling will change indicator loadings, and scaling will usually affect unstandardized structural coefficients in the model (standardized coefficients should be unchanged).

In summary, to test one or more hypothesized latent variable interaction or quadratic in a theoretical model, consider choosing an estimation technique that uses popular structural equation software and an estimator that reviewers will recognize. (Even though it unrealistically assumes multivariate normality, Maximum Likelihood estimates are generally preferred for theory testing in the Social Sciences. Fortunately, its coefficient estimates appear to be robust to "reasonable" departures from normality, but its standard errors are believed to be biased. Thus, the customary t-value cutoff of 2 in absolute value should probably be increased for an interaction/quadratic, or an estimator such as EQS's "Robust" ML estimator should be used to obtain a better estimate of coefficient standard errors.) Also consider using an estimation technique that has been formally evaluated, one that uses content- or face valid indicators, and one for which reliability can be computed and detailed interpretation is not problematical. The estimation technique chosen should also be likely to converge and produce acceptable or admissible parameter estimates, and adequate model-to-data fit. At the risk of appearing self promotional, the 2-step version of technique (4) was developed to meet these criteria.

Next, the reliability and validity of any interaction/quadratic to be estimated should be gauged. In theoretical model tests reliability and validity are important to avoid estimation results and thus interpretations that are an artifact of measurement (e.g., based primarily on measurement error). Interaction and quadratic reliabilities are available in the EXCEL spreadsheets on this web site. (However, these reliabilities assume *XZ*, for example, is itemized with all the indicators of *X* and all the indicators of *Z*--the reliability of *XZ* itemized with subsets of the Kenny and Judd product indicators is unknown.) A minimal demonstration of validity should probably include: content or face validity (how well the a latent variable's indicators match or tap its conceptual definition), construct validity (its correlations with other latent variables are theoretically sound), convergent validity (e.g., its average extracted variance is greater than .5--see Fornell and Larker 1981), and discriminant validity (e.g., its correlations with other measures are less than .7--also see Fornell and Larker 1981 for a more stringent criterion) (e.g., Bollen 1989, DeVellis 1991, Fornell and Larker 1981, Nunnally 1978). Thus, an interaction, *XZ* for example, might be judged content or face valid if *X* and *Z* are content valid and the specification of *XZ* includes all the indicators of *X* and *Z*. Interaction and quadratic Average Extracted Variances are available in the EXCEL spreadsheets on this web site. However, the construct (correlational) validity of an interaction or a quadratic is usually impossible to judge.

If the interaction(s) or quadratic(s) are not reliable (i.e., .7 or above), they probably should not be estimated because they will also not be convergent valid, suggesting they are composed of more than 50% error variance, and they may also be discriminant invalid using Fornell and Larker's (1981) criterion. If they are reliable but fail one or more validity tests, these validity deficiencies should be noted in the limitation portion of any final report.

Parenthetically, consider not deleting cases to reduce nonnormality--a necessary condition for interactions is nonnormality in the data. Then, mean or zero-center *all* the structural model variables, even the dependent/endogenous variables (e.g., *Y*) by subtracting each indicator's mean from its value in each of the cases.

Next, create the indicators of the interaction(s)/quadratic(s) in the data set, x_iz_j, for example, for most of the techniques in FAQ A, or x:z for technique (4). If "Direct Estimation" using technique (4) is to be used, the single indicator should not be formed using averages of the indicators of X or Z, for example, because averaging seems to produce estimation problems. Otherwise a summed or averaged indicator product, x:z for example, could be used for the "2-Step" version of technique (4), and summed indicators should be used for latent variable regression to be consistent with the EXCEL template on this web site, and the examples in *Latent Variable Interactions...* monograph (beginning with Chapter VIII: Suggestions/"Step-by-Step"), which is also on this web site. However, and **averaged** indicator product, x:z for example, is preferred because it produces an interaction/quadratic variance that does not overwhelm the input covariance matrix because of its magnitude (which can produce estimation problems).

Next, compute starting values for at least the interaction(s) and/or quadratic(s) parameters-- this includes estimates of all the free covariances with the other latent variables in the model-- and specify these in the structural model (measurement model estimates of variance and covariance should be sufficient). Be certain to estimate starting values consistently using summed indicators or averaged indications (e.g., avoid using summed starting values with averaged indicators). If more starting values are desirable (i.e., because the structural model estimates fail to pass the adequacy tests discussed below), additional measurement model parameter estimates and structural coefficient estimates from regression should be sufficient.

Consider analyzing a covariance matrix--analyzing correlation matrices should probably be avoided. Avoid correlating indicator measurement errors in the structural model--this violates the assumptions in most of the latent variable interaction and quadratics estimation approaches and, although corrected specification equations are available (e.g., *Latent Variable Interactions...* Chapter VIII: Intercorrelations on this web site), interaction/quadratic specification is much more complicated, and so many authors have warned against correlated measurement errors that their use may not be acceptable to reviewers.

Free the variance of the interaction(s) or quadratic(s). Similarly, allow the interaction(s) and/or quadratic(s) to correlate by freeing the correlational paths between them. Also allow the interaction(s) and/or quadratic(s) to correlate with the other exogenous latent variables in the model by freeing the correlational paths between them (e.g., *XZ* should be correlated with *X*, *Z* and the other exogenous latent variables). (interactions involving an endogenous LV, "endogenous interactions," have additional considerations and this matter is discussed in "Questions of the Moment" on this web site.) If there are large variances in the input covariance matrix or the covariance matrix implied by the raw input data (i.e., the variances are not all about the same size), scale the large variances or the raw data to lessen the chance of estimation difficulties.

For every interaction (e.g., *XZ*) to be estimated, consider also estimating the two related quadratics

(*XX* and *ZZ*). An interaction can be mistaken for a quadratic (see Lubinski and Humphreys 1990) (and vice versa, see *Latent Variable Interactions...* on this web site), and authors recommend estimating *XZ* in the presence of *XX* and *ZZ* as a stronger test of an hypothesized interaction (i.e., in competition with its related quadratics).

If more than one interaction or quadratic is to be estimated, they should all be estimated together in one model. However, just as adding Z and W, for example, to a model with X and Y can change the significance of the X-Y structural coefficient, adding XZ, XW and ZW can attenuate other structural coefficients, including those for XZ-Y, XW-Y and ZW-Y. This is especially true for XZ-Y and XW-Y, for example, because they share a common constituent variable X. Thus, estimating multiple interactions can produce significance difficulties among the interactions (the interested reader is directed to *Latent Variable Interactions...* Chapters VIII and IX on this web site for details on over coming this difficulty).

Examples of LISREL and other estimation "programs" are shown in Cortina, Chen and Dunlap (2001); Li, Harmer, Duncan et al (1998); Schumacker (2002); Schumacker and Marcoulides (1998) (e.g., Chapter 4); and *Latent Variable Interactions...* Chapter VIII: Suggestions/"Step-by-Step" (on this web site). Unfortunately, some of these examples contain errors. In addition, the variance of *XZ*, for example, is typically constrained rather than free, and *XZ* is typically not allowed to correlate with *X* and *Z* (while these are the correct recommendations for normally distributed data, real world survey data is seldom sufficiently normal for a constrained and uncorrelated *XZ* to avoid producing reduced model-to-data fit, and biased structural coefficients.

Once the model is estimated, examine model fit and the standardized structural coefficient estimates. Also examine the estimated variances of the model constructs, and the error terms of the structural equations (i.e., structural disturbances). Model fit should be acceptable using a sensitive fit index such as RMSEA (i.e., .05 or less suggests close fit, .051-.08 suggests acceptable fit-- see Brown and Cudeck 1993, and Jöreskog 1993). The standardized structural coefficient estimates should be between -1 and +1, and the structural disturbances should all be positive. The estimated variances and covariances of the model LV's should be positive and similar to than their error-attenuated (i.e., SAS, SPSS, etc.) counterparts. (Internally consistent LV's in the Gerbing and Anderson 1988 sense produce structural model LV variances and covariances that are nearly identical to their measurement model counterparts. This provides one of the strongest demonstrations of trustworthy structural model parameter estimates.) In addition, the structural model loadings and measurement error variances should all be between 0 and 1, and for internally consistent LV's they should all be within a few points of their measurement model values. Finally, a regression model of the interaction (e.g., XZ, X, Z, for example, and their predicted variable, Y) should produce unstandardized regression coefficients that are interpretationally equivalent to their corresponding structural coefficients in the SEM interaction model (corresponding t-values should have the same sign and roughly the same magnitude).

If the model estimation fails to pass any of these tests, the problem is almost always one of four things: multicollinearity, misspecification, incorrect or insufficient starting values, or empirical underidentification. To investigate these, verify mean centering: were the *indicators* mean centered *before* the interaction/quadratic indicator(s) were formed in the data set? Try mean centering *all* the *indicators* in the model, even those not involved in the interactions/quadratics. Next, check model specification. Are all the exogenous variables, including the interactions/quadratics, free to correlate? Are the indicator loadings between 0 and 1 (e.g., for each multiple indicator latent variable was one indicator fixed at one to provide a metric, and are there indicators larger than 1?--if so, fix the largest indicator to 1, etc.). Then check to see that *all* the starting values for the structural coefficients are non zero, and none of the variances and covariances of the constructs (e.g., PHI's in LISREL) are zero. Was the variance of *XZ* and/or *XX* constrained? If so, try freeing it. If direct estimation is being used, try using 2-Step estimation. If there are large variances in the indicator covariance matrix try scaling them. If none of these work, please send me an e-mail.

If 2-step estimation is used, the measurement parameter estimates for *X* and *Z*, for example, in the structural model should be very close to their counterparts in the measurement model (i.e., they should be the same to 2 or more decimal digits between the (full) measurement and structural models). If they are not, the structural model should be re-estimated replacing the calculated loadings and measurement error variances for *XZ* with loadings and measurement error variances for *X* and *Z* from the structural model.

If *XZ* is estimated with its related quadratics *XX* and *ZZ*, this should be done in two model estimations. In estimation 1, constrain the interactions/quadratics' structural coefficients to zero, and examine their resulting modification indices (LMTEST in EQS) for the largest one (i.e., a modification index above about 3.8, which roughly corresponds to a structural coefficient t-value of 2 with 1 degree of freedom). For emphasis, avoid jointly freeing the *XZ-Y*, *XX-Y*, and *ZZ-Y* associations--the frequent result is that all three associations are nonsignificant (NS) because *XZ*, *XX*, and *ZZ* are usually highly correlated. If the modification index for the hypothesized *XZ-Y* association is significant (about 3.8 or above), remove the related quadratics from the structural model and free the *XZ-Y* association in a second model estimation (estimation 2) to verify the *XZ-Y* significance, then report the results. Even if the modification indices for the *XX-Y* and/or the *ZZ-Y* associations are also "significant" (and/or larger), one's hypothesis predicts the *XZ-Y* association, and the *XX-Y* or *ZZ-Y* modification index may be "significant" by chance.

If the modification index/structural coefficient for the hypothesized interaction is non significant in estimation 1, there are several possibilities. To investigate further, obtain estimation 2. If the interaction is again non significant, this suggests the interaction is non significant and the interaction hypothesis is disconfirmed. However, if the interaction is significant in estimation 2 (i.e., without the related quadratics) and one or more related quadratics were significant in estimation 1, this suggests the interaction and one or more if its related quadratics may be "interchangeable." For example if the *XX-Y* structural coefficient was significant in estimation 1, this suggests that replacing the interaction with *XX* in estimation 2 is also significant. One could do one of four things at this point: Ignore the interchangeable quadratic because it was not hypothesized and report the significant estimation 2 interaction. Or, one could interpret the significant estimation 2 interaction and comment on the alternative equation 2 specification with *XX* and its implications. Specifically, an interchangeable *XX* suggests that Z moderates the *X-Y* association as hypothesized, but *X* also moderates itself in the *X-Y* association (but not both). This "self moderation" is also called "satiation" or "diminishing returns" because the *XX-Y* association is shaped like part of a horseshoe (see Chapter III: Visualizing... in *Latent Variable Interactions...* on this web site), and the *XX-Y* relationship may be much more "interesting" from theoretical and practical standpoints (see Howard 1989 for interesting examples of quadratic relationships). For example, experience suggests that the *XX-Y* specification will produce a larger standardized coefficient, suggesting that *XX* explains more variance in Y than XZ does. However, because XX may have been significant by chance, another study is indicated to sort this matter out further. The third alternative is to ignore the interchangeable quadratic and report the significant estimation 2 interaction, then design a new study pitting the interaction against the quadratic (see *Testing Latent Variable Models with Survey Data*, Step III on this web site for a comparatively easy study that could be used to sort this matter out further). The fourth alternative is to combine alternatives 2 and 3 by reporting the second study's results with the first.

For emphasis, if the structural coefficients for *XZ-Y*, *XX-Y*, and *ZZ-Y* are all non significant in estimation 1 (i.e., their modification indices are below about 3.8), this does not always mean that *XZ-Y* is non significant. Thus, estimation 2 should be obtained to verify the *XZ-Y* association.

Consider analyzing a covariance matrix if possible. Occasionally, a correlational matrix, rather than a covariance matrix, is used as the matrix to be analyzed in a theory test. Despite statements or implications in their user manuals, LISREL, EQS, AMOS, etc. all appear to assume that a covariance matrix is to be analyzed, and analyzing a correlation matrix usually changes model-to-data fit (chi-square is typically incorrect), and it produces incorrect standard errors (which introduces Type I and Type II errors-- see Cudeck 1989, Jöreskog and Sörbom 1996). If a correlation matrix must be analyzed in a theory test, consider comparing the results with those from a covariance matrix estimation. If both models fit the data and they are interpretationally equivalent (i.e., the set of significant variables and their interpretations are the same in both estimations), then the use of correlations probably will not be misleading.

If an hypothesized interaction and/or quadratic is nonsignificant, it is usually because of model problems, its reliability is too low, the data set is too small, the inclusion of other interactions/quadratics, and/or they hypothesized moderated relationship is actually quadratic/cubic. It is always a good idea to verify that model specification and estimation are not contributing to nonsignificance. Specifically, consider re-verifying that the model is properly specified (e.g., the correlations among the exogenous variables, including the interactions, are free; the correlations between exogenous variables and endogenous variables are not free; structural disturbances are not correlated, etc.). Then, check that the interactions are properly specified (e.g., the "essential" correlations between *X-XZ* and *Z-XZ*, for example, are free, the variance of *XZ* is also free, and the values for the loading and measurement error variance values have been properly calculated and keyed into the estimation program), and the model indicators are all mean- or zero-centered. Next, verify that the structural model fits the data, all the coefficient estimates are admissible (see Step VI, "Admissible Solutions Revisited" in the *Testing Latent Variable Models Using Survey Data* monograph on this web site), and the measurement parameters of *X* and *Z* in the structural model (i.e., the loadings, measurement error variances and the variances of *X* and *Z*) are within a few points of their measurement model values. If the measurement parameters of *X* and *Z* in the structural model are different from their measurement model values, recalculate the interaction's loading and measurement error variance using the structural model measurement parameter values.

To improve reliability the interested reader is directed to Netemeyer, Johnson and Burton (1990). Insufficient sample size can be checked by recalculating the t-value that would result from using a larger sample size (N) in the equation t*SQRT(N/n), where t is the current t-value, "*" indicates multiplication, SQRT is the square root function and n is the current sample size. If the reliability of *XZ* is 0.7 or above, its Average Variance Extracted is above .5, and a few more cases would push the t-value above 2, it is tempting to simply declare that the interaction "approaches significance" and proceed as though the interaction were significant. This is because a t-value of 2 for significance is simply a convention. Stated differently, no statistical assumptions are violated by declaring that a t-value greater than 1.95 in absolute is likely to be non zero in the population. However, because the standard error of *XZ* is believed to be biased with Maximum Likelihood estimation, a t-value larger than 2 in absolute values should probably be used to gauge the significance of *XZ*.

There are other estimators and estimation approaches that may produce a significant interaction that is non significant with Maximum Likelihood. EQS for example provides a ROBUST ML estimator that is less affected by non normality in the data. Other estimation approaches include bootstrapping the interaction's structural coefficient (i.e., averaging the resulting coefficients and standard errors--see "Bootstrapping" in the monograph *Testing Latent Variable Models Using Survey Data* (Step V) on this web site), or removing influential case(s) (outliners that contribute most to "flattening" the XZ-Y regression line) using Cook's distance in regression, or a scatterplot of the interaction, then re-estimating the structural model. However, all of these estimators and approaches have their drawbacks, most telling of which is typically reviewer resistance to anything that is not simple, straightforward and familiar (a variation on parsimony or Occam's razor, see Charlesworth 1956). It may be easier to quickly conduct another study using Scenario Analysis, which is described next. A significant interaction in the second study would lend weight to an insufficient sample-size argument.

A less desirable alternative is of course to conduct another study. A Scenario Analysis using student subjects (see the *Testing Latent Variable Models Using Survey Data*, Step III on this web site) might provide a comparatively easily executed second study to obtain a larger sample. Because a scenario analysis is an experiment, it also increases nonnormality which McClelland and Judd 1993 suggest will increase the likelihood of a significant interaction. This second study could be reported with the first, as in done in several social science disciplines (e.g., Social Psychology and Consumer Behavior).

Occasionally, a non significant hypothesized interaction/quadratic is the result of the presence of other hypothesized interactions and/or quadratics--interactions and quadratics are typically highly intercorrelated and as a result they can mask each other. To investigate this, try constraining the interactions/quadratics' structural coefficients to zero, and examine the resulting modification indices (LMTEST in EQS) for a significant modification index (i.e., a modification index above about 3.8, which roughly corresponds to a path coefficient t-value of 2 with 1 degree of freedom). Next, free this structural coefficient and examine the resulting modification indices for the un-freed interaction/quadratic(s). This process of examining modification indices, freeing interaction/quadratic(s), and reexamining modification indices should identify the interaction/quadratic that is "suppressing" the others.

Assuming the suppressor interaction is also non significant, an approach that *might* be defensible would be to "trim" or remove the suppressing interaction(s). "Might" of course would depend on the reviewers. Trimming non significant associations was a common practice in theory tests years ago, especially in studies that might have an intervention component (e.g., Sociology). However, based on the research behind *Testing Latent Variable Models Using Survey Data* on this web site, its use in theory tests has declined. If non significant interactions are trimmed, an additional study is desirable to investigate the element of chance introduced by this trimming. A Scenario Analysis using student subjects (see the *Testing Latent Variable Models Using Survey Data*, Step III on this web site) might provide a comparatively easily executed second study of the trimmed model using the existing questionnaire. Specifically, the hypotheses involving the NS interactions could also be trimmed for the second study and it would investigate the model without the trimmed interactions. The result could become a paper with two studies. "Multiple study" papers are common in social science disciplines such as Social Psychology and Consumer Psychology and Consumer Behavior, and it might be instructive to examine a few of them to determine how best to present two-study results (see recent issues of *The J. of Consumer Research*, for example).

E. What about the assumptions behind these techniques, and violations of these assumptions in real-world data?

All the techniques just discussed in FAQ (C) above, except for (2) which assumes 2-stage least square estimation, assume that the indicators of *X* and *Z* are multivariate normal. All but technique (3) assume each latent variable indicator in the structural model is mean or zero centered (i.e., the latent variables each have a mean of zero). They all assume that indicator measurement errors are not correlated (however, see *Latent Variable Interactions...* Chapter VIII on this web site for corrected specifications in the presence of correlated measurement errors). These assumptions were made to simplify the algebra used to derive each technique.

However, because survey data is seldom multivariate normal, the multivariate normal (ML) assumption behind most of the latent variable (LV) interaction/quadratic specification techniques is seldom met. However, substantive researchers have generally ignored this same assumption behind OLS Regression for years (the implementation of OLS Regression in SAS, SPSS, etc. standard errors assume the variables are normally distributed). Nevertheless, studies suggest that ML coefficient estimates are robust to "moderate" departures from normality, but the standard errors are not (assuming proper model specification such as a free variance of *XZ* and free correlations between *XZ* and *X* and *Z*). If a structural coefficient has a t-value close to 2, EQS's ROBUST Maximum Likelihood estimator, which is less distributionally dependent, could be used to shed more light on significance. Experience suggests, however, that t-values are only slightly changed for these LV's and models in real-world data, sometimes in the "wrong" direction, and an alternative is to increase the *XZ* significant-t-value cutoff 5% to 2.1 in absolute value.

The mean or zero centering assumption, however, is typically essential to successful estimation. Technique (3) does not make this assumption, and technique (6) discusses un-centered variables, but no guidance on how to use (6) with un-centered variables is provided. Although there are proposals to use an input correlation matrix to avoid zero or mean centering, correlational structural analysis can alter the model structure, change model-to-data fit, and produce incorrect standard errors--see Cudeck 1989 and Jöreskog and Sörbom (1996).

Experience suggests that it is sometimes possible with real-word data not to mean-center one or more variables, and still be able to estimate an interaction involving the un-centered variable. In this case success appears to depend on the amount of collinearity between the un-centered variable and the interaction. If this collinearity is high, as it usually is, successful estimation is frequently impossible without "ridge" estimation and unreasonably large structural coefficient(s).

However, there may be another alternative to mean centering, which is to use median splits of the data to detect an interaction or a quadratic with un-centered data. This approach is criticized because it *can* produce false negative or, occasionally, false positive interactions. However, the results reported in Ping (1996b) suggest that for sufficiently reliable latent variables, median splits could be relied upon when the differences across the split are sufficiently significant. There are no hard and fast rules, but reliability should be as high as possible and probably above .8, and significance should be high and probably above t = 2.5 (or very low and below t = 1). There are some drawbacks to this approach, however. Median splits do not test the hypothesized latent variable model. The interaction is missing from the final test of the hypothesized model and the effects of its correlations with other predictors on the model's structural coefficients are not accounted for (the "missing variable" problem--see James 1980).

F. What if one or more measures have a natural zero point and mean or zero centering is inappropriate?

As previously mentioned, technique (3) does not make this assumption. In addition, it is occasionally possible with real-word data not to mean-center one or more variables, and still be able to estimate an interaction involving the un-centered variable. However, success appears to depend on the amount of collinearity between an un-centered variable and its related interaction (e.g., X and XZ). If this collinearity is high, successful estimation is frequently impossible using the techniques discussed in FAQ A above, and/or the structural coefficient(s) are impossibly large. Alternatives to mean centering are discussed in Paragraph 3 of FAQ (E) above and in *Latent Variable Interactions...*, Chapter VIII: Mean Centering, on this web site.

G. How does one investigate the possibility that a significant but unmodeled interaction or quadratic might be responsible for a nonsignificant hypothesized association?

If the Z-Y association in $Y = b_1X + b_2Z + b_3W$ is hypothesized to be significant (i.e., b_2 should be significant) but it turned out to be non significant, one could ask, is there an interaction or quadratic suppressing b_2 (i.e., XZ, ZZ or ZW-- a relevant suppressor of b_2 will involve Z)? This occurs more often than substantive researchers realize, and it is one explanation for an hypothesized association being significant in one study and not significant or with the opposite sign in another study (i.e., inconsistent results across studies).

To investigate the possibility of unhypothesized but significant interactions and/or quadratics, several non equivalent approaches could be taken. One approach would be to perform the overall F test described in FAQ G and then examine the interactions and quadratics used for this F test for significant relevant suppressor(s). However it is very likely in real world data that there will be no significant interactions or quadratics in the set of interactions and quadratics that were involved in estimating this F. It is also likely that any significant suppressors in this set will become nonsignificant when the other nonsignificant interactions and quadratics are trimmed. Further, if any significant suppressors remain after trimming, it is also likely that one or more of them can become nonsignificant if a significant interaction or quadratic is temporarily removed (i.e., significance depends on the presence of other interactions or quadratics).

Another approach would be to use the technique discussed in *Latent Variable Interactions...*, Chapter IX on this web site) on the "relevant suppressors" (i.e., each nonsignificant association in the model could be probed for moderation-- is it being suppressed?). The interested reader is directed to *Latent Variable Interactions...*, Chapter IX on this web site for the details.

H. How does one interpret a significant interaction or quadratic?

Interpretation approaches such as graphing (see Aiken and West 1991), which are popular with ANOVA and categorical data, are one approach. Another approach that is more revealing involves factored coefficients (e.g., the structural equation $Y = b_1X + b_2Z + b_3XZ$ can be factored into $Y = b_2Z + (b_1 + b_3Z)X$, and the factored coefficient of X is $(b_1 + b_3Z)$) is interpreted). For more, the interested reader is directed to the paper "Interpreting Latent Variable Interactions" on this web site, and *Latent Variable Interactions...*, Chapter III also on this web site.

I. Can these interaction and quadratic estimation techniques be used with all of the popular structural equation modeling software packages?

With some exceptions, yes. Technique (13) requires a proprietary software package available from the authors of that technique. Direct estimation using LISREL's constraint equations or the CALIS equivalent can be used only with those software packages. The 2-step techniques can be used with any of the popular structural equation software packages (e.g., LISREL, EQS, CALIS, AMOS, etc.). Some time ago, AMOS was alleged to be having difficulty calculating some model fit indices with mean-centered data. Specifically, some AMOS fit indices are alleged to be incorrect (too large) with mean centered data, which may erroneously suggest lack of fit. However, RMSEA and CFI appeared to be correct, as did the model parameter estimates (e.g., loadings, errors, structural coefficients, etc.).

J. How should reviewer comments regarding interactions and/or quadratics be handled?

Reviewers occasionally ask, "are there any unmodeled significant interactions or quadratics?" This question is routinely asked in experimental studies analyzed with ANOVA. The procedure for answering this question is to specify all possible interactions and quadratics and perform an overall F test on any change in R^2 that results from adding them to the structural model. Unfortunately this cannot be done automatically as it is in ANOVA, and the procedure is discussed further in FAQ (G) above.

K. How are Latent Variable Cubics estimated?

There are other higher-order latent variables besides interactions and quadratics that may be important in model tests with survey data and I received the first request for specification of a cubic in 2003. When compared to interactions, related non-linear variables such as quadratics, XX and ZZ, and their cubic relatives, XXX and ZZZ, in

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XX + \beta_4 XZ + \beta_5 ZZ + \beta_6 XXX + \beta_7 ZZZ + \zeta_Y , \qquad (1$$

where β_1 through β_7 are unstandardized structural coefficients (also termed associations or, occasionally, effects), β_0 is an intercept, and ζ_Y is the estimation or prediction error, also termed the structural disturbance term, have received comparatively little methodological attention in survey research (however, see Aiken and West 1991). Perhaps as a result quadratics and cubics are seldom investigated in theoretical models involving survey data. However, they have been proposed and investigated in several social science literatures.

Specifying and estimating a cubic with Ordinary Least Squares (OLS) regression is easily accomplished when X, Z and Y are measured without error. Unfortunately when these variables are measured with error, the coefficient estimates from OLS regression in the above Equation 1 (i.e., the β's) will be biased in unknown directions, and they will be inefficient (i.e., they vary widely across replications) (Busemeyer and Jones 1983).

Latent Variable Interactions..., Chapter X and the working paper "Notes on Cubics, and Interactions and Quadratics in Latent Variables" on this web site discuss the specification, estimation and interpretation of latent variable cubics. While they both propose latent variable specifications of cubics involving latent variables, the working paper is more recent.

L. How is a "Second-Order" interaction estimated?

"Second-order" constructs were proposed by Jöreskog (1970). "Second-order" constructs are unobserved or latent variables that have other unobserved latent variables as their "indicators." Each of these "indicator" (first-order) latent variables has its respective observed indicators as usual.

These "Second-order" latent variables (LV's) have received attention recently, and "Second order" LV interactions have been of interest since at least 1997 (see Ping 1997). However, specifying and estimating these latent variables, although not difficult, is not a straightforward task using popular structural equation analysis such as LISREL, EQS, etc. In addition, there is little guidance for estimating an interaction involving a "Second-order" latent variable (e.g., XZ in

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \zeta_Y ,$$

where X or Z is a "Second-order" construct, β_1 through β_3 are unstandardized "regression" or structural coefficients [also termed associations or, occasionally, effects], β_0 is an intercept, and ζ_Y is the estimation or prediction error, also termed the structural disturbance term).

Chapter X in *Latent Variable Interactions...* on this web site discusses interactions involving "Second-order" constructs, specifically an interaction between a first-order latent variable and a "Second-order" latent variable. The working paper "Are their Second-Order Interactions? If So, How Are they Estimated?" also on this web site, also addresses "Second-order" interactions. Of the two the working paper is more recent.

REFERENCES

Aiken, Leona S. and Stephen G. West (1991), *Multiple Regression: Testing and Interpreting Interactions*, Newbury Park, CA: SAGE Publications.

Algina, James and Bradley C. Moulder (2001), "A Note on Estimating the Jöreskog-Yang Model for Latent Variable Interaction Using LISREL 8.3," Structural Equation Modeling, 8 (1) 40-52.

Anderson, James C. and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 103 (May), 411-23.

Bagozzi, Richard P. and Todd F. Heatherton (1994), "A General Approach to Representing Multifaceted Personality Constructs: Application to State Self-Esteem," *Structural Equation Modeling , 1 (1), 35-67.*

Bohrnstedt, G. W. and G. Marwell (1978), "The Reliability of Products of Two Random Variables," in Sociological Methodology, K. F. Schuessler ed., San Francisco: Jossy Bass, 254-273.

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.

Bollen, Kenneth A. (1995), "Structural Equation Models that are Nonlinear in Latent Variables: A Least Squares Estimator," *Sociological Methodology*, 25, 223-251.

Browne, Michael W. and Robert Cudeck (1993), "Alternative Ways of Assessing Model Fit," in *Testing Structural Equation Models*, K. A. Bollen et al. eds., Newbury Park CA: SAGE Publications.

Busemeyer, Jerome R. and Lawrence E. Jones (1983), "Analysis of Multiplicative Combination Rules When the Causal Variables are Measured With Error," *Psychological Bulletin*, 93 (May), 549-62.

Charlesworth, M. J. (1956), "Aristotle's Razor," *Philosophical Studies* (Ireland), 6: 105-112.

Cortina, Jose M. Gilad Chen and William P. Dunlap (2001), "Testing Interaction Effects in LISREL: Examination and Illustration of Available Procedures," Organizational Research Methods, 4 (4), 324-360.

Cudeck, Robert (1989), "Analysis of Correlation Matrices Using Covariance Structure Models," Psychological Bulletin, 105(2), 317-327.

DeVellis, Robert F. (1991), *Scale Development: Theory and Applications*, Newbury Park, CA: SAGE Publications.

Fornell, Claes and David F. Larker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.

Gerbing, David W. and James C. Anderson (1993), "Monte Carlo Evaluations of Goodness-of-Fit Indices for Structural

Equation Models," in *Testing Structural Equation Models*, K. A. Bollen and J. S. Long, eds., Newbury Park, CA: SAGE Publications.
Hayduk, Leslie A. (1987), *Structural Equation Modeling with LISREL: Essential and Advances*, Baltimore, MD: Johns Hopkins Press.
Howard, John A.(1989), *Consumer Behavior in Marketing Strategy*, Englewood Cliffs, NJ: Prentice Hall.
Jaccard, James and C. K. Wan (1995), "Measurement Error in the Analysis of Interaction Effects Between Continuous Predictors Using Multiple Regression: Multiple Indicator and Structural Equation Approaches," *Psychological Bulletin*, 117 (2), 348-357.
James, Lawrence R. (1980), "The Unmeasured Variables Problem in Path Analysis," *Journal of Applied Psychology*, 65 (4), 415-421.
Jöreskog, Karl G. (1970), "A General Method for Analysis of Covariance Structures," *Biometrika*, 57, 239-251.
Jöreskog, Karl G. (1993), "Testing Structural Equation Models," in *Testing Structural Equation Models*, Kenneth A. Bollen and J. Scott Long eds., Newbury Park, CA: SAGE.
Jöreskog, Karl G. and Fan Yang (1996), "Nonlinear Structural Equation Models: The Kenny and Judd Model with Interaction Effects," *Advances in Structural Equation Modeling Techniques*, G.A. Marcoulides, R.E. Schumacker, eds., Hillsdale, NJ: LEA.
Jöreskog, Karl G. and Dag Sörbom (1996), *Lisrel 8 User's Reference Guide*, Chicago: Scientific Software International, Inc.
Jöreskog, Karl G. (2000), "Latent Variable Scores and Their Uses," (on-line paper), http://www.ssicentral.com/lisrel/techdocs/lvscores.pdf.
Kenny, David and C. M. Judd (1984), "Estimating the Nonlinear and Interactive Effects of Latent Variables," *Psychological Bulletin*, 96, 201-10.
Klein, A. G. and H. Moosbrugger (2000), "Maximum Likelihood Estimation of Latent Interaction Effects with the LMS Method," *Psychometrika*, 65, 457-474.
Klein, A. G. and B. O. Muthén (2002), "Quasi Maximum Likelihood Estimation of Structural Equation Models with Multiple Interactions and Quadratic Effects," Unpublished ms., Graduate School of Education, UCLA.
Li, Fuzhong, Peter Harmer, Terry E. Duncan, Susan C. Duncan, Alan Acock and Shawn Boles (1998), "Approaches to Testing Interaction Effects Using Structural Equation Modeling Methodology," *Multivariate Behavioral Research*, 33 (1), 1-39.
Lubinski, D. and Humphreys, L.G. (1990), "Assessing Spurious Moderator Effects: Illustrated Substantively with the Hypothesized ("Synergistic") Relation Between Spatial and Mathematical Ability," *Psychological Bulletin*, 107, 385-393.
Marsh, Herbert W., Zhonglin Wen and Kit-Tai Hau (2004), "Structural Equation Models of Latent Interactions: Evaluation of Alternative Estimation Strategies and Indicator Construction," *Psychological Methods*, 9 (3), 275-300.
Mathieu, J. E., S. I. Tannenbaum and E. Salas (1992), "Influences of Individual and Situational Characteristics on Measuring of Training Effectiveness," *Academy of Management Journal*, 35, 828-847.
McClelland, G. H. and C. M. Judd (1993), "Statistical Difficulties of Detecting Interactions and Moderator Effects," *Psychological Bulletin*, 114 (2), 376-390.
Molder, Bradley C. and James Algina (2002), "Comparison of Methods for Estimating and Testing Latent Variable Interactions," *Structural Equation Modeling*, 9 (1), 1-19.
Netemeyer, Richard G., Mark W. Johnson and Scot Burton (1990), "Analysis of Role Conflict and Role Ambiguity in a Structural Equations Framework," *Journal of Applied Psychology*, 75 (2), 148-157.
Nunnally, Jum C. (1978), *Psychometric Theory*, 2nd Ed., New York: McGraw-Hill.
Ping, R. (1995), "A Parsimonious Estimating Technique for Interaction and Quadratic Latent Variables," *The Journal of Marketing Research*, 32 (August), 336-347.
Ping, R. (1996a), "Latent Variable Interaction and Quadratic Effect Estimation: A Two-Step Technique Using Structural Equation Analysis," *Psychological Bulletin*, 119 (January), 166-175.
Ping, R. (1996b), "Improving the Detection of Interactions in Selling and Sales Management Research," *Journal of Personal Selling and Sales Management*, 16 (Winter), 53-64.
Ping, R. (1996c), "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients," *Multivariate Behavioral Research*, 31 (1), 95-120.
Ping, R. (1997), "Voice in Business-to-Business Relationships: Cost-of-Exit and Demographic Antecedents," *Journal of Retailing*, 73 (Summer), 261-281.

Ping, R. (1999b), "Latent Variable Interactions: Substantive Considerations," *1999 Winter American Marketing Association Educators' Conference Proceedings*, Chicago: American Marketing Association.
Ping, R. (2001), "A Suggested Standard Error for Interaction Coefficients in Latent Variable Regression," *2001 Academy of Marketing Science Conference Proceedings*, Miami: Academy of Marketing Science.
Schermelleh-Engle, K., A. Kein and H. Moosbrugger (1998), "Estimating Nonlinear Effects using a Latent Moderated Structural Equations Approach," in *Interaction and Nonlinear Effects in Structural Equation Modeling*, R. E, Schumacker and G. A. Marcoulides, eds., Mahwah, NJ: Erlbaum.
Schumacker, Randall E. and George A. Marcoulides (1998), *Interaction and Nonlinear Effects in Structural Equation Modeling*, Mahwah, NJ: Erlbaum.
Schumacker, Randall E. (2002), "Latent Variable Interaction Modeling," *Structural Equation Modeling*, 9 (1), 40-54.
Wall, M. M. and Y. Amemiya (2001), "Generalized Appended Product Indicator Procedure for Nonlinear Structural Equation Analysis," *Journal of Educational and Behavioral Statistics*, 26, 1-29.
Wong, S. K. and J. S. Long (1987), "Reparameterizing Nonlinear Constraints in Models with Latent Variables," Technical Report, Washington State University.

TMPLTJMR.xls

The excel spreadsheet file titled “TMPLTJMR.xls” has been added to the metadata record for “Theoretical Model Testing with Latent Variables” as an additional file.

Please download this additional file to access the original spreadsheet.

EXCEL Template for Computing Starting or Fixed Values for Latent Variable (LV) Interactions and Quadratics Using a Single Indicator Interaction Specification

The APA citation for this paper is Ping, R.A. (2017). "EXCEL template for computing starting or fixed values for latent variable (LV) interactions and quadratics using a single indicator interaction specification." [on-line paper]. http://www.wright.edu/~robert.ping/jmr(1).doc .

(Note a previous version of this paper is available at www.wright.edu/~robert.ping/jmr.doc .)

The EXCEL spreadsheet is intended to assist in the specification of the single interaction/quadratic indicators, x:z, x:x and z:z (i.e., Ping 1995, 1996 single indicators). It uses measurement model parameter estimates for the loadings, measurement error variances, the variances associated with the latent variables X and Z, and the correlations between X and Z; along with the SAS, SPSS, etc. covariances among X, Z, XZ, XX and ZZ. The spreadsheet assumes that X and Z are unidimensional, preferably consistent (i.e., their measurement models fit the data) with mean-centered indicators, and it assumes that there are no correlated measurement errors involving X or Z.

To use this spreadsheet, plan to estimate a measurement model containing at least X and Z (a larger or a full model measurement model could be used as long as the model latent variables are all unidimensional). If starting values for all the phi's (covariances) associated with XX, XZ and ZZ are desired, also obtain SAS, SPSS, etc. estimates of the covariance matrix for X, Z, XX, XZ and ZZ.

The spreadsheet may be a bit confusing/vague about 1) “averaging,” about 2) the “phi’s” on the spreadsheet and the “Error Attenuated Cov’s” on the spreadsheet, and 3) “two-steps.” I prefer “averaging” the XX, XZ and ZZ loadings even though Ping 1995 and 1996 describe unaveraged results—the spreadsheet assumes averaging because otherwise the model covariance matrix can produce determinants that are “too large” for some computers.

Regarding 2), the “Error Attenuated Cov's” on the spreadsheet are optional unless “phi” starting values for XX, XZ or ZZ are desired. (I usually use phi starting values—they are formally correct for multivariate normal X and Z and they seem to speed up estimation. Note that X=(x1+...+xn)/n and Z=(z1+...+zm)/m are used to create the “Error Attenuated Cov’s” for averaging.)

About 3): Ping 1996 suggested estimating XX, XZ, etc. in “two steps.” Step 1 is to use the spreadsheet’s (fixed) loadings, error variances, etc. for XX, XZ, etc. from the measurement model for X and Z. Step 2 is to use fixed loadings, measurement error variances, etc. for XX, XZ from the structural model estimated using the spreadsheet in a second structural model. The loadings, error variances, etc. for XX, XZ, etc. will be different between the two structural models (due to non normality), and the second structural model results should be reported, unless all the model’s t-values are practically unchanged between the two structural model estimations (e.g., different in the second decimal place or less). (Experience suggests a third estimation or more is usually never required, unless there are errors in the XX, XZ, etc. specifications. If more than two structural models are necessary to obtain “convergence” in model estimates, please email me for suggestions.)

The bold entries, and the italicized entries, on the spreadsheet should be deleted (to avoid contamination with the example data). The result should be error messages or zeroes in most of the non-blank areas of the spreadsheet that should correct themselves once new data is entered. (Note much of the spreadsheet is “protected” to ensure the calculations, etc. cannot be erased.)

Next, the measurement model loadings, measurement error variances, and variances for X and Z should be entered (or copied and pasted) into the appropriate locations on the spreadsheet (i.e., loadings go in the "lambda" lines, measurement error variances go in the "theta" lines, and measurement model variances/covariances for X and Z go in the "phi" matrix). (Note there are locations for 10 loadings and measurment errors—the example has fewer than 10 loadings and measurment errors with blanks because the measures had 5 and 4 items. These entries will all appear on the EXCEL spreadsheet in bold font as like in the example—again, the unbolded cells are unrelated to entering data). At this point the loadings (lambda) and measurement error variances (theta) for XX, XZ and ZZ will be available near the bottom of the spreadsheet, along with reliabilities and AVE’s for X, Z, XX, XZ, .and ZZ.

If starting “phi” values for X, Z, XX, XZ, and ZZ are desired, also enter the SAS, SPSS, etc. covariances for X, Z, XX, XZ, and ZZ in the "Error Attenuated Cov's" matrix (again, don’t forget to average X=(x1+...+xn)/n and Z=(z1+...+zm)/m). These entries will all appear on the EXCEL spreadsheet in italicized font as they did in the example values. Once this is accomplished the rest of the covariances in the "phi" matrix should be nonzero.

Obviously deleting old data is important when using this spreadsheet, and it is probably a good idea to always delete the "Error Attenuated Cov's" entries even if starting values for the balance of the "phi's" are not desired to avoid accidentally using incorrect phi’s later.

When the spreadsheet has downloaded, it can be saved for repeated later use (i.e., without going back on line). Thus, it is possible to save a “master” copy of the on-line version of this EXCEL spreadsheet locally for modification, subsequent calculations, saving modified copies, etc. For example, I use the spreadsheet to calculate X and Z reliabilities and AVE’s even for models with no XZ, etc. (I keep track of what model variables X and Z represent in the saved file name--e.g., “XisSAT_YisALt.xls, etc.)

At the risk of overdoing it, I have one more word(s) about Latent Variable (LV) Interaction and Quadratic validity (a disinterested reader could skip to the bottom of the text). Authors in the Social Sciences disagree on what constitutes an adequate demonstration of validity. Nevertheless, a minimal demonstration of the validity of any LV should probably include the content or face validity of its indicators (how well they tap into the conceptual definition of the second-order construct), the LV's construct validity, and its convergent and discriminant validity (e.g., Bollen, 1989; DeVellis, 1991; Nunnally, 1993). The "validity" of this LV would then be qualitatively assessed considering its reliability and its performance over this minimal set of validity criteria.

Construct validity is concerned in part with an LV's correspondence or correlation with other LV's. The other LV's in the study should be valid and reliable, then their correlations with the target LV (e.g., significance, direction and magnitude) should be theoretically sound. Convergent and discriminant validity are Campbell and Fiske's (1959) proposals involving the measurement of multiple constructs with multiple methods, and they are frequently considered to be additional facets of construct validity. Convergent measures are highly correspondent (e.g., correlated) across different methods. Discriminant measures are internally convergent. However, convergent and discriminant validity are frequently not assessed in substantive articles as Campbell and Fiske (1959) intended (i.e., using multiple traits and multiple methods). Perhaps because constructs are frequently measured with a single method (i.e., the study at hand), reliability is frequently substituted for convergent validity, and LV correlational distinctness (e.g., the target LV's correlations with other measures are less than about 0.7) is substituted for discriminant validity.

However, LV reliability is a measure of the correspondence between the items and their LV, the correlation between an LV and its items, and "correlations less than 0.7" ignores measurement error. Fornell and Larker (1981) suggested that adequately convergent LV's should have measures that contain more than 50% explained or common variance in the factor analytic sense (less than 50% error variance, also see Dillon and Goldstein 1984), and they proposed a statistic they termed Average Variance Extracted (AVE) as measure of convergent validity. AVE is a measure of the shared or common variance in an LV, the amount of variance that is captured by the LV in relation to the amount of variance due to its measurement error (Dillon and Goldstein 1984). In different terms, AVE is a measure of the error-free variance of a set of items (AVE and its computation are discussed in detail elsewhere on this web site).

AVE can also be used to gauge discriminant validity (Fornell and Larker 1981). If the *squared* (error-disattenuated or structural equation model) correlation between two LV's is less than either of their individual AVE's, this suggests the LV's each have more internal (extracted) variance than variance shared between the LV's. If this is true for the target LV and all the other LV's, this suggests the discriminant validity of the target LV.

Unfortunately, experience suggests that AVE in LV Interactions and Quadratics is typically low, frequently less than 50%. For example, while they are not below 50% see the lower LV Interaction and Quadratic AVE's in the EXCEL Spreadsheet example when compared to high AVE’s of X and Z. Thus, to judge the validity of an LV Interaction or Quadratic, first it must be acceptably reliable (validity assumes reliability). Content or face validity is usually assumed unless fewer than all the indicators of the constituent variables are used to itemize the LV Interaction or Quadratic. Construct or correlational validity is usually difficult to judge, and it might be ignored. Convergent validity (AVE) should be 0.50 or above (the LV Interaction or Quadratic should be composed of 50% or less error) and it should be discriminant valid with the other model LV's, except perhaps its constituent variables (X or Z) (i.e., it is empirically distinct from the other model LV's--its AVE is larger than the squared correlations of the other LV's). In summary, while there are no hard and fast rules, reliability, and content, convergent and discriminant validity are probably sufficient to suggest the validity of an LV Interaction or Quadratic. Reliability, and content and convergent validity would be necessary, and construct (correlational) validity is usually ignored. With an AVE near 0.50 an LV Interaction or Quadratic *might* be argued to be empirically indistinct from 5-10% of the other model LV's by chance (depending on reviewers). More than that would suggest the LV Interaction or Quadratic is discriminant invalid, and its validity is impugned.

Experience suggests the substantive effect of the typically low AVE's in LV Interactions and Quadratics is their structural coefficients and their significances vary widely across replications. Specifically, with an AVE near 0.50 an hypothesized interaction or quadratic can be significant in one study but nonsignificant in a replication or near-replication. As a result, replication of a model test with hypothesized interactions or quadratics becomes comparatively more important. Specifically, an hypothesized interaction or quadratic that is NS in a model test could be significant in a replication, or vice versa.

For an LV Interaction or Quadratic with an AVE below 0.50, the alternatives besides ignoring AVE and hoping reviewers do likewise are to improve AVE in the LV Interaction or Quadratic. Low AVE in XZ is caused by low correlation between X and Z and/or comparatively large measurement errors in the items of X and or Z (i.e., low X and/or Z reliability). (Please see www.wright.edu/~robert.ping/ImprovXZ_AVEa.doc for more on improving XZ and XX reliability and validity.)

REFERENCES

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.

Campbell, Donald T. and Donald W. Fiske (1959), "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin*, 56, 81-105.

DeVellis, Robert F. (1991), *Scale Development: Theory and Applications*, Newbury Park, CA: SAGE Publications.

Dillon, William R. and Matthew Goldstein (1984), *Multivariate Analysis: Methods and Applications*, New York: Wiley.

Fornell, Claes and David F. Larker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.

Nunnally, Jum C. (1993), *Psychometric Theory*, 3rd Edition, New York, NY: McGraw-Hill.

Ping, R. A. (1995) “A Parsimonious Estimating Technique for Interaction and Quadratic Latent Variables,” *Journal of Marketing Research*, 32 (August), 336-347.

Ping, R. A. (1996), “Latent Variable Interaction and Quadratic Effect Estimation: A Two-step Technique Using Structural Equation Analysis,” *Psychological Bulletin*, 119 (January), 166-175.

Cubic.xls

The excel spreadsheet file titled “Cubic.xls” has been added to the metadata record for “Theoretical Model Testing with Latent Variables” as an additional file.

Please download this additional file to access the original spreadsheet.

weeding1.xls

The excel spreadsheet file titled “weeding1.xls” has been added to the metadata record for “Theoretical Model Testing with Latent Variables” as an additional file.

Please download this additional file to access the original spreadsheet.

EXCEL Template for Obtaining Internally Consistent Subsets of Items

This EXCEL template is intended to help delete items from a measure to produce two or more sets of items that "fit the data" (are internally consistent).

The APA citation for these instructions is Ping, R.A. (2006). "More about the template for obtaining an internally consistent set of items." [on-line paper]. http://www.wright.edu/~robert.ping/weeding.doc.

New measures will almost never "fit the data" using a single construct measurement model without dropping items to attain model-to-data fit. In addition, most well established measures developed before covariant structure analysis (LISREL, AMOS, etc.) became popular also will not fit the data without item weeding.

It turns out that measures used with covariant structure analysis are limited to about six items (see discussions in Anderson and Gerbing 1984, Gerbing and Anderson 1993, Bagozzi and Heatherton 1994, and Ping 2008). One explanation is that correlated measurement errors, ubiquitous in survey data but customarily not specified in covariant structure analysis, eventually overwhelm model-to-data fit in single-construct and full measurement models as indicators are added to the specification of a construct. And, that usually happens with about 6 items per construct.

There are ways around item weeding, such as various item aggregation techniques (see Bagozzi and Heatherton 1994), but many reviewers in the Social Sciences do not like these approaches. Unfortunately, reviewers also may not like dropping items from measures because of concerns over face- or content validity (how well the items "tap" the conceptual and operational definitions of their target construct). One "compromise" is to show the full measure's items in the paper, and assuming the full measure does not fit a single construct measurement model, show one submeasure that does fit the data and is maximally "equivalent" to the full measure in face or content validity. However, to do that, several submeasures are usually required, and finding even one is frequently a tedious task.

This template will assist in finding at least two subsets of items from the target measure that fit the data in a single construct measurement model of the items. The process is as follows. First, exploratory (common) factor analyze the target measure with its items using Maximum Likelihood estimation and varimax rotation. If the measure is multidimensional, start with the Factor 1 items. The other factors and the full measure can be used later.

Next, estimate a single construct (confirmatory) measurement model using the Factor 1 items (if the measure is unidimensional Factor 1 is the full measure). If the first measurement model fits the data item omission is not required. If this measurement model does not fit the data, find the "First Order Derivatives" in the output. (I will assume LISREL 8, which requires "all" on the OU line to produce First Order Derivatives. As far as I know, most other estimation packages produce statistics equivalent to First Order Derivatives. For example in SIMPLIS “First Order Derivatives” are available by adding the line “LISREL Output: FD.”). Paste the lower triangle of First Order Derivatives for "THETA-EPS" into the template making sure you retain the item names so you can figure out which item to drop (see the example on the template). Then find the largest value in the "Overall Sum" column--it will be the same as the "Max =" value in the lower right corner of the matrix.

Now, reestimate the measurement model with the item having the largest "Overall Sum" omitted (call this Reestimation 1). Record the Chi Square and RMSEA values on the spreadsheet for reference. If they are acceptable, use the items in this measurement model as submeasure 1.

There is no agreement on acceptable single construct measurement model fit. I use either a Chi Square that is slightly nonzero for single construct measurement models (e.g., 1E-07, not 0), or an RMSEA that is .08 or slightly below, but many authors would suggest much stronger fit criteria for single construct measurement models.[1]

If the unomitted items do not fit the data, find the "First Order Derivatives" for "Theta-Eps" in the Reestimation 1 output. Paste these into the second matrix in the template, record the Chi-Square and RMSEA values, and reestimate the single construct measurement model (Reestimation 2).

Repeating this process, eventually Chi Square will become nonzero, and after that RMSEA will decline to 0.08 or less (the recommended minimum for fit in full measurement and structural models--see Brown and Cudeck 1993, Jöreskog 1993). This should happen with about 7 or 8, down to about 5, remaining items. If acceptable fit does not happen by about 4 items, an error has probably been made, usually by omitting the wrong item.

Each subset after Chi Square becomes non zero is a candidate subset for "best," but because items are disappearing with each step, these smaller subsets are usually less face valid, and thus the first acceptable subset is usually the preferred one.

To find another subset of items, repeat the above process using "Modification Indices" for "Theta Epsilon." (The SIMPLIS command line is “LISREL Output: MI.”) The theory behind Modification Indices is different from First Derivatives, and a different subset usually results.

Another subset of items usually can be found using reliability. The reliability of all the Factor 1 items is computed using SAS, SPSS, etc., the item that contributes least to reliability is deleted, and the reliability of the remaining items is computed. This process is continued until deleting any item reduces reliability. The remaining items usually will fit the data in a single construct measurement model.

If the full measure was multidimensional, there may be several more subsets found by repeating the above procedures using the full measure's items instead of the Factor 1 items, then using the reliability procedure just mentioned. Experience suggests these subsets are smaller, but they frequently include items from Factor 2, etc. and thus they may be more face valid. This process can also be used on any Factor 2 items, Factor 3, etc.

There are many more subsets that can be found by omitting the next largest "Overall Sum" item instead of the "Max =" item. Specifically, the second largest item in Reestimation 1 could be omitted in place of the largest. Then, continuing as before omitting the largest "Overall Sum" items, The result is frequently a different subset of items that fits the data. Another subset can usually be found using this "Second Largest" approach using modification indices instead of first derivatives. Others can be found omitting the second largest overall sum item in Reestimation 2, instead of Reestimation 1, etc., with or without deleting the second largest in Reestimation 1. This "Second Largest" strategy can also be used on the full set of items.

Experience suggests that there are about N-things-taken-6-at-a-time combinations of items with real world data that will fit the data, where N is the number of items in the full measure (more, if 5, 4 and 3 item subsets are counted). For example, if the original measure has 8 items, with real world data there are about 8!(8-6)!/6! = 112 6-tem subsets of items that might fit the data. While the above strategies will not find all of them, experience suggests they should identify several two subsets that are usually attractive because they are comparatively large (again however, usually with about 6 items) and they should appear to tap the target construct comparatively well.

The above spreadsheet approaches may not always identify the highest reliability subsets of items, but experience suggests the resulting subsets are usually larger and as, or more, face valid than those produced by other approaches. However, with low reliability measures, even though the "First Derivative" or "Modification Indices" subsets should be only a few points lower in reliability than a subset found by, for example, dropping items that contribute lest to reliability, the higher reliability subset may be preferred to a higher face validity subset.

It may be instructive to (re)submit all the subsets found to an item-judging panel for their selection of the "best" subset for each construct.

Other comments: There are exceptions to several of the assertions made above, but this is probably not the place for an exhaustive exposition on item deletion strategies. For emphasis, the template assumes lower triangular matrices. There is an additional example in Appendix E of the monograph, *Testing Latent Variable Models*..., on the web site.

REFERENCES

Anderson, James C. and David W. Gerbing (1984), "The Effect of Sampling Error on Convergence, Improper Solutions, and Goodness of Fit Indices for Maximum Likelihood Confirmatory Factor Analysis," *Psychometrika*, 49, 155-73.

Bagozzi, Richard P. and Todd F. Heatherton (1994), "A General Approach to Representing Multifaceted Personality Constructs: Application to Self Esteem," *Structural Equation Modeling*, 1 (1), 35-67.

Browne, Michael W. and Robert Cudeck (1993), "Alternative Ways of Assessing Model Fit," in *Testing Structural Equation Models*, K. A. Bollen et al. eds, Newbury Park CA: SAGE Publications.

Gerbing, David W. and James C. Anderson (1993), "Monte Carlo Evaluations of Goodness-of-Fit Indices for Structural Equation Models," in *Testing Structural Equation Models*, K. A. Bollen and J. S. Long, eds., Newbury Park, CA: SAGE Publications.

Jöreskog, Karl G. (1993), "Testing Structural Equation Models," in *Testing Structural Equation Models*, Kenneth A. Bollen and J. Scott Long eds., Newbury Park, CA: SAGE.

_______ (2004), "On Assuring Valid Measures for Theoretical Models Using Survey Data," *Journal of Business Research*, 57 (2), 125-41.

ENDNOTES

[1] In my opinion, some authors go too far in real world data with single construct measurement model fit, resulting in unnecessarily small submeasures. There are several issues here, including model fit versus face or content validity, and experience suggests that with real-world data, "barely fits" in single construct measurement models is almost

always sufficient to attain full measurement model fit. Thus, in real world data, subsets of items that each produce a comparatively small but nonzero Chi Square or an RMSEA that is just below .08 are usually "consistent enough" to later produce a full measurement model that fits the data. I prefer the RMSEA criterion because it seems to produce fewer problems later. Again, however, many authors would not agree with this strategy. Later, if it turns out that the full measurement model does not adequately fit the data, simply estimate the next item weeding single construct measurement model and drop the next largest "Overall Sum" items to improve full measurement model fit.

TMPLTPB.xls

The excel spreadsheet file titled “TMPLTPB.xls” has been added to the metadata record for “Theoretical Model Testing with Latent Variables” as an additional file.

Please download this additional file to access the original spreadsheet.

This EXCEL spreadsheet is intended to assist with the specification of Kenny and Judd (1984) product indicators for the interactions/quadratics, XX, XZ and ZZ, using measurement model parameter estimates for the loadings, measurement error variances and variances associated with the latent variables X and Z, and SAS, SPSS, etc. covariances among X, Z, XZ, XX znd ZZ. The spreadsheet assumes that X and Z are unidimensional (i.e., consistent-- they fit the data) with mean centered indicators, and that there are no correlated measurement errors involving X or Z.

To use the spreadsheet, estimate a measurement model containing at least X and Z (a larger or a full model measurement model could be used as long as all the model latent variables are unidmensional). If starting values for the PHI's associated with XX, XZ and ZZ are desired, also obtain SAS, SPSS, etc. estimates of the covariance matrix for X, Z, XX, ZX and XX, in that order (i.e., if starting values for XX, XZ and/or ZZ are not required, skip the SAS, SPSS, etc. estimates). Next, the bold entries and the italicized entries on the spreadsheet should be deleted to avoid mixing old data with new data, and the result should be error messages or zeroes in most of the non-blank areas of the spreadsheet (that should correct themselves once new data is entered). Then the measurement model loadings, measurement error variances, and variances for X and Z should be entered into the appropriate locations on the spreadsheet (i.e., loadings go in the "lambda" lines, measurement error variances go in the "theta" lines, and measurement model variances/covariances for X and Z go in the "Phi" matrix). These entries will all appear in bold font-- unbolded cells are unrelated to entering measurement model parameter estimates). At this point the loadings (lambda) and measurement error variances (theta) for XX, XZ and ZZ will be available near the bottom of the spreadsheet.

If starting values for for the PHI matrix are desired, also enter the SAS, SPSS, etc. covariances for X, Z, XX, XZ, and ZZ in the "Error Attenuated Cov's" matrix. These entries will all appear on the EXCEL spreadsheet in italicized font. Once this is accomplished the rest of the covariances in the "Phi" matrix should be nonzero.

Obviously deleting old data is important to using this spreadsheet, and it is probably a good idea to always delete the "Error Attenuated Cov's" entries even if starting values for the balance of the "Phi's" are not desired.

For emphasis, when this spreadsheet (and the others) are visible on a local computer, it can be saved on that computer for later use (i.e., without going back on line). Thus, it is possible to save a copy of the on line version of the EXCEL spreadsheet locally to be used as a "master copy" for modification, subsequent calculations, saving modified copies, etc.

Tmplmvbr.xls

The excel spreadsheet file titled “Tmplmvbr.xls” has been added to the metadata record for “Theoretical Model Testing with Latent Variables” as an additional file.

Please download this additional file to access the original spreadsheet.

EXCEL Template for Computing the (Measurement Error) Adjusted Covariance Matrix for Latent Variable Regression

(The APA citation for this paper is Ping, R. A. (2008). "EXCEL Template for Computing the (Measurement Error) Adjusted Covariance Matrix for Latent Variable Regression." [on-line paper]. http://www.wright.edu/~rping/ItemWeed.doc) .

This EXCEL spreadsheet adjusts a covariance matrix from SAS, SPSS, etc. involving the latent variable Y, a set of up to 5 other latent variables, A through E, and, optionally, all possible interactions and quadratics involving A through E (i.e., AA, BB, AB, CC, AC, BC, DD, AD, BD, CD, EE, AE, BE, CE, DE) for use in error-adjusted-OLS regression ("latent variable" regression). This may seem like a step backward in structural equation analysis, but there are situations involving latent variables where LISREL, EQS, AMOS, etc. are difficult to impossible to use and (error-adjusted) OLS regression is helpful (e.g., model building where OLS regression's forward selection and backward selection are useful, latent variable models with one or more categorical variables, etc.).

The adjustment uses measurement model parameter estimates for the loadings, measurement error variances and variances associated with the latent variables Y, and A through E. The spreadsheet assumes that Y, and A through E are internally consistent (each of their single construct measurement models fit the data), they have mean centered indicators, and that there are no correlated measurement errors involving any of the latent variables A through E.

To use the spreadsheet, a (full) measurement model containing Y and up to five latent variables of interest should be estimated. Next, the bold entries and the italicized entries on the spreadsheet should be deleted to avoid mixing old data with new data, and the result should be zeroes in most of the non- blank areas of the spreadsheet (these values should correct themselves once new data is entered). Then, the covariance matrix to be adjusted should be created using SAS, SPSS, etc. and the variables of interest. Note that this covariance matrix should be created with Y, the dependent/endogenous variable named first. Next, the measurement model loadings, measurement error variances, and variances for Y and the variables of interest should be entered into the appropriate locations on the spreadsheet (i.e., loadings go in the "lambda" lines, measurement error variances go in the "theta" lines, and measurement model variances/covariances for X and Z go in the "Phi" matrix). These entries will all appear in bold font--un-bolded cells are unrelated to entering measurement model parameter estimates). At this point the adjusted covariance matrix will be available beneath the covariance matrix to be adjusted in the middle of the spreadsheet.

Several comments may be of interest. Obviously deleting old data is important in using this spreadsheet. For emphasis, when this spreadsheet (and the others) are visible on a local computer, it can be saved on that computer for later use (i.e., without going back on line). Thus, it is possible to save a copy of the on line version of the EXCEL spreadsheet locally to be used as a "master copy" for modification, subsequent calculations, saving modified copies, etc. The data that appears in the website version of this spreadsheet is also shown in a reordered form in Tables AE1 and AE2 of the monograph *INTERACTIONS AND QUADRATICS IN SURVEY DATA: A SOURCE BOOK FOR THEORETICAL MODEL TESTING (2nd Edition)*, on this web site. Several entries in

Table AE2 are slightly different from the spreadsheet "Adjusted Covariance Matrix..." entries (e.g., Var(SxA) which is Var(AB) in the "Adjusted Covariance Matrix..." of the spreadsheet) for unknown reasons (possibly transcription errors from the spreadsheet to the Table AE2 matrix-- however, the Table AE2 matrix was used to create the latent variable regression results shown in Tables E, G and H, not the spreadsheet).

NOTES ON “USED DATA”--
REUSING A DATA SET TO CREATE
A SECOND THEORY-TEST PAPER

Ping R.A. (2013). "Notes on 'Used Data.'" Am. Mktng. Assoc. (Summer) Educators' Conf. Proc.

ABSTRACT

There is no published guidance for using the same data set in more than one theory-test paper. Reusing data may reduce the “time-to-publication” for a second paper and conserve funds as the “clock ticks” for an untenured faculty member. Anecdotally however, there are reviewers who may reject a theory-test paper that admits to reusing data. The paper critically discusses this matter, and provides suggestions.

INTRODUCTION

Anecdotally, there is confusion among Ph.D. students about whether or not the same data set ought to be used in more than one theory-test paper. Some believe that data should be used in only one such paper. Others believe that data may be reused.

In a small and informal survey of journal editors, none was found to be opposed to reusing data, even when their journals’ “instructions to the writers” stated or implied that the study, and presumably its data, should be original.

In an anecdote from this survey, an editor summarized his experience with a paper that used data from a previous article. One reviewer rejected the paper because the data was not “original,” while the other reviewers saw no difficulty with a paper that relied on “used data.” This anecdote hints there also may be confusion about used data among some reviewers, and, since they are likely authors, presumably among some authors.

In a small pretest of a study of faculty at Research 1 universities who had Ph.D. students, none could recall the topic of reusing data in theory tests ever being discussed.

Because the consequences of any such confusion might include that the diffusion of knowledge may be impeded (e,g., an important study could be delayed, or go unpublished, because the author(s) had difficulty funding a second study), the paper critically discusses the reuse of data in theory tests, and provides suggestions. Along the way, several matters are raised for possible future discussion and pursuit.

USED DATA

"Used data" is ubiquitous. Secondary data from, for example, the US Census Bureau, and the Bureau of Labor Statistics, are in use almost everywhere. The advantages of (re)using this data include reduced costs and time. But data collected by governments/non-governmental-organizations/commercial firms may not be ideal for a theory test. (It tends to be descriptive, and multi-item measures typical in theory tests may be unavailable; raw secondary data may be difficult to obtain; or it may not measure all the variables that are important to the researcher.)

This paper will focus on the initial reuse of primary data; typically with formative/reflective (multi item) measures intended or used for theory testing. Theory-testing situations that might be judged to involve the initial reuse(s) of data include creating two or more papers based on a single data set gathered by the author(s). Other situations include creating a paper based on data that was previously collected for commercial purposes. (Anecdotally, in Europe, Ph.D. candidates' dissertation data may have been gathered and used by a "sponsoring company" for the company's commercial purposes that are unrelated to the dissertation.) They also include reanalyzing a published data set for illustrative or pedagogical purposes (typically for a suggested methodology), and reanalyzing a paper's data to further understand or "probe" a result observed in the paper. Less obviously, improving measure psychometrics (e.g., deleting measure items to improve reliability), and model-building also involve reusing data.

The advantages and disadvantages of reusing data are discussed next. Then, suggestions for theory testing are provided, and avenues for future research are sketched.

ADVANTAGES OF RESUSING A THEORY-TEST DATA SET

One advantage of reusing data is that it can reduce the elapsed time between theory generation and analysis, the resources required for data gathering (e.g., costs), and in some cases (e.g., data gathered by others) the expertise required to gather data. For example, in a model with several variables, after a paper that tests hypothesized links among (exogenous) model antecedents and their (endogenous) consequences, more papers in which the antecedents (or the consequences) are themselves linked, might be theoretically interesting enough for submission without gathering additional data. (Criteria for "theoretically interesting" might include new theory that either extends, or fills a gap in, extant theory.)

Reusing data may enable the division of a large paper into two or more papers, in order to satisfy a journal's page limit. For example, in a model with multiple final endogenous (consequence) variables, these variables might be divided into two sets of consequence variables (with their antecedents), and thus two papers, one for each resulting model. In each paper, this might reduce the number of hypotheses and their justifications, and the discussion and implications sections.

Stated differently, it might mean that an important study would not be delayed, or go unpublished, because of paper size, or difficulty funding an additional study.

Other advantages of reusing data might include:

- "Piggy backing" a theory test onto a commercial survey. This and using data already gathered by a commercial firm also may save time and costs.

- o Combining two surveys into a single survey. Unrelated surveys may not be easily combined, but, for example, when two models have some of the same latent variables, time and money might be conserved.
- o Publication of a dissertation with changes. (These changes should be based on additional theory, such as an additional path(s), that was developed prior to any data analysis beyond that for the dissertation. Stated differently, the logic of science (e.g., Hunt 1983) permits empirical discovery, hypothesis, then testing; but testing must be conducted using different data from that used in empirical discovery—see Kerr 1998 (I thank a reviewer for this citation)).
- o The use of secondary data.

Although it is now less popular that it was, meta analysis (e.g., Glass 1976) uses previously gathered data. In addition, methodologists and others also have used previously published data sets to illustrate a suggested methodology (e.g., Jöreskog and Sörbom 1996, and Bentler 2006).

Reuse of a paper's data includes estimating associations "Post Hoc"--after the model has been estimated (see Friedrich 1982)--to further understand or explain an observed association(s). It also includes reanalysis of the paper's data to illustrate different model assumptions. (For example, Ping 2007 reported results with and without Organizational Commitment in the proposed model for discussion purposes.)

Reusing data also enables psychometric improvement of measures. Measure items are routinely deleted serially with measure (or model) reestimation to improve reliability and facets of validity (e.g., average extracted variance—see Fornell and Larker 1981). This might be argued to be reuse of the data set (i.e., data snooping) to find the "best" itemization of a measure.

DISADVANTAGES OF RESUSING A THEORY-TEST DATA SET

Reusing data to produce more "hits" may not be viewed others as a worthy endeavor. Absent a compelling explanation such as reducing paper size, or sharpening the focus of a paper (e.g., a previous paper was on the antecedent-consequences links, and the next paper is about the links among the consequences), a reviewer (or reader) might judge data reuse as opportunism rather than "proper" science.

A second paper that, for example, replaces correlations in a previously published model's antecedents with paths, may be judged conceptually too similar to the first paper for publication. Thus, instead of conserving time, time may be wasted on a second paper that experiences rejections because of its insufficient contribution beyond the first paper.

Further, papers that are variations on a single model, and that reuse not only data but theory/hypotheses, measures, and methods, and share some results that are identical to a previous paper could be judged idioplagaristic. As a result, time and effort may be lost in rewriting to perceptually separate papers that use the same data set.

Care must be taken in how a model is divided into submodels. For example, omitting one or more significant exogenous variables in a model may bias the path coefficients of an endogenous variable to which they are linked (i.e., the "missing variable problem"--James 1980). And, it is easy to show that omitting one or more dependent variables in a model may change model fit, and thus standard errors and model paths' significance.

"Piggy backing" onto commercial survey (or using commercial data) may save time and costs, but an academic researcher may have difficulty controlling some of the project. For example, overall questionnaire design and its testing may not be under the control of the academic researcher. Similarly the sampling frame, sampling, and activities to increase response rates also may not be under the direction of the academic researcher. Further, the appearance of an academic researcher's "independence" from the survey "issues" (i.e., the researcher is not "up to something") may be lost by not using university letterhead or return address. (Or arguably worse: using university letterhead and return address to collect data that also will be analyzed by a commercial firm). Finally, having someone else "doing some of the work" can deprive a researcher of valuable experience in data gathering. (This could be an important disadvantage: for a dissertation, demonstrating data gathering expertise is typically required.)

Last, a questionnaire that combines several surveys may be too large for its respondents: it may increase their fatigue, and it may produce echeloning, respondent irritation over similarly worded items, etc., that can increase response errors, and produce low response rates.

DISCUSSION

It may not be apparent that a model might contain candidate submodels for additional papers. Several examples might help suggest a framework for finding candidate submodels.

Finding Submodels

In Figure 1, a disguised (but actual) theoretical latent variable model (Model 1), the blank (fixed at zero) paths (e.g., A2 -> A3) could be freed to help produce submodels. To improve readability, several Model 1 latent variables were rearranged, and exogenous (antecedent) latent variables (those without an antecedent) were relabeled "A" (see Figure 3). Terminal (endogenous) consequences (latent variables that are not antecedents) were relabeled "TC," and intermediate (endogenous) latent variables were relabeled "E."

Next, each blank (fixed at zero) path was considered for being freed, then in which direction it might be freed. Then, several of these new paths were discarded because they were theoretically implausible, of little interest theoretically, or directionality could not be established (bidirectional/non recursive paths were not considered). Next, several A's were relabeled as E's.

The results included Model 1 and the (full) Figure 3 model, plus several submodels involving the A's and E's that were judged interesting enough for possible submission. For example, a submodel involving E5, and the other E's and A's (to avoid missing variable problems—A4, for example is an indirect antecedent of E5) (Submodel 1) was judged to have submission potential (E5 was judged to be an important consequence) (see Figure 4). (Submodel 1 could be abbreviated E5 = f (E4, E6, E7, Ei, Ea, Eb, A2, A4 | i = 1-3, paths among E's free as shown in Figure 3, paths among Ea, Eb, A2 and A4 free as shown in Figure 3), where "f" denotes "function of, as shown in Figure 4" and "|" means "where.")

A "hierarchy of effects" (serial) respecification of Figure 3 also was considered. Specifically, a second-order latent variable S1 was specified using Ea, A2, Eb and A4 (see Figure 2, and see Jöreskog 1971). Similarly, second-order latent variables S2 and S3 were specified using E1-E7 (see Figure 2), and the proposed sequence S1, S2, S3 then TC was specified. (Experience suggests that a second-order latent variable can be useful to combine, and thus simplify, latent variables in a model (e.g., Dwyer and Oh 1987)).

Similarly, there was an interesting submodel involving Eb (Eb = f (Ea, A2, A4)) (not shown, but see Figure 3), and another interesting submodel involving E1-E3 (Submodel 2) ({Ei} = f (A2, A4, Ea, Eb | i = 1-3, paths among A2, A4, Ea and Eb free as shown in Figure 3, paths among Ei free as shown in Figure 3), where "{}" means "set of ") (not shown, but see Figure 3). In summary, several models were found, each having a "focal consequence" latent variable(s) that was judged to be important enough to have submission potential.

Figure 6 shows a different disguised theoretical latent variable model (Model 2) where antecedent (exogenous) latent variables have been labeled “A,” and terminal consequences (latent variables that are not antecedents) have been labeled “TC.” In Figure 7, Model 2 was rearranged for clarity, bolded paths were added to replace the originally blank (fixed at zero) paths in Model 2, and intermediate latent variables were (re)labeled E (Model 3). Because much of the theory and many of the measures in Model 2 were new, the first paper (with Figure 6’s Model 2 and no bolded paths) was too large for journal acceptance. As a result, TC3 (itself an interesting focal variable) was excised for placement in a second paper (i.e., TC3 = f (A3, Ei | i = 1-7, all paths among A3 and Ei fixed at zero) (not shown, but see Figure 7). An additional model with the focal variable E2 = f (A3, E1, E3 | bolded paths among A3, and E1 and E3 free as shown in Figure 7) (Submodel 3) was judged interesting enough for journal submission (A3 is an indirect antecedent of E2 and is specified to avoid the missing variable problem) (not shown, but see Figure 7). Another interesting model was discovered, with the bolded Figure 7 paths among E4-E7 (with A3 and E1-E3 without their bolded paths, and without TC3), that was judged to be a “hierarchy of effects” (sequential) model (i.e., first E4, next E5 or E7, then E6, then E7) (Submodel 4) (not shown, but see Figure 7).

An additional model with a theoretically plausible and interesting non-recursive (bi-directional) path between E6 and E7 (see Figure 5, and see Bagozzi 1980) also was discovered using Figure 7. (A non-recursive model that was identified—see for example Dillon and Goldstein 1984, p.447—was not immediately obvious. At least two variables were required for identification of the bi-directional path between E6 and E7: one that should significantly affect E6 but should not be linked to E7, and another that should significantly affect E7 but should not be linked to E6. Because nearly all the Figure 7 latent variables were theoretically linked to both

E6 and E7 (and could not be omitted without risking the missing variable problem), theoretically plausible demographic variables D1 and D2 were added to attain identification). Finally, a comparison of the Figure 7 model's estimates for males versus those for females was considered.

In summary, after rearranging and re-labeling the Figure 6 latent variables for clarity, previously fixed but theoretically plausible paths were freed. Then, interesting focal variables were found and submodels with as many of the Figure 6 variables as antecedents as possible (to avoid the missing variable problem) were estimated (to determine if the results were still "interesting"). In addition, the Figure 7 model was found to contain a hierarchy of effects submodel, and at least one of the paths was plausibly non-recursive. Finally, the Figure 6 model was estimated for males, then reestimated for females, and the results were compared.

Experience suggests that models with many variables may contain "interesting" submodels. Models with several "intermediate" variables (e.g., Figure 3), and those with multiple antecedents or several terminal consequences (e.g., Figure 7) also are likely to contain interesting submodels. As the examples suggested, in addition to "single consequence" submodels, linked antecedent and linked consequence submodels (e.g., Figure 7), second order, hierarchy-of-effects and non-recursive submodels are possible. Comparing model results for categories of a demographic(s) variable also might produce interesting results.

Irregularities

Unfortunately, data reuse may provide opportunities for "irregularities." For example, combining two surveys into a single survey provides an opportunity to "data snoop" across surveys. While this might generate interesting theory, it also might result in a paper that "positions" exploratory research (data snooping, then theory/hypotheses, and then a theory disconfirmation test using the data-snooped data) as confirmatory research (theory/hypotheses prior to any data analysis involving these hypotheses, then disconfirmation).

Data reuse also may provide a temptation to "position" the results of post hoc analysis as though they were originally hypothesized. For example, care must be taken that paths discovered by post hoc data analysis (e.g., to explain an hypothesized but non-significant association) are not then hypothesized as though they were not the results of data snooping.

(Parenthetically, "data snooping" also might be acceptable using a split sample, or a simulated data set. With a split sample, half of the original data set might be used for data snooping, and the other half could be used to test any resulting hypotheses. Similarly, a simulated data set might be generated using the input item-covariance matrix from the original data set, then used for data snooping. Then, the original data set could be used to test any resulting hypotheses. In both cases, the additional hypotheses, and the split half or simulated data set procedure should be mentioned in the interest of full disclosure.

Improving Psychometrics

Viewing sequentially dropping items (item weeding) to improve measure psychometrics as reanalysis of a data set, thus reusing data, may require additional discussion. Item weeding is routinely done in structural equation analysis to improve internal and external consistency, and reliability and validity in measures. These activities have been criticized (e.g., Cattell 1973; Fornell and Yi 1992; Gerbing, Hamilton and Freeman 1994; Kumar and Dillon 1987a, 1987b), however these complaints did not involve data reuse, and these objections are now seldom heard.

Item weeding is (implicitly) justified as required to separate measurement from model structure (e.g., Anderson and Gerbing 1988). (Ideally it produces a compromise between measurement model "fit" and face validity). However, it is easy to show that in real-world data these efforts can reduce the standard errors of the structural model's path coefficients. Stated differently, item weeding could be viewed as data snooping to (perhaps inadvertently) weaken the desired disconfirmation test of a proposed model by finding itemizations that are more likely to improve the chances of "confirming" the model.

Alternatives to weeding are few. In real-world data, summing unweeded indicators may not be acceptable because the resulting measure may be unreliable. However, Gerbing and Anderson (1984) suggested in effect that deleted items could be specified as a second factor in a second-order latent variable (e.g., Jöreskog 1971). The software they suggested to expedite this task, ITAN (Gerbing and Hunter 1988) is no longer readily available, but experience suggests that in real-world data exploratory factor analysis could be used to create second-order latent variables from the "factors" (to likely reduce both the "data snooping," and to reduce the item deletions and thus improve measure face validity).

SUGGESTIONS FOR THEORY TESTING

Authors may want to be more aware of the opportunities attending data reuse. Even if they elect not to reuse their data for publication, finding submodels might be used as way to discover additional interesting research topics. Authors could then write a second paper on an interesting submodel while conducting a new data gathering activity to test that submodel. They also might estimate the submodel using the "old" data before the new data are available, to develop at least a framework for several sections of the new paper, including possibly the reliability and validity of the submodels' measures (these should be reconfirmed using the new data), and the results and discussion sections.

Once the new data are available, the second paper could be revised based on the new data. The used-data issue would be avoided, and time might be conserved by the parallel activities of writing a new paper while collecting data for its test.

However, given the risks that the new paper might be judged too similar to any previous paper, or it may be judged idioplagaristic, authors may elect to conserve time and funds by constructing a new paper based on the used data. In that event, the editor of any target journal probably should be contacted, to gauge their reaction to reusing data (there is the obvious matter of possibly compromising the double blind review process, even if the editor instructs the reviewers that the authors are not necessarily the same as before).

In addition, to anticipate any reviewer objections, authors should consider a "full disclosure" of the history of the data, and the paper. Specifically, any prior publication, such as publication of a previous paper involving the data, publication of the paper as an abstract, a conference paper, etc. probably should be noted to address any reviewer questions about the paper's relationship to any other published papers.

Any previous use of the data briefly should be described in the first submission of a paper that reuses data, to address any reviewer questions about the originality of the data given the sample appears to be identical to a previously published article(s). If reuse becomes an issue during review, additional details, previous paper descriptions, and assurances such as "analysis of the data for the present paper was conducted after theorizing," and "theorizing was not revised to fit the data," etc. could be provided. Further, any valid justifications, such as "the present paper is the result of pruning the prior paper to meet the page limitation," could be stated.

In addition, in a combined survey, it could be stated that extensive pretesting was conducted to reduce survey recipient fatigue; or in a study that piggybacked onto a commercial study, that the lead researcher was careful to maintain strict control of all phases of the study. Further, it could be stated that every effort was made to reduce idioplagarism, that care was taken in creating submodels to eliminate the missing variable problem, and that the model was tested with and without omitted consequent variable to estimate any bias due to model fit. (parenthetically, this "data history" also may be important after paper acceptance, so readers can gauge the acceptability of the paper for themselves).

Ideally, if data are to be reused, that decision should be made prior to any data gathering. Specifically, after the initial model is developed, any additional submodels and their hypotheses should be developed before any data are gathered. This should reduce any temptation to develop hypotheses then insert them in the original paper based on data snooping.

If the decision to reuse data is made after data has been gathered, all submodel(s) and their hypotheses should be developed before any submodel is estimated. Again, this may reduce any temptation to insert "data snooped" hypotheses in the same paper.

Addressing the matter of multiple papers with many of the same variables, and the same hypotheses for these variables, the same measures and sample, many of the same findings, etc. being judged too similar, or even idioplagaristic, may require effort. Similarity might be reduced by emphasizing that, although the new paper involves previously studied constructs, it provides important new theory about the relationships among them. For example, Submodel 1 in Figure 3 proposed previously unexplored antecedents (E4, E6 and E7) of an important variable (E5).

Reducing the appearance of idioplagarism may require writing a fresh paper, instead of rewording (or cutting and pasting), for example, the hypotheses justifications, and the descriptions of the measures, sampling, data gathering, the results, etc. of a prior paper.

Finally, if multiple papers using the same data set are jointly submitted for review, ideally each paper should acknowledge the existence of the other(s). A (brief) explanation of each could be provided, and copies might be placed on a commercial web site, for the reviewers.

Several comments may deserve emphasis: publishing similar versions of a paper, for example a conference version or an "earlier" version of a paper, could be argued to be idioplagarism. An alternative may be to consider publishing an abstract rather than a full paper. Similarly, submitting an unaltered or slightly altered paper to multiple outlets also could be viewed as idioplagaristic. (This is proscribed by many publication outlets. Typically it is discovered by having a common reviewer, and anecdotally, violation can be grounds for rejection, or desk rejection of any future submission.) One should resist the temptation to hide any reuse of data. (A reviewer who is familiar with any previous paper may question the originality of the data.)

At the risk of overdoing it, theory should always precede data analysis. Specifically, while hypotheses may be developed or revised using data, they should not be tested using the same data. (However, hypotheses developed after post hoc analysis of the data are appropriate for the paper's discussion or future research sections--with a caveat that these results may be an artifact of the present data set, and thus are exploratory and are in need of disconfirmation in a future study.)

FUTURE RESEARCH

It may be instructive to survey Ph.D. students, journal editors, and faculty for their attitudes about reusing data. If students have either no attitude, or a weakly held one, while some journal editors and reviewers do not object, this might suggest an additional publication strategy for untenured faculty "while the P&T clock ticks." (However, it is plausible that "top tier" journal editors and reviewers, when reviewing for these journals, might covertly object to reusing data—indeed a comment from a reviewer in the present venue hinted that they may object to reusing data.)

A similar study of these attitudes in the European Union also might be interesting. If Ph.D. students and others are encouraged, in effect, to seek a "sponsoring company" for their research (with the possibility that their academic research may become part of the sponsoring company's commercial research), this might suggest at the very least, topics for debate, if not avenues for research and publication.

SUMMARY

Because there is no published guidance concerning the use of the same data set in several theoretical model-test papers, and there may be confusion among Ph.D. students and reviewers about whether this is appropriate in theory tests, the paper critically discussed reused data in theoretical model tests, and provided suggestions.

Experience suggests that models are likely contain at least one submodel that might be a candidate for an additional paper. And, although it was anecdotal, some editors and reviewers had no objection to "used data" in theory tests. However, authors should be aware of the risks that attend used data in theory tests: reviewers may not approve of reusing data, and any subsequent paper based on used data may be judged conceptually too similar to the first paper for publication. Papers based on used data also may be judged idioplagaristic when compared to other papers to use the data. Further, care must be taken in specifying submodels to avoid the "missing variable" problem.

Suggestions for authors included that they may want to contact the editors of target journals to gauge the acceptability of a paper based on used data. And, that if data is to be reused, that decision ideally should be made prior to data collection, to reduce any temptation to add additional hypotheses to the paper based on “data snooping” the data once it was collected. And, if data are reused, authors should consider a “full disclosure” of the history of the data set.

REFERENCES

Anderson, James C. and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 103 (May), 411-23.

Bagozzi, R. P. (1980), "Performance and Satisfaction in an Industrial Sales Force: An Examination of their Antecedents and Simultaneity," *Journal of Marketing*, 44 (Spring), 65-77.

Bentler, Peter M (2006), *EQS 6 Structural Equations Program Manual*, Encino, CA: Multivariate Software, Inc.

Cattell, R. B. (1973), *Personality and Mood by Questionnaire*, San Francisco: Jossey-Bass.

Dillon, William R. and Matthew Goldstein (1984), *Multivariate Analysis, Methods and Applications*, New York: Wiley.

Dwyer, F. Robert and Sejo Oh (1987), "Output Sector Munificence Effects on the Internal Political Economy of Marketing Channels," *Journal of Marketing Research*, XXIV (November), 347-58.

Fornell, Claes and David F. Larker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.

Fornell, Claes and Youjae Yi (1992), "Assumptions of the Two-Step Approach to Latent Variable Modeling," *Sociological Methods and Research*, 20 (Spring), 291-320.

Friedrich, R. J. (1982), "In Defense of Multiplicative Terms in Multiple Regression Equations," *American Journal of Political Science*, 26, 297-833.

Gerbing, D.W. and J. C. Anderson (1984), "On the Meaning of Within-Factor Correlated Measurement Errors," *Journal of Consumer Research*, 11 (June), 572-50.

Gerbing, David W. and John E. Hunter (1988), *ITAN, A Statistical Package for Item Analysis*, (Available from D. W. Gerbing, School of Business, Portland State University, Portland, OR 97207).

Gerbing, David W., Janet G. Hamilton and Elizabeth B. Freeman (1994), "A Large-scale Second-order Structural Equation Model of the Influence of Management Participation on Organizational Planning Benefits," *Journal of Management*, 20, 859-85.

Glass G. V (1976), "Primary, Secondary, and Meta-Analysis of Research," *Educational Researcher*, 5 (10), 3–8.

Hunt, Shelby D. (1983), *Marketing Theory, The Philosophy of Marketing Science*, Homewood, IL: Irwin.

James, Lawrence R. (1980), "The Unmeasured Variables Problem in Path Analysis," *Journal of Applied Psychology*, 65 (4), 415-421.

Jöreskog, K. G. (1971), "Statistical Analysis of Sets of Congeneric Tests," *Psychometrika*, 36, 109-133.

_____ and D. Sörbom (1996), *LISREL 8 User's Reference Guide*, Chicago: Scientific Software International.

Kerr, Norbert L. (1998), "HARKing: Hypothesizing After the Results are Known," *Personality and Social Psychology Review*, 2 (3), 196-217.

Kumar, Ajith and William R. Dillon (1987a), "The Interaction of Measurement and Structure in Simultaneous Equation Models with Unobservable Variables," *Journal of Marketing Research*, XXIV (February), 98-105.

_____ (1987b), “Some Further Remarks on Measurement-Structure Interaction and the Unidimensionality of Constructs,” *Journal of Marketing Research,* XXIV (November), 438-444.

Ping Robert (2007), “Salesperson-Employer Relationships: Salesperson Responses to Relationship Problems and their Antecedents," Journal of Personal Selling and Sales Management, XXVII, 1 (Winter), 39-57.

Figure 1—Abbreviated Latent Variable Model (Model 1) (Disguised) (see p. 6)

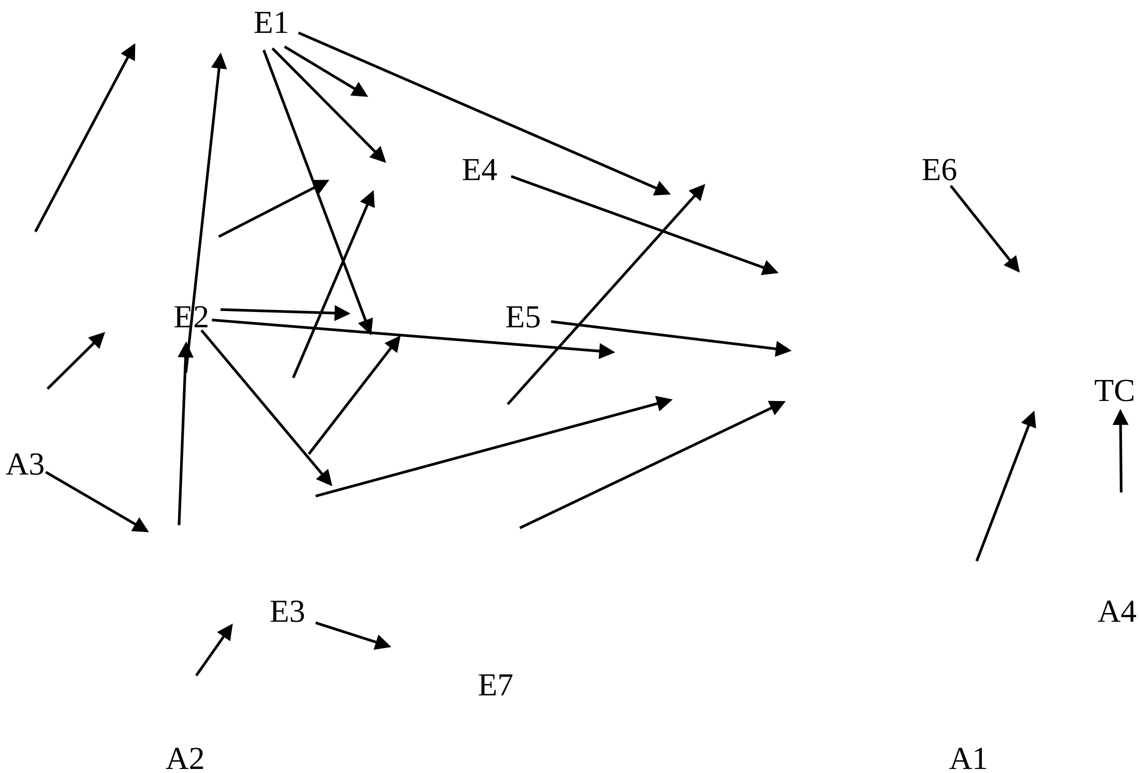


Figure 2—Respecified Figure 3 Model (see p. 7)

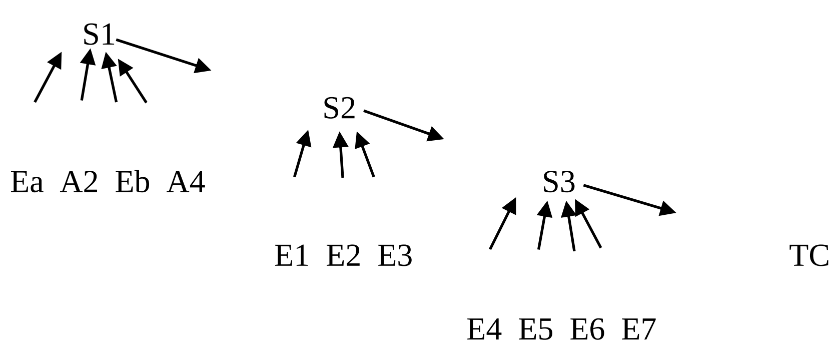


Figure 3—Rearranged Figure 1 Model with Plausible Additional Paths (in bold) (see p. 6)

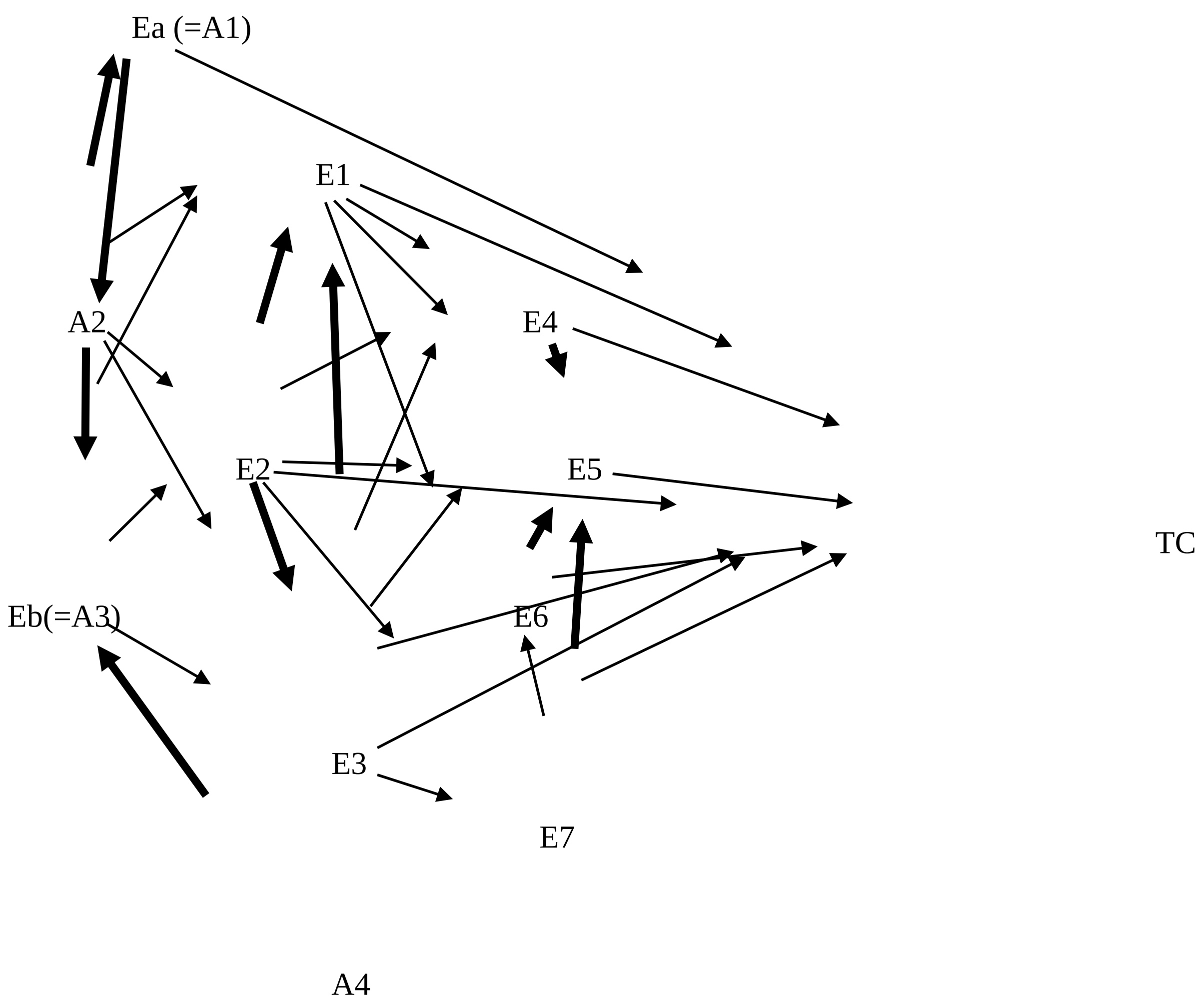


Figure 4—Submodel 1 (of Figure 3) (see p. 7)

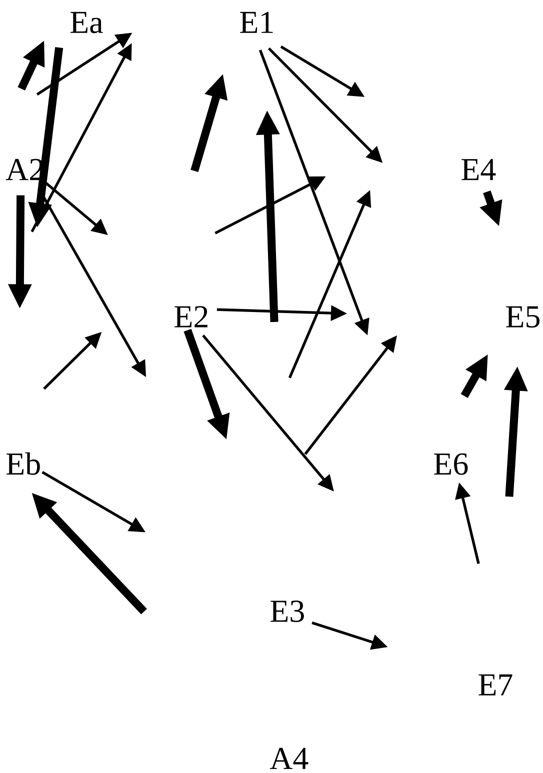


Figure 5—An Abbreviated Non-Recursive Respecification of Figure 7 (see p. 8)

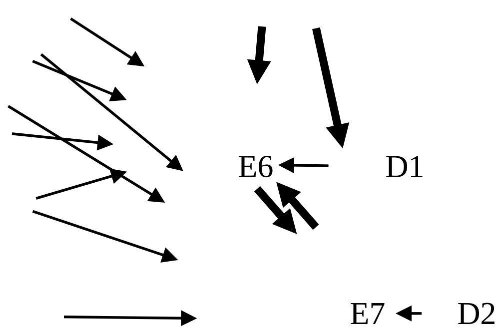


Figure 6—Abbreviated Latent Variable Model (Model 2) (Disguised) (see p. 8)

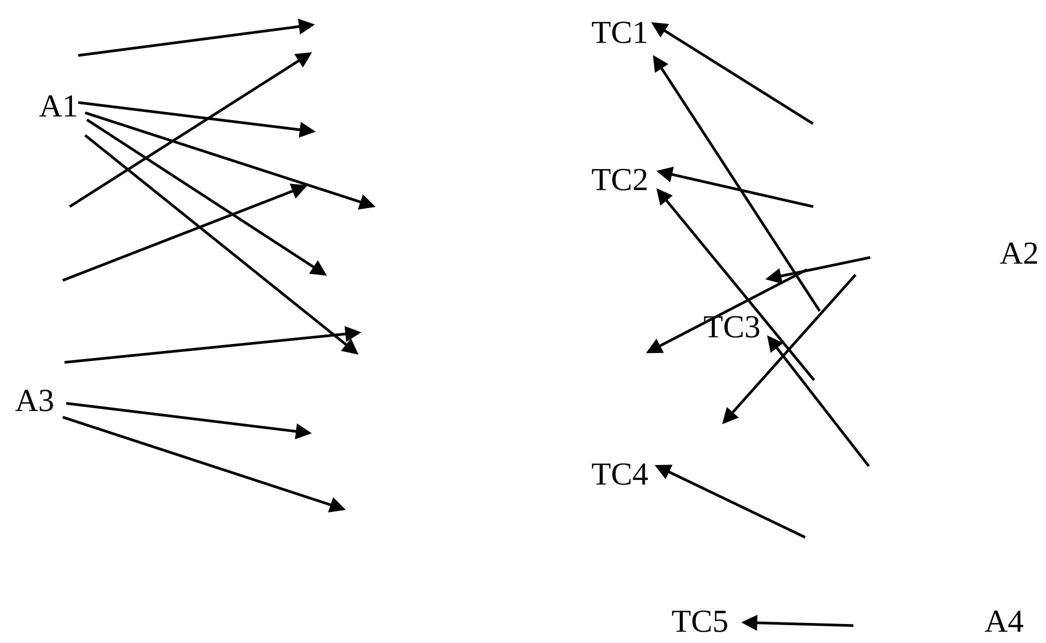
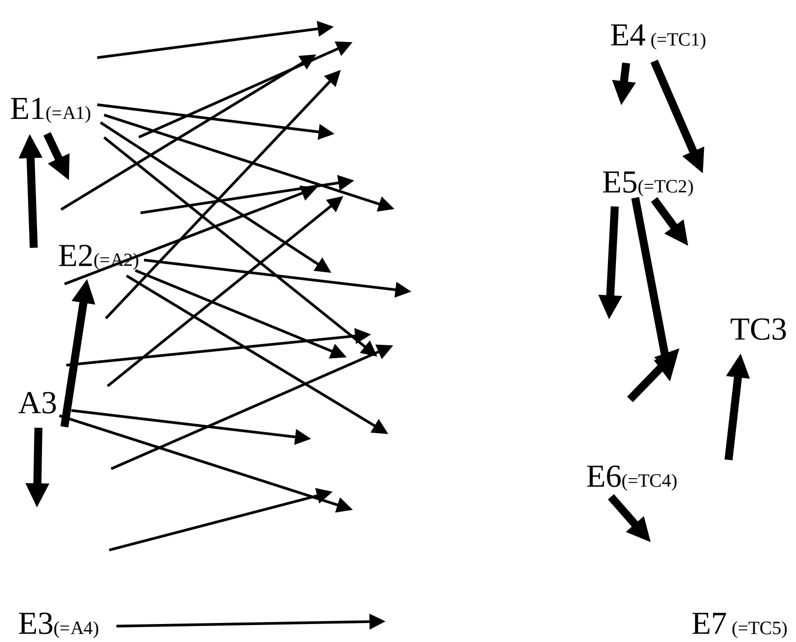


Figure 7—Rearranged Abbreviated Model 2 with Plausible Additional Paths (in bold) (see p. 8)



WHY ARE THE HYPOTHESIZED ASSOCIATIONS NOT SIGNIFICANT? A *THREE-WAY* INTERACTION?

Robert A. Ping, Wright State University, Dayton, OH

ABSTRACT

There is little guidance for estimating a latent variable (LV) "three-way" interaction (e.g., XZW). The paper explores these variables, and suggests their specification. It also provides a pedagogical example to suggest the utility of three-way interactions. Hypothesizing these LV's is discussed, their reliability is derived, a remedy for their nonessential ill-conditioning (their high correlations with X, Z and W) in real-world data is suggested, and an approach to interpreting them is illustrated.

INTRODUCTION

"Two-way" interactions in structural equation analysis (SEM) such as XZ, XW and ZW in

1) $$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 W + \beta_4 XZ + \beta_5 XW + \beta_6 ZW + \beta_7 XZW + \zeta_Y ,$$

where X, Z, W and Y are non-categorical variables, β_1 through β_7 are unstandardized structural coefficients, β_0 is an intercept (typically ignored in SEM), and ζ_Y is the structural disturbance (estimation error) term, have received considerable theoretical attention (see Aiken and West 1991). They also have been investigated with survey data in several substantive literatures (see Aiken and West 1991, p. 2; Bohrnstedt and Marwell 1978; Jaccard, Turissi and Wan 1990, p. 79; Lubinski and Humphreys 1990; and Podsakoff, Tudor, Grover and Huber 1984 for partial lists of citations).

However, non-categorical "three-way" interactions in survey data (e.g., XZW in Equation 1) have received little attention. They also have yet to appear in published SEM models, perhaps because there is little guidance for estimating them. This paper sheds additional light on these LV's and their estimation. Specifically, it discusses their specification, estimation and interpretation. Along the way their utility is illustrated, and a remedy for a property of these LV's in real world data that apparently is not well known, their nonessential ill-conditioning, is proposed. Hypothesizing a three-way interaction is discussed, their reliability is derived, and an approach to interpreting these LV's is illustrated.

To help motivate this topic, we will skip ahead to a pedagogical example. In studies of firms' Reactions to Dissatisfaction in Business-to-Business relationships, the relationship of the subject's Switching Costs (SC's) (costs to replace the primary supplier) with the subject's Opportunism (OPP) (guileful self-interest seeking) was observed to be non-significant (NS) in Ping (1993), and positive in Ping (2007). Similarly, the OPP association with the subject's Investment (INV) (expenditures to maintain the relationship) was NS in Ping (1993), and positive in Ping (2007). This suggested the possibility that INV and SC were being moderated (Ping 1996d).[1] Subsequently, it was judged plausible that INV moderated SC (argument omitted).[2] In a reanalysis of one of the above studies' data sets, however, INVxSC was not significant.

Another possibility was that Alternatives (ALT) (attractive replacement relationships) moderated an interaction between INV and SC.[3] Specifically, it was plausible that there was a three-way interaction among ALT, INV and SC: ALTxINVxSC. In the reanalysis data set ALTxINVxSC was significant.[4]

Next, we will discuss two-way interactions, which will lead to a proposed specification of a three-way interaction involving LV's, then the details of the above pedagogical example (that will illustrate their estimation and interpretation).

INTERACTIONS IN SURVEY MODELS

It will be important later to briefly discuss two-way interaction specification, in order to lay the groundwork for specifying XZW. There have been several proposals for specifying two-way LV interactions including (1) Kenny and Judd 1984; (2) Bollen 1995; (3) Jöreskog and Yang 1996; (4) Ping 1995; (5) Ping 1996a; (6) Ping 1996b; (7) Jaccard and Wan 1995; (8) Jöreskog 2000; (9) Wall and Amemiya 2001; (10) Mathieu, Tannenbaum and Salas 1992; (11) Algina and Moulder 2001; (12) Marsh, Wen and Hau 2004; (13) Klein and Moosbrugger 2000/Schermelleh-Engle, Kein and Moosbrugger 1998/Klein and Muthén 2002; and (14) Moulder and Algina 2002.

These proposed techniques are based on the Kenny and Judd (1984) product-of-indicators proposal (x_1z_1, x_1z_2, ... x_1z_m, x_2z_1, x_2z_2, ... x_2z_m, ... x_nz_m, where n and m are the number of indicators of X and Z respectively). However, in theoretical model tests using real world survey data, where models with several, usually over-determined, LV's (i.e., LV's with four or more indicators), are the rule, specifying XZ with all the Kenny and Judd product indicators typically produces model-to-data fit problems. Specifically, in

---

[1] A moderated variable's structural coefficient estimate occurs at the mean of the moderator variable in a study (see Aiken and West 1991). When a significant moderation is unspecified, variation in the moderator variable's mean across studies can produce wide variation in the moderated variable's structural coefficients.

[2] Details of the study will be omitted to sidestep matters that are unimportant to the methodological matters at hand. And, hereafter Y will be used instead of OPP.

[3] An additional possibility was that ALT moderated INV and it moderated SC. A three way interaction is possible in this case also—ALT may moderate INVxSC.

[4] In theory testing, such explorations are within the logic (science) of discovery (see Hunt 1983). Specifically, a "discovery," such as a three-way interaction, that can be theoretically supported is proposed in the Discussion section of the study at hand for disconfirmation in a subsequent study.

Techniques 1 and 5, the resulting specification of XZ in its single construct measurement model usually will not fit the data (i.e., this specification of XZ is inconsistent with the data), and full measurement and structural models containing this specification of XZ can exhibit unacceptable model-to-data fit.

Several proposals use subsets of the Kenny and Judd (1984) product indicators, or indicator aggregation, to avoid these inconsistency problems (Techniques 3, 5, 7, 9, 11, 12 and 14). Unfortunately, omitting Kenny and Judd product indicators raises questions about the face or content validity of the resulting interaction. Specifically, if all the indicators of X are not present in the itemization of XZ, is XZ still the "product of the LV X and the LV X"? (techniques 3, 7, 9, 11, 12 and 14). This specification has additional drawbacks: the reliability of XZ is unknown for a partially itemized XZ. As we shall see, the formula for the reliability of XZ is a function of X and Z with all their items (see Bohrnstedt and Marwell 1978). Further, a procedure for determining which product indicators to retain is unknown. And, deleting Kenny and Judd product indicators can produce interpretation problems because the X in XZ is no longer operationally the same as X in Equation 1, for example.

Some proposed techniques do not involve Maximum Likelihood estimation, or commercially available estimation software (Techniques 2, 6 and 13). And, several of these proposals have not been evaluated for possible bias and lack of efficiency (i.e., Techniques 8 and 10).

A SINGLE INDICATOR

The following will rely on the Ping (1995) proposal for specifying XZ because it has the fewest of the above drawbacks. This proposed specification uses a single indicator for XZ that is the product of sums of the indicators for X and Z. Specifically, for X with the indicators x_1 and x_2, and Z with indicators z_1 and z_2 the single indicator of XZ would be $xz = (x_1+x_2)(z_1+z_2)$. Ping (1995) suggested that under the Kenny and Judd (1984) normality assumptions,[5] a loading, λ_{xz}, and measurement error variance, $\theta\varepsilon_{xz}$, for this single indicator are

2) $\qquad \lambda_{xz} = \Lambda_X\Lambda_Z$,

and

2a) $\qquad \theta\varepsilon_{xz} = \Lambda_X{}^2\mathrm{Var}(X)\theta_Z + \Lambda_Z{}^2\mathrm{Var}(Z)\theta_X + \theta_X\theta_Z$,

where $\Lambda_X = \lambda_{x1} + \lambda_{x2}$, Var indicates error disattenuated variance, $\theta_X = \mathrm{Var}(\varepsilon_{x1}) + \mathrm{Var}(\varepsilon_{x2})$, ε_{x1} is the measurement error of x_1, ε_{x2} is the measurement error of x_2, $\Lambda_z = \lambda_{z1} + \lambda_{z2}$, $\theta_z = \mathrm{Var}(\varepsilon_{z1}) + \mathrm{Var}(\varepsilon_{z2})$, $\lambda_{xz} = \Lambda_X\Lambda_Z$, $\theta\varepsilon_{xz} = (\Lambda_X)^2\mathrm{Var}(X)\theta_Z + (\Lambda_Z)^2\mathrm{Var}(Z)\theta_X + \theta_X\theta_Z$, and λ

[5] X and Z are assumed to be independent of their measurement errors (ε_{x1}, ε_{x2}, ε_{z1}, and ε_{z2}), their measurement errors are mutually independent, the indicators x_1, x_2, z_1, and z_2, and the measurement errors (ε_{x1}, ε_{x2}, ε_{z1}, and ε_{z2}) are multivariate normal with mean zero.

and θ are loadings and measurement error variances. The indicators x_i and z_j are mean-centered by subtracting the mean of x_i, for example, from x_i in each case, and the single indicator of XZ, xz, becomes

$$x_c z_c = [\Sigma(x_i^u - \mathbf{M}(x_i^u))]\ [\Sigma(z_j^u - \mathbf{M}(z_j^u))]\ ,$$

where x_i^u and z_j^u are uncentered indicators (denoted by the superscript "u"), **M** denotes a mean, and Σ is a sum taken before any multiplication. Centering x_i and z_j not only helps provide simplified Equations 2) and 2a), it reduces the high correlation or nonessential ill-conditioning (Marquardt, 1980; see Aiken and West, 1991) of X and Z with XZ that produces unstable (inefficient) structural coefficient estimates that can vary widely across studies.

Using simulated data sets and data conditions that were representative of those encountered in surveys, Ping's (1995) results suggested that the proposed single indicator for an interaction produced unbiased and consistent coefficient estimates.

This single-indicator specification can be estimated in two steps. First, the data for the single indicator of XZ is created by computing the sum of the indicators of X times the sum of the indicators of Z in each case. Next, the measurement parameters in Equations 2 and 2a (i.e., λ_{x1}, λ_{x2}, etc., Var(ε_{x1}), etc., Var(X), etc.) are estimated in a measurement model (MM) that excludes XZ. Then, the loadings and measurement error variances for XZ's (λ_{xz} and θ_{xz}) are computed using equations 2 and 2a, and using these parameter estimates. Finally, specifying the calculated loadings and error variances λ_{xz} and θ_{xz} for the product indicator as fixed values, the structural model is estimated.

If the structural model estimates of the measurement parameters for X and Z (i.e., λ_{x1}, λ_{x2}, etc., Var(ε_{x1}), etc., Var(X), etc.) do not approximate those from the MM (i.e., equality in the first two decimal places) the loadings and error variances of the product indicator are recomputed using the structural model estimates of the equation 2 and 2a measurement parameters. Experience suggests that with consistent LV's zero to two of these iterations are sufficient to produce exact estimates (i.e., equal to "direct" estimates of XZ-- see Ping 1995).

THREE WAY INTERACTIONS

XZW also could be specified as the product of sums of indicators (e.g., xzw = $(\Sigma x_i)(\Sigma z_j)(\Sigma w_k)$). However, mean centering x_i, z_j and w_k does not reduce the multicollinearity (nonessential ill-conditioning--see Marquardt 1980, and Aiken and West 1991) between XZW and X, Z and W that typically occurs in real-world data.[6] Unfortunately, the bias from this multicollinearity can produce an apparently non-

[6] Authors seem to be unaware of this multicollinearity in real world data (e.g., Aiken and West 1991). (See for example, the uncomfortably high correlations between ALTxINVxSC, in several specifications, and ALT, INV and SC in Table A.)

significant (NS) three-way interaction. For example, in the pedagogical example ALTxINVxSC specified with mean-centered X, Z and W was NS.

A SPECIFICATION

An alternative specification that avoids multicollinearity bias is to use the indicator

3) $x_c(zw)_c = [\Sigma x_i^u$ **-** $\mathbf{M}(\Sigma x_i^u)][(\Sigma z_j^u)(\Sigma w_k^u)$ - $\mathbf{M}((\Sigma z_j^u)(\Sigma w_k^u))]$,

where x_i^u, z_j^u and w_k^u are uncentered indicators (denoted by the superscript "u"), **M** denotes a mean, and Σ is a sum taken before any multiplication, to spec the 3 way $X_c(Z_uxW_u)_c$.[7]

The loading and error variance of $x_c(zw)_c$ is derived in Appendix A..

AN EXAMPLE

Returning to the pedagogical example, after developing a plausible argument for ALT moderating the Y-INVxSC association, the three-way interaction $ALT_c(INV_uxSC_u)_c$ in

5) $Y = aSAT + bALT_c + cINV_u + dSC_u + gALT_c(INV_uxSC_u)_c + h(INV_uxSC_u)_c + \zeta$,

where a, b, etc. are structural coefficients, was specified in a measurement model for Equation 5 to gage each LV's psychometrics.[8]

$ALT_c(INV_uxSC_u)_c$ was specified by computing the Equation 3 single indicator, $alt_c(inv{\cdot}sc)_c = [\Sigma alt_i^u$ **-** $\mathbf{M}(\Sigma alt_i^u)][(\Sigma inv_j^u)(\Sigma sc_k^u)$ - $\mathbf{M}((\Sigma inv_j^u)(\Sigma sc_k^u))]$ in each case, then the loading and measurement error variance was computed using the Equations 4e and 4f (alt_i^u, inv_j^u and sc_k are the uncentered indicators of ALT, INC and SC respectively, and **M** denotes a mean).

---

[7] In the following, x, z, w, x_c, z_c, w_c, $(zw)_c$, x_i^u, z_j^u, w_k^u denote unstandardized variables.

[8] Some authors recommend including the two-way interactions, ALTxINV, ALTxSC and INVxSC, in Equation 5, unless theory suggests they are zero (e.g., Aiken and West 1991). However, theory typically suggests that associations should be non-zero, and in a theory test this is the usual rationale for including variables. Thus, because we could not develop plausible arguments for including ALTxINV or ALTxSC, Equation 5 was specified without them. Theory appeared to support including INVxSC, and, although it was previously observed to be non-significant, it was specified because $ALT_c(INV_uxSC_u)_c$ is the interaction of two (hypothesized) antecedents of Y, ALT and (INVxSC). (This matter is discussed later.)

Specifically, estimates of the loading and measurement error variance of $ALT_c(INV_uxSC_u)_c$ were computed using the loadings, variances, covariances and measurement error variances of ALT_c, INV_u and SC_u from earlier measurement models (MM's) without the interactions,[9] and SPSS estimates of the means.

Then, the measurement model for Equation 5 was estimated, using LISREL 8 and Maximum Likelihood estimation. The resulting loadings and measurement error variances of SAT, ALT, INV, SC and Y were sufficiently similar to those from previous MM's containing just SAT, ALT, INV, SC and Y, that a second Equation 5 measurement model estimation (to revise the computed interaction loadings and measurement error variances) was judged not necessary (see Ping 1996a). Because the Equation 5 measurement model fit the data, $ALT_c(INV_uxSC_u)_c$ was judged to be externally consistent. Finally, $ALT_c(INV_uxSC_u)_c$ was judged to be trivially internally consistent.

RELIABILITY AND VALIDITY

The reliability of $X_c(ZW)_c$ is unknown, and it is derived in Appendix B.

The reliability of $ALT_c(INV_uxSC_u)_c$ was computed to be 0.89. Specifically, SPSS values for the means of INV and SC, the square root of the MM variances for the standard deviations of INV and SC, the MM value for the correlation between INV and SC, SPSS reliabilities of ALT, INV and SC,[10] and the calculated error-disattenuated correlation of ALT_c and $(INVxSC)_c$ (see Ping 1996c) were substituted into Equation 6. Then, $ALT_c(INV_uxSC_u)_c$ was judged to be valid.[11]

---

[9] If ALT, INV and SC are internally and externally consistent their measurement parameters should not change materially with the addition of other latent variables, including two- and three-way interactions--see Ping 1996a.

[10] The latent variable reliabilities of ALT, INV and SC are practically equivalent to their coefficient alpha reliabilities (Anderson and Gerbing 1988).

[11] There is little agreement on validity criteria. A minimal demonstration of validity might include content or face validity (how well an LV's indicators tap its conceptual definition), construct validity (the target LV's correlations with other LV's are theoretically sound), convergent validity (e.g., its average extracted variance (Fornell and Larker 1981) is greater than .5), and discriminant validity (e.g., its correlations with other measures are less than some cutoff value) (e.g., Bollen 1989, DeVellis 1991, Fornell and Larker 1981, Nunnally 1978) (see Ping 2004a). $ALT_c(INV^*SC)_c$ was judged to be content valid because ALT, INV and SC were content valid, and the specification of $ALT_c(INV^*SC)_c$ included all the indicators of ALT, INV and SC. To gage convergent and discriminant validities, the formula for the Average Variance Extracted (AVE) of $X_c(ZW)_c$,

$$(\Sigma\lambda_{xi}^2)(\Sigma\lambda_{zj}^2)(\Sigma\lambda_{wk}^2)Var(XZW)/[(\Sigma\lambda_{xi}^2)(\Sigma\lambda_{zj}^2)(\Sigma\lambda_{wk}^2)Var(XZW) + \Theta_{xc}\Theta_{(zw)c}]$$

(Fornell and Larker 1981), where $(\Sigma\lambda_{xi}^2)$ is the sum of the squares of λ_{xi}, etc., Var(XZW)

Next, the structural model was estimated using LISREL 8 and Maximum Likelihood estimation, and the abbreviated results are shown in Table B. Then, the Y associations with ALT, INV and SC were interpreted to account for their moderation. First, the plausible moderation of INVxSC by ALT was trivially "confirmed"[12] by the significant $ALT_c(INV_u xSC_u)_c$ coefficient in Table B.) Then, the (now moderated) Y-ALT association was interpreted. Specifically, Equation 5 was "factored"[13] to produce the (full) structural coefficient ("simple slope"--see Aiken and West 1991) of ALT_c:

$$Y = aSAT + (b + g(INV_u xSC_u)_c)ALT_c + cINV_u + dSC_u + h(INV_u xSC_u)_c + \zeta$$

and the coefficient of ALT was interpreted using $(b' + g(INV_u xSC_u)_c)$--the results are shown in Table C.[14]

Next, Equation 5 was re-factored to produce the moderated coefficient of INV,

$$\begin{aligned} Y &= aSAT + bALT_c + cINV_u + dSC_u + gALT_c(INV_u xSC_u)_c + h(INV_u xSC_u)_c + \zeta \\ &= aSAT + bALT_c + cINV_u + dSC_u + gALT_c(INV_u xSC_u\text{-}m)_c + h(INV_u xSC_u\text{-}m)_c + \zeta \\ &= aSAT + bALT_c + dSC_u \\ &\qquad + cINV_u + gALT_c(SC_u\text{-}m/INV_u)_c INV_u + h(SC_u\text{-}m/INV_u)_c\, INV_u + \zeta \\ &= aSAT + bALT_c + cINV_u + dSC_u \\ &\qquad + (c + gALT_c(SC_u\text{-}m/INV_u)_c + h(SC_u\text{-}m/INV_u)_c)INV_u + \zeta\ , \end{aligned}$$

where INV is non zero and $m = \mathbf{M}((\Sigma z_j^u)(\Sigma w_k^u))$, and INV_u was interpreted (see Table D).

Finally, Equation 5 was again re-factored to produce a moderated coefficient of SC in,

$$Y = aSAT + bALT_c + cINV_u$$

---

is the variance of XZW (available in the structural model), and $\Theta_{xc}\Theta_{(zw)}$ is the measurement error variance of $x_c(zw)_c$ (see Equations 4a through 4d in Appendix A), was used, and $ALT_c(INV{*}SC)_c$ was judged to be convergent and discriminant valid using Fornell and Larker's criteria (see Fornell and Larker 1981). Finally, the construct (correlational) validity of $ALT_c(INV{*}SC)_c$ was impossible to judge.

[12] Specifically, the test failed to disconfirm the moderation. True "confirmation" would be suggested inductively by many such "confirmation" results.

[13] In this case the "factored" coefficient of ALT was the partial derivative of Y with respect to ALT_c. The term "factored" emphasizes that ALT is the same (i.e., centered and fully itemized) in all of its occurrences in Equation 5.

[14] Graphs also can be used for interaction interpretation (e.g., Aiken and West (1991)). However, the significances required for interpreting the originally hypothesized individual effects of ALT, INV and SC are not available using graphs.

$$+ (d + gALT_c(INV_u\text{-}m/SC_u) + h(INV_u\text{-}m/SC_u)_c)SC_u + \zeta$$

for interpretation, and the results (not reported) were similar to Table D.

DISCUSSION

Because these specifications of a three-way interaction have not been formally evaluated for possible bias and inconsistency, their threshold for significance proabably should be conservative (e.g., |t-value| > 2.10).

$ALT_cxINV_cxSC_{c-}$ was non significant while $ALT_c(INV_uxSC_u)_c$ was significant, as previously reported. In addition, the $ALT_c(INV_uxSC_u)_c$ structural coefficient was different from the $INV_c(ALT_uxSC_u)_c$ and $SC_c(INV_uxALT_u)_c$ structural coefficients (unstandardized beta=0.23, 0.01 and 0.29; t=2.64, 0.83 and 1.90 respectively). This suggests there are several specifications of a three-way interaction: an "all centered" specification ($ALT_cxINV_cxSC_c$), and three "permutation" three-way interaction specifications, $ALT_c(INV_uxSC_u)_c$, $INV_c(ALT_uxSC_u)_c$ and $SC_c(INV_uxALT_u)_c$. This, plus some authors' preference for including ALTxINV, ALTxSC and INVxSC, appears to suggest that a "proper" disconfirmation test of a three-way interaction should include all the relevant two-way interactions, ALTxINV, ALTxSC and INVxSC, plus the "all centered" three-way interactions, and $ALT_cxINV_cxSC_c$, plus the "permutation" three-ways interactions $ALT_c(INV_uxSC_u)_c$, $INV_c(ALT_uxSC_u)_c$ and $SC_c(INV_uxALT_u)_c$.

However, experience suggests that justifying all the relevant two-way interactions and all the "permutation" three-way interactions, may be difficult. Because in theory testing, variables probably should not be added to a model without theoretical justification, some two- and three-way interactions ways may not be candidates for the model.

In addition, "permutation" three-way interactions may be sufficiently correlated to suppress each other. For example, in the pedagogical example, the "permutation" three-ways were all non-significant when they were jointly specified.

Further, experience suggests that the two-way interactions corresponding to $ALT_cxINV_cxSC_c$—the "all centered" two-way interactions ALT_cxINV_c, ALT_cxSC_c and INV_cxSC_c—typically can not be specified jointly with the "permutation" three-way interactions. For example, in the pedagogical example, two of these two-way interactions were highly correlated with $ALT_c(INV_uxSC_u)_c$ (0.85 or higher) and INV_cxSC_c was uncomfortably, and negatively (-0.46), correlated with $ALT_c(INV_uxSC_u)_c$. Thus, in a three-way interaction specification, the "all centered" two-way interactions should be excluded if a "permutation" three-way interaction is specified.

This usually limits the specification alternatives for Equation 5 to one "permutation" three-way interaction with its corresponding two-way interaction, or the "all centered" three-way interaction, with its corresponding two-way interactions.

Nevertheless, $X_cZ_cW_c$ may be the preferred specification in theory testing. While it can be non-essentially ill-conditioned, (in which case a plausible "permutation" three-way interaction such as $X_c(Z_uxW_u)_c$ should be tested instead), this three-way interaction form is recommended in regression (Aiken and West 1991), all the two-way interactions can be included, and it also could be argued to provide the stronger disconfirmation test--it jointly tests all three moderation hypothesis, ALT moderates INVxSC, INV moderates ALTxSC and SC moderates ALTxINV. Parenthetically, all the two-way interactions probably should be included to improve detailed interpretation, unless they cannot be adequately theoretically justified. For example, in Equation 5 with all the two-way interactions and $X_cZ_cW_c$ instead of $X_c(Z_uW_u)_c$ the factored coefficient of X becomes

$$Y = a'SAT + b'X_c + c'Z_c + d'W_c + g'X_cxZ_cxW_c + hX_cxZ_c + iX_cxW_c + jZ_cxW_c + \zeta' ,$$

$$= a'SAT + c'Z_c + d'W_c + jZ_cxW_c + b'X_c + hX_cxZ_c + iX_cxW_c + g'X_cxZ_cxW_c + \zeta' ,$$

$$= a'SAT + c'Z_c + d'W_c + jZ_cxW_c + (b' + hZ_c + iW_c + g'Z_cxW_c)X_c + \zeta' ,$$

the significances of which could be at least slightly different from Equation 5 without the two-way interactions, even if the coefficients h and i are non-significant. (Note that X_u, Z_u and W_u have been replaced by X_u, Z_u and W_u respectively).

Thus, one approach might be to test an "all centered" three-way interaction first. Then, if the "all centered" three-way interaction is non-significant, this could be argued to imply that not all three-way moderations are significant. The next step might be to test one or more plausible "permutation" three-way interactions, as discussed above.

The loading, measurement error variance and reliability of an "all centered" three-way interaction are derived in Appendix C.

Spreadsheets for estimating the loading, measurement error variance and reliability for the "all centered" and "permutation" interaction specifications are available by e-mail.

The paper assumes that the three-way interaction was discovered (i.e., it was not hypothesized before the data was collected without having been previously discovered). In an informal, and so far incomplete, survey of articles, most articles replicate this assumption: a three-way interaction is discovered, or the hypothesized three-way interaction was discovered in a prior study. Thus, the above may be a guide to most three-way interaction situations.

However, there are several other plausible situations: an hypothesized three-way interaction that was discovered in a previous study, and two hypothesized two-way interactions that involve three first-order latent variables. In the second situation, a proper test would be to include the associated three-way interaction because the two-way interactions all could be conditional (i.e., moderated by a third latent variable). The procedure would be as follows:

1) Gage the reliability and validity, and the internal and external consistency, of X, Z and W, and the other model latent variables. For emphasis, all the model's latent variables must be internally consistent (i.e., X's single construct measurement model fits the data, Z's single construct measurement model fits the data, etc.); and the full measurement model with all the model's latent variables, including X, Z, W, XZ, ZW, XW and XZW should be shown to be externally consistent. Consistency is particularly important because the parameter estimates in the interaction specifications are assumed to be trivially different between the measurement model without XZW and the measurement and structural models with XZW (which consistency ensures).
2) Since an "all centered" specification of XZW, $X_cZ_cW_c$, where X_c, Z_c and W_c are mean- or zero-centered (e.g., the values of X_c, for example, in each case are equal to the case value x minus the mean of all the values of x), provides the strongest disconformation test--it can jointly test all three moderation hypotheses (X moderates ZW, Z moderates XW, and W moderates XZ)--$X_cZ_cW_c$ is specified and tested first. X_c, Z_c and W_c are zero-centered (e.g., the values of X_c, for example, in each case are equal to the case value x minus the mean of all the values of x). Then, $X_cZ_cW_c$ and the relevant two-way interactions, X_cZ_c, X_cW_c and Z_cW_c, are jointly specified. (The two-way X_cZ_c, for example, is required because $X_cZ_cW_c$ is the moderation of the X_cZ_c-Y association by W_c and both X_c and Z_cW_c should be present to avoid creating an additional missing variable problem for Y.)
3) Next, estimates of the loading and measurement error variance for $X_cZ_cW_c$, along with those for X_cZ_c, X_cW_c and Z_cW_c,[15] should be computed using the loadings, variances, covariances and measurement error variances of X_c, Z_c and W_c from their external consistency measurement models (MM's) without the interactions (see Footnote 8), and SPSS estimates of the means.
4) Then, the reliability and validity of $X_cZ_cW_c$, X_cZ_c, X_cW_c and Z_cW_c, should be verified, and the external consistency of the model with $X_cZ_cW_c$, X_cZ_c, X_cW_c and Z_cW_c should be gaged.
5) Next, the structural model should be estimated, and if the structural model parameter estimates for X, Z and W are not sufficiently similar to those from the external consistency measurement model, the structural model parameter estimates should be used to revise the computed interaction loadings and measurement error variances, and the structural model should be re-estimated.

In the first situation, estimating XZW should proceed using steps 1-5 above. Then,

6) If the "all centered" three-way interaction is non significant, the next step would be to test one or more permutation three ways (e.g., $X_c(ZW)_c$, etc.)

[15] X_cZ_c and X_cW_c were ignored in the pedagogical example because they were judged implausible.

Specifically, the estimates of the loading and measurement error variance for $X_c(ZW)_c$, for example, are computed using the loadings, variances, covariances and measurement error variances of X_c, and $(ZW)_c$ in the consistency measurement models without the interactions (see Footnote 8), and SPSS estimates of the means.

7) Next, the reliability and validity of $X_c(ZW)_c$ and $(Z_cW)_c$ should be verified, and the external consistency of the model with $X_c(ZW)_c$ and $(Z_cW)_c$ should be gauged.
8) Then, the structural model should be estimated, and if the structural model parameter estimates for X, Z and W are not sufficiently similar to those from the external consistency measurement model, the structural model paramater estimates should be used to revise the computed interaction loadings and measurement error variances, and the structural model should be re-estimated.

In Step 2 X_cZ_c, X_cW_c and Z_cW_c, were specified for completeness. However, in theory testing, variables probably should not be added to a model without theoretical justification. Thus, any two-way and three-way interactions should be theoretically justified (argued to be plausible). Some of the two-way interactions were omitted in the example because they could not all be theoretically justified. However, any two-way interaction(s) omitted limits the test to those interaction that are present. For example, if X_cZ_c for example is omitted, $X_cZ_cW_c$ no longer tests the moderation of X_cZ_c by W_c.

Because the ALTxINVxSC interaction was significant, an ALTxINVxSC-->Y hypothesis for a subsequent study might be, "ALT moderates the investments-switching costs interaction." On the other hand, it might be argued that INVxSC jointly amplifies the ALT-->Y association (see Table C—at low INV and SC ALT was non-significant, while at higher INV and SC, ALT was significant). This latter hypothesis (and its analog which was not observed in the present study, ALT is negative/positive when INVxSC is low, but positive/negative when it is high--which would apply if the Table C structural coefficients changed sign) in effect treats INVxSC as a single variable (which of course it is). Choosing among these alternative hypotheses might depend on which is easier to argue.

For emphasis, in the pedagogical example a significant three-way interaction was discovered after the data was gathered. Thus, any hypotheses and justifications involving it should not be added to any paper except in the Discussion section. Specifically, misreporting the above three-way interaction as hypothesized before the data was collected not only would be intellectually questionable, it would ruin the model (disconfirmation) test--the three-way hypothesis can no longer formally be disconfirmed by the data at hand.[16]

---

[16] The three-way interaction could be re-tested using random subsets of data and simulated data sets generated from the covariance matrix of the study data set. However, the three-way interaction could still be an artifact of the data set at hand, and these retests could be argued to be opportunistic substitutes for new data from a subsequent study to formally test the interaction.

Parenthetically, there are several “shortcuts” that may save time and specification effort. For example, Equation 5 could be estimated using OLS regression, and if the “all centered” three-way interaction (with the plausible two-way interactions) is significant, experience suggests the structural equation results for the “all centered” three-way interaction are likely to be significant as well. If the “all centered” three-way interaction is not significant in regression, Equation 5 could be estimated using OLS regression and a plausible “permutation” three-way interaction. If this is significant, experience suggests the structural equation results for the “permutation” three-way interaction should also be significant. While a three-way interaction may still be significant in structural equation analysis, experience suggests that if the “all centered” and the plausible “permutation” three-way interactions are not significant in OLS regression, the structural equation specifications of these three-way interactions also may be non-significant.

In addition, while specifying all three “permutation” three-way interactions jointly typically suppress each other, testing them in pairs may suggest which one to attempt to justify. For example in the pedagogical example, when all three “permutation” three-way interactions were tested jointly in regression, and the least significant “permutation” three-way interaction and its related two-way interaction were dropped, this suggested that ALT(INVxSC) was significant.

In the pedagogical example ALT moderated INVxSC, and the results are shown in Table E.

The original motivation for interactions were hypothesized associations of INV and SC with Y that were unstable across studies. Stated differently it was the plausible moderation of INV and SC that was of interest. Thus, for example, Table D presented the INV-Y associations (the SC-Y were similar).

Table C presented the moderation of ALT by the INVxSC interaction. This also was of interest because the significant three-way interaction suggested the ALT-->Y association hypothesized in the studies was in fact conditional.

In summary, the significant three-way interaction $ALT_c(INV_u xSC_u)_c$ implies that all constituent variables' associations with Y (e.g., ALT_c-->Y, INV_u-->Y and SC_u-->Y) are conditional.

Equation 4c and 4d could be used to specify desired two-way interactions. These calculations are also included in the three-way interaction spreadsheets available by e-mail.

SUMMARY AND CONCLUSION

Because there is little guidance for estimating a latent variable (LV) "three-way" interaction (e.g., XZW), the paper explored these variables, and the results suggested that a "three-way" interaction may have several non-equivalent specifications. The paper also provided a pedagogical example to suggest the utility of these LV's. Hypothesizing a "three-way" interaction was discussed, their reliability was derived, a remedy for their nonessential ill-conditioning (their high correlations with X, Z and W) in real-world data was suggested, and an approach to interpreting them was illustrated.

In summary, a procedure for estimating XZW would follow steps 1 through 8 in the Discussion section.

REFERENCES

Aiken, Leona S. and Stephen G. West (1991), *Multiple Regression: Testing and Interpreting Interactions*, Newbury Park, CA: SAGE Publications.

Algina, James and Bradley C. Moulder (2001), "A Note on Estimating the Jöreskog-Yang Model for Latent Variable Interaction Using LISREL 8.3," Structural Equation Modeling, 8 (1) 40-52.

Anderson, J. C. and Gerbing, D. W. (1988). Structural Equation Modeling in Practice: a Review and Recommended Two-Step Approach. *Psychological Bulletin*, 103, 411-23.

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.

______ (1995), "Structural Equation Models that are Nonlinear in Latent Variables: A Least Squares Estimator," *Sociological Methodology*, 25, 223-251.

Bohrnstedt, G.W. and G. Marwell (1978), "The Reliability of Products of Two Random Variables," in Sociological Methodology, K. F. Schuessler ed., San Francisco: Jossy Bass, 254-273.

DeVellis, Robert F. (1991), *Scale Development: Theory and Applications*, Newbury Park, CA: SAGE Publications.

Fiske, S.T. and S.E. Taylor (1991). (2nd Ed.). Social Cognition. New York: McGraw-Hill.

Fornell, Claes and David F. Larker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.

Hackman, J.R., E. E. Lawler and L.W. Porter (1977), *Perspectives on Behavior in Organizations,* NY: McGraw-Hill.

Hunt, Shelby D. (1983). *Marketing Theory: The Philosophy of Marketing Science.* Homewood, IL: Irwin.

Jaccard, James, Robert Turrisi and Choi K. Wan (1990), *Interaction Effects in Multiple Regression*, Newbury Park, CA: SAGE Publications.

Jaccard, James and C. K. Wan (1995), "Measurement Error in the Analysis of Interaction Effects Between Continuous Predictors Using Multiple Regression: Multiple Indicator and Structural Equation Approaches," *Psychological Bulletin*, 117 (2), 348-357.

Jöreskog, Karl G. and Fan Yang (1996), "Nonlinear Structural Equation Models: The Kenny and Judd Model with Interaction Effects," *Advances in Structural Equation Modeling Techniques*, G.A. Marcoulides, R.E. Schumacker, eds., Hillsdale, NJ: LEA.

Jöreskog, Karl G. (2000), "Latent Variable Scores and Their Uses," (on-line paper), http://www.ssicentral.com/lisrel/ techdocs/lvscores.pdf.

Kahneman, D. and A. Tversky (1979), "Prospect Theory: An Analysis of Decisions under Risk," *Econometrica*, 47, 313–327.

Kenny, D. and C. M. Judd (1984), "Estimating the Nonlinear and Interactive Effects of Latent Variables," *Psychological Bulletin*, 96, 201-10.

Klein, A.G. and H. Moosbrugger (2000), "Maximum Likelihood Estimation of Latent Interaction Effects with the LMS Method," *Psychometrika*, 65, 457-474.

Klein, A.G. and B.O. Muthén (2002), "Quasi Maximum Likelihood Estimation of Structural Equation Models with Multiple Interactions and Quadratic Effects," Unpublished ms., Graduate School of Education, UCLA.

Lord, F.M. and M.R. Novic (1968), *Statistical Theories of Mental Test Scores*, Reading, PA: Addison-Wesley.

Lubinski, D. and Humphreys, L.G. (1990), "Assessing Spurious Moderator Effects: Illustrated Substantively with the Hypothesized ("Synergistic") Relation Between Spatial and Mathematical Ability," *Psychological Bulletin*, 107, 385-393.

Marquardt, D.W. (1980), "You Should Standardize the Predictor Variables in your Regression Models," *Journal of the American Statistical Association*, 75, 87-91.

Marsh, Herbert W., Zhonglin Wen and Kit-Tai Hau (2004), "Structural Equation Models of Latent Interactions: Evaluation of Alternative Estimation Strategies and Indicator Construction," *Psychological Methods*, 9 (3), 275-300.

Mathieu, J.E., S.I. Tannenbaum and E. Salas (1992), "Influences of Individual and Situational Characteristics on Measuring of Training Effectiveness," *Academy of Management Journal*, 35, 828-847.

Molder, Bradley C. and James Algina (2002), "Comparison of Methods for Estimating and Testing Latent Variable Interactions," *Structural Equation Modeling*, 9 (1), 1-19.

Mølgaard C, Lykke B, Michaelsen K. Thomsen (1999), "Whole Body Bone Mineral Accretion in Healthy Children and Adolescents," *Archives of Disease in Childhood*, 81, 10–15.

Nunnally, Jum C. (1978), *Psychometric Theory*, 2nd Ed., New York: McGraw-Hill.

Ping, Robert (1993a), "The Effects of Satisfaction and Structural Constraints on Retailer Exiting, Voice, Loyalty, Opportunism, and Neglect," *Journal of Retailing*, 69 (Fall), 320-352. (13)

______ (1995), "A Parsimonious Estimating Technique for Interaction and Quadratic Latent Variables," *The Journal of Marketing Research*, 32 (August), 336-347.

______ (1996a), "Latent Variable Interaction and Quadratic Effect Estimation: A Two-Step Technique Using Structural Equation Analysis," *Psychological Bulletin*, 119 (January), 166-175.

Ping, R.A. (1996b), "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients," *Multivariate Behavioral Research*, 31 (1), 95-120.

______ (1996c), "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients," *Multivariate Behavioral Research*, 31 (1), 95-120.

______ (1996d), "Estimating Latent Variable Interactions and Quadratics: The State of This Art," *Journal of Management*, 22 (1), 163-183.

______ (2004a) "On Assuring Valid Measures for Theoretical Models Using Survey Data," *Journal of Business Research*, 57 (2) (February), 125-141.

Ping, Robert (2007a) "Salesperson-Employer Relationships: Salesperson Responses to Relationship Problems and their Antecedents," the *Journal of Personal Selling and Sales Management*, 27 (1) (Winter), p39-57. (13)

Podsakoff, Philip M., William D. Tudor, Richard A. Grover and Vandra L. Huber (1984), “Situational Moderators of Leader Reward and Punishment Behaviors: Fact or Fiction?” *Organizational Behavior and Human Performance*, 34, 21-63.

Rusbult, Caryl E., Dan Farrell, Glen Rogers and Arch G. Mainous III (1988), "Impact of Exchange Variables on Exit, Voice, Loyalty, and Neglect: An Integrative Model of Responses to Declining Job Satisfaction," *Academy of Management Journal*, 31 (September), 599-627.

Schermelleh-Engle, K., A. Kein and H. Moosbrugger (1998), "Estimating Nonlinear Effects using a Latent Moderated Structural Equations Approach," in *Interaction and Nonlinear Effects in Structural Equation Modeling*, R. E, Schumacker and G. A. Marcoulides, eds., Mahwah, NJ: Erlbaum.

Shaw, Marvin E. and Philip R. Costanzo (1982), *Theories of Social Psychology*, NY: McGraw-Hill

Wall, M.M. and Y. Amemiya (2001), "Generalized Appended Product Indicator Procedure for Nonlinear Structural Equation Analysis," *Journal of Educational and Behavioral Statistics*, 26, 1-29.

Table A--Selected Correlations[a]

	SATu	SATc	ALTu	ALTc	INVu	INVc	Scu	SCc
AcIcSCc	-0.3641	-0.3641	0.5657	0.5657	-0.4513	-0.4513	-0.4662	-0.4662
AuIuSCu	-0.0382	-0.0382	0.4059	0.4059	0.5499	0.5499	0.5792	0.5792
Ac(ISC)c	-0.0094	-0.0094	0.0930	0.0930	-0.0501	-0.0501	-0.1397	-0.1397

[a] SATu denotes an uncentered SAT, SATc denotes a mean-centered SAT, etc.; AcIcSCc is the product of mean-centered (MC) ALT, MC INV and MC SC.
AuIuSCu is the product of uncentered (UC) ALT, UC INV and UC SC.
Ac(ISC)c is the product of centered ALT and the mean-centered product of UC INV and UC SC.

Table B--Equation 5 Unstandardized Structural Coefficient Estimates

$$Y = aSAT_u + bALT_c + cINV_u + dSC_u + gALT_c(INV_u xSC_u)_c + h(INV_u xSC_u)_c$$

a	b	c	d	g	h	
-.06	.14	-.08	.24	.02	-.02	Unstd Beta
-.85	2.43	-.42	.97	2.11	-.41	T-value

Table C--Interpretation: (Factored Coefficient) ALT Associations with Y Due to the INVxSC Interaction[a]

		SC =					
		1	2	3	3.26 [b]	4	5
INV =	1	0.1672	0.1889	0.2106	0.2162	0.2323	0.2540
	2	0.1889	0.2323	0.2757	0.2870	0.3191	0.3626
	2.51 [b]	0.2000	0.2544	0.3089	0.3231	0.3634	0.4179
	3	0.2106	0.2757	0.3409	0.3578	0.4060	0.4711
	4	0.2323	0.3191	0.4060	0.4286	0.4928	0.5797
	5	0.2540	0.3626	0.4711	0.4993	0.5797	0.6882
ALT-structural coefficient standard errors[c]							
INV =	1	0.1382	0.1290	0.1199	0.1176	0.1111	0.1026
	2	0.1290	0.1111	0.0944	0.0903	0.0795	0.0677
	2.51 [b]	0.1243	0.1024	0.0828	0.0783	0.0675	0.0598
	3	0.1199	0.0944	0.0731	0.0688	0.0607	0.0626
	4	0.1111	0.0795	0.0607	0.0596	0.0665	0.0925
	5	0.1026	0.0677	0.0626	0.0679	0.0925	0.1359
ALT-structural coefficient t-values							
INV =	1	1.2100	1.4647	1.7560	1.8385	2.0906	2.4762
	2	1.4647	2.0906	2.9209	3.1780	4.0142	5.3592
	2.51 [b]	1.6083	2.4844	3.7318	4.1280	5.3875	6.9841
	3	1.7560	2.9209	4.6631	5.2038	6.6931	7.5250
	4	2.0906	4.0142	6.6931	7.1959	7.4104	6.2685
	5	2.4762	5.3592	7.5250	7.3503	6.2685	5.0637

[a] The unstandardized structural coefficients suggest that the ALT-Y association was positive over the range of INV and SC in the study. The structural coefficient t-values suggest, however, that the moderated Y-ALT association was NS for low values of INV and SC in the study.

[b] Mean value in the study.

[c] The SQUARE of the Standard Error (SE) of the Equation 5 factored coefficient of ALT is:

$$\begin{aligned} Var(b + g(INVxSC-M_{INVxSCT})) &= Var(b) + (INVxSC- M_{INVxSCT})^2 Var(g) + 2(INVxSC-M_{INVxSCT})Cov(b,g) \\ &= SE(b)^2 + (INVxSC-M_{INVxSCT})^2 SE(g)^2 + 2(INVxSC- M_{INVxSCT})Cov(b,g). \end{aligned}$$

Table D--Interpretation: (Factored Coefficient) INV Associations with Y
Due to the ALTxSC Interaction (unstandardized betas and se's omitted)[a]

				INV (T-values)				
	Centrd	----------------------------SC=----------------------------						
ALT=	ALT=	1	2	3	3.26 [b]	4	5	INV=
1	-1.51	0.71	0.68	0.66	0.65	0.63	0.59	1
		0.47	0.39	0.30	0.27	0.17	-0.01	2
		0.37	0.27	0.12	0.08	-0.08	-0.39	2.51
		0.29	0.15	-0.04	-0.10	-0.33	-0.81	3
		0.14	-0.05	-0.35	-0.46	-0.85	-1.74	4
		0.03	-0.23	-0.64	-0.79	-1.35	-2.57	5
2	-0.51	0.67	0.65	0.63	0.62	0.60	0.56	1
		0.45	0.38	0.29	0.26	0.17	-0.01	2
		0.36	0.26	0.12	0.08	-0.08	-0.39	2.51
		0.28	0.15	-0.04	-0.10	-0.33	-0.83	3
		0.14	-0.05	-0.35	-0.46	-0.87	-1.90	4
		0.02	-0.23	-0.65	-0.80	-1.44	-2.98	5
2.51 [b]	0	0.65	0.63	0.61	0.60	0.58	0.55	1
		0.44	0.37	0.28	0.26	0.16	-0.01	2
		0.35	0.25	0.12	0.07	-0.08	-0.39	2.51
		0.27	0.14	-0.04	-0.10	-0.33	-0.84	3
		0.13	-0.05	-0.35	-0.46	-0.88	-1.97	4
		0.02	-0.23	-0.65	-0.81	-1.48	-3.14	5
3	0.49	0.63	0.61	0.59	0.58	0.56	0.53	1
		0.43	0.36	0.28	0.25	0.16	-0.01	2
		0.34	0.25	0.12	0.07	-0.08	-0.39	2.51
		0.26	0.14	-0.04	-0.10	-0.33	-0.85	3
		0.13	-0.05	-0.35	-0.46	-0.89	-2.04	4
		0.02	-0.22	-0.65	-0.82	-1.52	-3.21	5
4	1.49	0.59	0.57	0.55	0.55	0.53	0.50	1
		0.40	0.35	0.27	0.24	0.15	-0.01	2
		0.32	0.24	0.11	0.07	-0.08	-0.39	2.51
		0.25	0.14	-0.04	-0.10	-0.32	-0.86	3
		0.13	-0.05	-0.35	-0.46	-0.91	-2.11	4
		0.02	-0.22	-0.66	-0.84	-1.58	-3.09	5
5	2.49	0.55	0.53	0.52	0.51	0.49	0.47	1
		0.38	0.33	0.25	0.23	0.15	-0.01	2
		0.31	0.23	0.11	0.07	-0.07	-0.39	2.51
		0.24	0.13	-0.04	-0.10	-0.32	-0.88	3
		0.12	-0.05	-0.35	-0.46	-0.93	-2.11	4
		0.02	-0.22	-0.67	-0.85	-1.61	-2.71	5

[a] The unstandardized structural coefficients suggest that at low values of ALT, INV and SC the INV-Y association was positive, while at high values of ALT, INV and SC it was negative. However, the structural coefficient t-values suggest that only the negative associations were significant in the study (as originally hypothesized). (These results also hint that significant positive associations might occur only if the average ALT, INV and SC were materially higher in the study.) The "band of significance" in the study extended from the lower left corner (high ALT and low SC) to the upper right corner (low ALT and high SC). The maximum associations and t-values in the study were at high ALT.

[b] Mean value in the study.

[c] The SQUARE of the Standard Error (SE) of the Equation 5 factored coefficient of INV is:

$$\mathrm{Var}(c + g(ALTxSC\text{-}M_{INVxSCT}/INV))$$
$$= \mathrm{Var}(c) + (ALTxSC\text{-}M_{INVxSCT}/INV)^2\mathrm{Var}(g) + 2(ALTxSC\text{-}M_{INVxSCT}/INV)\mathrm{Cov}(c,g).$$
$$= SE(c)^2 + (ALTxSC\text{-}M_{INVxSCT}/INV)^2 SE(g)^2 + 2(ALTxSC\text{-}M_{INVxSCT}/INV)\mathrm{Cov}(c,g).$$

Table E--Interpretation: (Factored Coefficient) INVxSC Associations with Y Due to the ALT(INVxSC) Interaction[a]

ALT Value[b]	Centered ALT Value	(INVxSC)c Coefficient[c]	SE of (INVxSC)c Coefficient[d]	t-Value (INVxSC)c Coefficient
1	-1.51	-0.060	0.003	-17.47
2	-0.51	-0.038	0.004	-9.89
2.51	0	-0.027	0.004	-6.51
3	0.49	-0.016	0.004	-3.65
4	1.49	0.005	0.005	1.00
5	2.49	0.027	0.006	4.23

[a] The structural coefficient t-values suggest that the ALT moderated the INVxSC-Y association as suspected. They also reveal that at low ALT the INVxSC-Y association was negative, but at very high ALT it was positive.

[b] Mean value in the study.

[c] The SQUARE of the Standard Error (SE) of the Equation 5 factored coefficient of (INVxSC)c is:

Var(b + gINVxSC)

$= Var(b) + (INVxSC)^2Var(g) + 2(INVxSC)Cov(b,g).$

$= SE(b)^2 + (INVxSC)^2SE(g)^2 + 2(INVxSC)Cov(b,g).$

Appendix A—The Derivation of the Loading and Measurement Error Variance of $x_c(zw)_c$

Under the Kenny and Judd (1984) normality assumptions the variance of the product of $x_c = [\Sigma x_i^u - \mathbf{M}(\Sigma x_i^u)]$ and $(zw)_c = [(\Sigma z_j^u)(\Sigma w_k^u) - \mathbf{M}((\Sigma z_j^u)(\Sigma w_k^u))]$ is

4a) $$\begin{aligned}Var(x_c(zw)_c) = (\Lambda_X\Lambda_{ZW})^2 Var(X_c(Z_u x W_u)_c) + \Lambda_X^2 Var(X_c)\Theta_{(zw)c} \\ + \Lambda_{ZW}^2 Var((Z_u W_u)_c)\Theta_{xc} + \Theta_{xc}\Theta_{(zw)c},\end{aligned}$$

(e.g., Ping 1995) where $\Lambda_X = \Sigma\lambda_{xi}$ is the loading of X on x_c, Λ_{ZW} is the loading of ZW on $(zw)_c$, Var(XZW) is the error disattenuated variance of XZW, Var(X) is the error disattenuated variance of X, and $\Theta_{(zw)c}$ is the measurement error variance of $(zw)_c$. Var(ZW) is the error disattenuated variance of ZW, and $\Theta_{xc} = \Sigma Var(\varepsilon_{xi})$ is the measurement error variance of x_c. This provides a specification of the XZW using the indicator $x_c(zw)_c$ with the loading $\Lambda_X\Lambda_{ZW}$ and the measurement error variance of $\Lambda_X^2 Var(X)\Theta_{(zw)c} + \Lambda_{ZW}^2 Var(ZW)\Theta_{xc} + \Theta_{xc}\Theta_{(zw)c}$, if estimates of the parameters associated with ZW are available (the parameters associated with X are available in the measurement model) (see Ping 1995).

To provide an estimate of $Var((Z_u W_u)_c)$,

4b) $$\begin{aligned}Var((Z_u W_u)_c) &= Var((Z_u W_u) - \mathbf{E}(Z_u W_u)) \\ &= Var(Z_u W_u) \\ &= \mathbf{E}^2(Z)Var(W) + \mathbf{E}^2(W)Var(Z) + 2\mathbf{E}(Z)\mathbf{E}(W)Cov(Z,W) \\ &\quad + Var(Z)Var(W) + Cov^2(Z,W)\end{aligned}$$

(Bohrnstedt and Goldberger 1969), where **E**(*) denotes the expectation (mean) of * and $\mathbf{E}^2$(*) its square, Cov(Z,W) is the covariance of Z and W, and $Cov^2(Z,W)$ is its square (Z and W are uncentered and have non-zero means). Since first moments (means) are unaffected by measurement error, **E**(Z), for example, can be estimated using $\mathbf{E}(\Sigma z_j^u)$ (the SAS, SPSS, etc. mean of Σz_j^u) (Lord and Novic 1968), and the variances of Z and W, and their covariance, can be estimated in a measurement model without $ALT_c(INV_u x SC_u)_c$ (see Footnote 8).

To provide an estimate of Λ_{ZW} and $\Theta_{(zw)c}$,

$$\begin{aligned}(zw)_c &= (\Sigma z_j^u)(\Sigma w_k^u) - \mathbf{M}((\Sigma z_j^u)(\Sigma w_k^u)) \\ &= [(\Sigma\lambda_{zj})Z + \Sigma\varepsilon_{zj}][(\Sigma\lambda_{wk})W + \Sigma\varepsilon_{wk}] - \mathbf{M}((\Sigma z_j^u)(\Sigma w_k^u)) \\ &= (\Lambda_Z Z + \Sigma\varepsilon_{zj})(\Lambda_W W + \Sigma\varepsilon_{wk}) - \mathbf{M}((\Sigma z_j^u)(\Sigma w_k^u)) \\ &= \Lambda_Z Z\Lambda_W W + \Lambda_Z Z\Sigma\varepsilon_{wk} + \Lambda_W W\Sigma\varepsilon_{zj} + \Sigma\varepsilon_{zj}\Sigma\varepsilon_{wk} - \mathbf{M}((\Sigma x_i^u)(\Sigma x_i^u)),\end{aligned}$$

with the usual assumptions that the LV's are independent of measurement errors, measurement errors have zero expectations, and measurement errors are independent.

Thus,

$$
\begin{aligned}
\mathrm{Var}((zw)_c) &= \mathrm{Var}(\Lambda_Z Z \Lambda_W W + \Lambda_Z Z \Sigma\varepsilon_{wk} + \Lambda_W W \Sigma\varepsilon_{zj} + \Sigma\varepsilon_{zj}\Sigma\varepsilon_{wk} - \mathbf{M}((\Sigma x_i^u)(\Sigma x_i^u))) \\
&= \mathrm{Var}(\Lambda_Z Z \Lambda_W W + \Lambda_Z Z \Sigma\varepsilon_{wk} + \Lambda_W W \Sigma\varepsilon_{zj} + \Sigma\varepsilon_{zj}\Sigma\varepsilon_{wk}) \\
&= \mathrm{Var}(\Lambda_Z Z \Lambda_W W) + \mathrm{Var}(\Lambda_Z Z \Sigma\varepsilon_{wk} + \Lambda_W W \Sigma\varepsilon_{zj} + \Sigma\varepsilon_{zj}\Sigma\varepsilon_{wk}) \\
&\quad + 2\mathrm{Cov}(\Lambda_Z Z \Lambda_W W, \Lambda_Z Z \Sigma\varepsilon_{wk} + \Lambda_W W \Sigma\varepsilon_{zj} + \Sigma\varepsilon_{zj}\Sigma\varepsilon_{wk}) \\
&= \mathrm{Var}(\Lambda_Z Z \Lambda_W W) + \mathrm{Var}(\Lambda_Z Z \Sigma\varepsilon_{wk}) + \mathrm{Var}(\Lambda_W W \Sigma\varepsilon_{zj}) + \mathrm{Var}(\Sigma\varepsilon_{zj}\Sigma\varepsilon_{wk}) \\
&= (\Lambda_Z \Lambda_W)^2 \mathrm{Var}(ZW) + \Lambda_Z^2 \mathrm{Var}(Z\Sigma\varepsilon_{wk}) + \Lambda_W^2 \mathrm{Var}(W\Sigma\varepsilon_{zj}) \\
&\quad + \mathrm{Var}(\Sigma\varepsilon_{zj}\Sigma\varepsilon_{wk}) \ .
\end{aligned}
$$

From Equation 4b

$$
\begin{aligned}
\mathrm{Var}((zw)_c) &= (\Lambda_Z \Lambda_W)^2 \mathrm{Var}(ZW) + \Lambda_Z^2 (\mathbf{E}^2(Z)\mathrm{Var}(\Sigma\varepsilon_{wk}) + \mathrm{Var}(Z)\mathrm{Var}(\Sigma\varepsilon_{wk})) \\
&\quad + \Lambda_W^2 (\mathbf{E}^2(W)\mathrm{Var}(\Sigma\varepsilon_{zj}) + \mathrm{Var}(W)\mathrm{Var}(\Sigma\varepsilon_{zj})) + \mathrm{Var}(\Sigma\varepsilon_{zj})\mathrm{Var}(\Sigma\varepsilon_{wk}) \\
&= (\Lambda_Z \Lambda_W)^2 \mathrm{Var}(ZW) + \Lambda_Z^2 (\mathbf{E}^2(Z)\Theta_w + \mathrm{Var}(Z)\Theta_w) \\
&\quad + \Lambda_W^2 (\mathbf{E}^2(W)\Theta_z + \mathrm{Var}(W)\Theta_z + \Theta_w\Theta_z \ ,
\end{aligned}
$$

where $\Theta_w = \Sigma\mathrm{Var}(\varepsilon_{wk})$ and $\Theta_z = \Sigma\mathrm{Var}(\Sigma\varepsilon_{zj})$, because the covariances involving measurement errors are zero, and the expansions of $\mathrm{Var}(X_u\varepsilon_{Xu})$ and $\mathrm{Var}(\varepsilon_{Xu}\varepsilon_{Xu})$ contain zero $\mathbf{E}(\varepsilon_{Xu})$ terms.

As a result, in Equation 4a

4c) $\Lambda_{ZW} = (\Sigma\lambda_{zj})(\Sigma\lambda_{wk})$

and

4d) $\Theta_{(zw)c} = \Lambda_Z^2(\mathbf{E}^2(Z)\Theta_w + \mathrm{Var}(Z)\Theta_w) + \Lambda_W^2(\mathbf{E}^2(W)\Theta_z + \mathrm{Var}(W)\Theta_z + \Theta_w\Theta_z$

and the parameters in Equation 4a all can be estimated. Specifically, the loading of XZW,

4e) $\Lambda_X \Lambda_{ZW} = (\Sigma\lambda_{xi})(\Sigma\lambda_{zj})(\Sigma\lambda_{wk})$

and the measurement error variance of $x_c(zw)_c$,

4f) $\Lambda_X^2 \mathrm{Var}(X)\Theta_{(zw)c} + \Lambda_{ZW}^2 \mathrm{Var}(ZW)\Theta_{xc} + \Theta_{xc}\Theta_{(zw)c}$,

where Λ_{ZW}, Var(ZW) and $\Theta_{(zw)c}$ are determined using Equations 4b, 4c and 4d.

Appendix B—The Derivation of the Reliability of $X_c(ZW)_c$

The reliability of $X_c(ZW)_c$ is

$$\rho_{XcZcWc} = \frac{Var(X_c(ZW)_c)}{Var(x_c(zw)_c)} ,$$

where X_c, Z and W are true scores (Z and W arew uncentered), and x_c, z_c and w_c are observed scores. This might be calculated from the structural model, and SPSS using the cases for the indicator $x_c(zw)_c$.

The following is more elegant: because

6) $$\rho_{Xc(ZW)c} = \frac{C_{xc,(zw)c}^2 + V_{zc}V_{(zw)c}\rho_{xc}\rho_{(zw)c}}{C_{xc,(zw)c}^2 + V_{xc}V_{(zw)c}} ,$$

(Bohrnstedt and Marwell 1978) where x_c and $(zw)_c$ are observed (SPSS, SAS, etc.) variables (i.e., sums of indicators), $C_{xc,(zw)c}$ denotes the (observed) covariance of x_c and $(zw)_c$, V_{xc}, for example, denotes the (observed) variance of x_c, and ρ denotes reliability (Bohrnstedt and Marwell 1978). The reliability of $(zw)_c$ is,

6a) $$\rho_{(zw)c} = \frac{\mathbf{E}(z)^2V_W\rho_W + \mathbf{E}(w)^2V_z\rho_z + 2\mathbf{E}(z)\mathbf{E}(w)C_{z,w} + V_zV_w\rho_z\rho_w + C_{z,w}^2}{\mathbf{E}(z)^2V_w + \mathbf{E}(w)^2V_z + 2\mathbf{E}(z)\mathbf{E}(w)C_{z,w} + V_zV_w + C_{z,w}^2} ,$$

(Bohrnstedt and Marwell 1978). $\mathbf{E}(Z)$, for example, in Equation 6a can be estimated using $\mathbf{E}(\Sigma z_j^u)$ (Lord and Novic 1968), and ρ_Z, for example can be estimated using coefficient alpha (Anderson and Gerbing (1988), and the other parameters in Equation 6 and 6a can be estimated using structural model parameter estimates.

Appendix C—The Derivation of the Loading, Measurement Error Variance and Reliability of an "All-Centered" Three-Way Interaction, $X_cZ_cW_c$

Loading and Measurement Error Variance

The loading, measurement error variance and the reliability of the indicator, $x_cz_cw_c$, of an "all centered" interaction, $X_cZ_cW_c$, are as follows: under the Kenny and Judd (1984) normality assumptions (see Footnote 4)

7) $$\begin{aligned}Var(x_cz_cw_c) = {} & Var(x_c)Var(z_c)Var(w_c) + 2Var(x_c)Cov^2(z_c,w_c) \\ & + 2Var(z_c)Cov^2(x_c,w_c) + 2Var(w_c)Cov^2(x_c,z_c) \\ & + 8Cov(x_c,z_c)Cov(x_c,w_c)Cov(z_c,w_c)\end{aligned}$$

(Kendall and Stewart 1958) where x_c, z_c and w_c are sums of indicators. In Equation 7)

7a) $$\begin{aligned}& Var(x_c)Var(z_c)Var(w_c) \\ & = (\Lambda_{Xc}^2Var(X_c) + \Theta_{Xc})(\Lambda_{Zc}^2Var(Z_c) + \Theta_{Zc}) (\Lambda_{Wc}^2Var(W_c) + \Theta_{Wc}) \\ & = \Lambda_{Xc}^2Var(X_c)\Lambda_{Zc}^2Var(Z_c)\Lambda_{Wc}^2Var(W_c) + \Lambda_{Zc}^2\Lambda_{Wc}^2Var(Z_c)Var(W_c)\Theta_{Xc} \\ & \quad + \Lambda_{Xc}^2\Lambda_{Wc}^2Var(X_c)Var(W_c)\Theta_{Zc} + \Lambda_{Wc}^2Var(W_c)\Theta_{Xc}\Theta_{Zc} \\ & \quad + \Lambda_{Xc}^2\Lambda_{Zc}^2Var(X_c)Var(Z_c)\Theta_{Wc} + \Lambda_{Zc}^2Var(Z_c)\Theta_{Xc}\Theta_{Wc} \\ & \quad + \Lambda_{Xc}^2Var(X_c)\Theta_{Zc}\Theta_{Wc} + \Theta_{Xc}\Theta_{Zc}\Theta_{Wc},\end{aligned}$$

where x_c, for example, equals $\Lambda_{Xc}^2Var(X_c) + \Theta_{Xc}$, and $\Lambda_X = \Sigma\lambda_{xi}$ is the loading of X on x_c, $\Theta_{xc} = \Sigma Var(\varepsilon_{xi})$ is the measurement error variance of x_c and ε_{xi} is the measurement error of x_i

Also in Equation 7,

7b) $$\begin{aligned}& 2Var(x_c)Cov^2(z_c,w_c) \\ & = 2(\Lambda_{Xc}^2Var(X_c)+\Theta_{Xc})Cov^2(\Lambda_{zc}Z_c+\varepsilon_{zc},\Lambda_{wc}W_c+\varepsilon_{wc}) \\ & = 2\Lambda_{Xc}^2\Lambda_{Zc}^2\Lambda_{Wc}^2Var(X_c)Cov^2(z_c,w_c)+\Theta_{Xc}\Lambda_{zc}^2\Lambda_{Wc}^2Cov^2(Z_c,W_c).\end{aligned}$$

Similarly,

7c) $$\begin{aligned}& 2Var(z_c)Cov^2(x_c,w_c) \\ & = 2\Lambda_{Xc}^2\Lambda_{Zc}^2\Lambda_{Wc}^2Var(Z_c)Cov^2(x_c,w_c)+\Theta_{Zc}\Lambda_{zc}^2\Lambda_{Wc}^2Cov^2(X_c,W_c)\end{aligned}$$

7d) $$\begin{aligned}& 2Var(w_c)Cov^2(x_c,z_c) \\ & = 2\Lambda_{Xc}^2\Lambda_{Zc}^2\Lambda_{Wc}^2Var(W_c)Cov^2(x_c,z_c)+\Theta_{wc}\Lambda_{zc}^2\Lambda_{Wc}^2Cov^2(X_c,Z_c),\end{aligned}$$

and

7e) $$8Cov(x_c,z_c)Cov(x_c,w_c)Cov(z_c,w_c)$$

$$= 8[\mathrm{Cov}(\Lambda_{Xc}X_c + \Sigma\varepsilon_{xj},\Lambda_{Zc}Z_c + \Sigma\varepsilon_{zj})*$$
$$\mathrm{Cov}(\Lambda_{Zc}Z_c + \Sigma\varepsilon_{zj},\Lambda_{Wc}W_c + \Sigma\varepsilon_{wj})*$$
$$\mathrm{Cov}(\Lambda_{Xc}X_c + \Sigma\varepsilon_{xj},\Lambda_{Wc}W_c + \Sigma\varepsilon_{wj})]$$

$$= 8\Lambda_{Xc}^2\Lambda_{Zc}^2\Lambda_{Wc}^2\mathrm{Cov}(X_c,Z_c)\mathrm{Cov}(X_c,Z_c)\mathrm{Cov}(X_c,Z_c),$$

where $\Sigma\varepsilon_{xj}$, for example, is the sum of the measurement errors of x_j.

Substituting Equations 7a-7e into Equation 7, then simplifying by replacing the sum of the first term in Equations 7a-7d, plus Equation 7e, with their equivalent, $\Lambda_{Xc}^2\Lambda_{Zc}^2\Lambda_{Wc}^2\mathrm{Var}(X_cZ_cW_c)$ (Kendall and Stewart 1958), Equation 7 becomes

7f) $$\mathrm{Var}(x_cz_cw_c) = \Lambda_{Xc}^2\Lambda_{Zc}^2\Lambda_{Wc}^2\mathrm{Var}(X_cZ_cW_c)$$

$$+ \Lambda_{Zc}^2\Lambda_{Wc}^2\mathrm{Var}(Z_c)\mathrm{Var}(W_c)\Theta_{Xc}$$
$$+ \Lambda_{Xc}^2\Lambda_{Wc}^2\mathrm{Var}(X_c)\mathrm{Var}(W_c)\Theta_{Zc} + \Lambda_{Wc}^2\mathrm{Var}(W_c)\Theta_{Xc}\Theta_{Zc}$$
$$+ \Lambda_{Xc}^2\Lambda_{Zc}^2\mathrm{Var}(X_c)\mathrm{Var}(Z_c)\Theta_{Wc} + \Lambda_{Zc}^2\mathrm{Var}(Z_c)\Theta_{Xc}\Theta_{Wc}$$
$$+ \Lambda_{Xc}^2\mathrm{Var}(X_c)\Theta_{Zc}\Theta_{Wc} + \Theta_{Xc}\Theta_{Zc}\Theta_{Wc}$$
$$+ \Theta_{Xc}\Lambda_{zc}^2\Lambda_{Wc}^2\mathrm{Cov}^2(Z_c,W_c)$$
$$+ \Theta_{Zc}\Lambda_{zc}^2\Lambda_{Wc}^2\mathrm{Cov}^2(X_c,W_c)$$
$$+ \Theta_{wc}\Lambda_{zc}^2\Lambda_{Wc}^2\mathrm{Cov}^2(X_c,Z_c)\ .$$

Thus, the loading of $X_cZ_cW_c$ is $\Lambda_{Xc}\Lambda_{Zc}\Lambda_{Wc}$ and the measurement error variance of $X_cZ_cW_c$ is given by the sum of the second through the last terms of Eqn 7f.

Reliability

The reliability of $X_cZ_cW_c$ is

$$\rho_{XcZcWc} = \frac{\mathrm{Var}(X_cZ_cW_c)}{\mathrm{Var}(x_cz_cw_c)}\ ,$$

where X_c, Z_c and W_c are true scores, and x_c, z_c and w_c are observed scores. This might be calculated from the Structural model, and SPSS using the cases for the indicator $x_cz_cw_c$.

The following is more elegant: because

$$\mathrm{Cov}(a^T,b^T) = \mathrm{Cov}(a,b)$$

(Lord and Novic 1968), where a^T and b^T are true scores and a and b are observed scores,

$$\mathrm{Var}(X_c) = \rho_{Xc}\mathrm{Var}(x_c)$$

and

$Cov^2(a,b) = r^2(a,b)Var(a)Var(b)$,

where r(a,b) is the correlation of a and b, Equation 7 can be rewritten as

$$\text{8a)}\quad Var(x_c z_c w_c) = Var(x_c)Var(z_c)Var(w_c)[1 + 2r^2(z_c,w_c) + 2r^2(x_c,w_c) + 2r^2(x_c,z_c) + 8r^2(x_c,z_c)r^2(x_c,w_c)r^2(w_c,z_c)Var(x_c)Var(z_c)Var(w_c)] .$$

Similarly, replacing x_c, z_c and w_c in Equation 7 with X_c, Z_c and W_c

$$\begin{aligned}\text{8b)}\quad Var(X_c Z_c W_c) &= Var(X_c)Var(Z_c)Var(W_c) + 2Var(X_c)Cov^2(Z_c,W_c) \\ &\quad + 2Var(Z_c)Cov^2(X_c,W_c) + 2Var(W_c)Cov^2(X_c,Z_c) \\ &\quad + 8Cov(X_c,Z_c)Cov(X_c,W_c)Cov(Z_c,W_c) \\ &= Var(x_c)Var(z_c)Var(w_c)[\rho_{Xc}\rho_{Zc}\rho_{Wc} + 2\rho_{Xc}r^2(z_c,w_c) \\ &\quad + 2\rho_{Zc}r^2(x_c,w_c) + 2\rho_{Wc}r^2(x_c,z_c) \\ &\quad + 8r^2(x_c,z_c)r^2(x_c,w_c)r^2(w_c,z_c)Var(x_c)Var(z_c)Var(w_c)]\end{aligned}$$

and

$$\rho_{XcZcWc} = \frac{\rho_{Xc}\rho_{Zc}\rho_{Wc} + 2\rho_{Xc}r^2(z_c,w_c) + 2\rho_{Zc}r^2(x_c,w_c) + 2\rho_{Wc}r^2(x_c,z_c) + 8r^2(x_c,z_c)r^2(x_c,w_c)r^2(w_c,z_c)Var(x_c)Var(z_c)Var(w_c)}{1 + 2r^2(z_c,w_c) + 2r^2(z_c,w_c) + 2r^2(x_c,z_c) + 8r^2(x_c,z_c)r^2(x_c,w_c)r^2(w_c,z_c)Var(x_c)Var(z_c)Var(w_c)} .$$

BUT WHAT ABOUT CATEGORICAL (NOMINAL) VARIABLES IN LATENT VARIABLE MODELS?

ABSTRACT

The paper suggests an approach for specification, estimation and interpretation of a categorical or nominal exogenous (independent) variable in theory or hypothesis tests of latent variable models with survey data. An example using survey data is provided.

Anecdotally, categorical variables (e.g., Marital Status) are ubiquitous in applied marketing research. However, they are absent from published theory (hypothesis) tests of latent variable models using survey data.

One plausible explanation is there is no explicit provision for "truly" categorical variables in the popular structural equation (covariant structure) analysis software packages (e.g., LISREL, EQS, AMOS, etc.). There, the term "categorical variable" means an ordinal variable (e.g., an attitude measured by a Likert scale) (e.g., Jöreskog and Sörbom 1996), rather than a nominal variable such as Marital Status.

Further, normality is a fundamental assumption in covariance structural analysis (e.g., in LISREL, EQS, Amos, etc.) for maximum likelihood estimation, the preferred estimator for hypothesis tests involving latent variables and survey data (e.g., Jöreskog and Sörbom 1996). A (truly) categorical independent variable is typically estimated using "dummy" variables that are not normally distributed. (For example, while case values for the categorical variable Marital Status, for example, might be 1 for Single, 2 for Married, 3 for Divorced, etc., new variables, for example Dummy_Single, Dummy_Married, etc., are created and estimated instead of Marital Status. Dummy_Single might have a case value of 1 if Marital Status = Single, and 0 otherwise, Dummy_Married might have cases that have the value 1 if Marital Status = Married, and 0 otherwise, etc.)

There are other less obvious barriers in survey-data theory tests to adding (truly) categorical exogenous variables to models that also contain latent variables, including determining the significance of a categorical variable when its dummy variables are estimated instead. If each dummy variable is significant (or nonsignificant), it seems reasonable to conclude that the categorical variable from which the dummies were created is significant (or nonsignificant). However, if some dummy variables are significant but some are not, there is no guidance for estimating the significance of the categorical variable from which they were created. In addition, interpreting a significant categorical variable can involve interpreting changes in intercepts, parameters that are not usually estimated in theoretical model testing.

Several approaches have been suggested for estimating categorical variables (e.g., dummy variable regression, logistic regression, latent category models, etc.). However, there is no guidance for estimating a "mixed" model--one that combines a categorical exogenous variable with latent variables--in theory (hypothesis) tests involving survey data.

This paper addresses these matters. It suggests a specification for a categorical variable in theory (hypothesis) tests of latent variable models involving survey data. It also discusses the estimation and interpretation of a categorical variable in these "mixed" models.

AN EXAMPLE

To expedite the presentation of these matters, we will use a real-world data set involving buyer-seller relationship Satisfaction (SAT), Alternative relationship attractiveness (ALT), and Exit propensity (EXI). Data (200+ cases) was collected in a survey to test a larger model in which Satisfaction and Alternative Unattractiveness were hypothesized to be negatively associated with Exiting. SAT, ALT and EXI were measured using multiple item measures (Likert scales). The resulting latent variables, SAT, ALT and EXI were judged to be valid and reliable, and the itemizations of each were judged to be internally and externally consistent in the Anderson and Gerbing (1988) sense.[1]

The structural equation

$$EXI = b_1SAT + b_2ALT + \zeta, \tag{1}$$

where b_i are structural coefficients and ζ is structural disturbance, was estimated using LISREL and maximum likelihood estimation, and SAT and ALT were significantly (negatively) associated with EXI as shown in Table A.

SAT was measured with five-point Likert-scaled items that each could be analyzed as a categorical variable. (E.g., the SAT indicator Sa2 had 5 categories: those respondents who strongly agreed they were satisfied, those who agreed they were satisfied, those who were neutral, etc.) For pedagogical purposes SAT was replaced in Equation 1 with one of its high loading indicators, Sa2. The resulting model was estimated, and Sa2 and ALT were significantly and negatively associated with EXI, also as shown in Table A.

Next, dummy variables for the categories of Sa2 (i.e., category 1 = strongly dissatisfied, category 2 = dissatisfied, category 3 = neutral, category 4 = satisfied, and category 5 = very satisfied) were created. Specifically, in each case, Sat_Dummyi = 1 if Sa2 = i (i = 1 to 5) in that case. Otherwise, Sat_Dummyi = 0. Thus for example, in each

[1] Other study details are omitted to skirt matters such as conceptual and operational definitions, etc. that were judged to be of minimal importance to the present purposes.

case where Sa2 = 1, Sat_Dummy1 was assigned the value of 1. For those cases where Sa2 equaled some other value (i.e., 2, 3, 4 or 5), Sat_Dummy1 was assigned 0.

Equation 1 was altered to produce the structural model

$$\begin{aligned} EXI = {} & b_{11}Sat\_Dummy1 + b_{12}Sat\_Dummy2 + b_{13}Sat\_Dummy3 \\ & + b_{14}Sat\_Dummy4 + b_{15}Sat\_Dummy5 + b_2'ALT + \zeta', \end{aligned} \quad (2$$

where b_{1j} and b_2' (j = 1 to 5) are structural coefficients, and ζ' is structural disturbance.

Unfortunately, Equation 2 currently cannot be estimated satisfactorily using LISREL (or other popular covariant structure packages such as EQS, AMOS, etc.). The covariance matrix produced by the dummy variables is not positive definite.

The usual "remedy" is to estimate Equation 2 with one dummy variable omitted (e.g., Blalock 1979). However, this approach is unsatisfactory for theory testing because omitting a dummy variable in Equation 2 alters the significances of the remaining dummy variables, depending on which dummy variable is omitted. For example, see the significances in Tables B and C of Sat_Dummy2 or Sat_Dummy4 when Sat_Dummy1 or SAT_Dummy5 was omitted from Equation 2.

Ping (1996) proposed a latent variable estimation approach that will estimate Equation 2 without omitting dummy variables. The approach, Latent Variable Regression, adjusts the Equation 2 variance-covariance matrix for the measurement errors in ALT and EXI using Equation 2 measurement model loadings and measurement error variances. The resulting error-disattenuated variance-covariance matrix is then input to OLS regression. This approach was judged to be acceptably unbiased and consistent in the Ping (1996) article.

In order to use Latent Variable Regression to estimate Equation 2, the error-adjusted covariance matrix for Equation 2, and "regression through the origin" was used (to accommodate the collinearity of the dummy variables--see Blalock 1979). Specifically, the (error attenuated) variances and covariances of ALT, EXI and the five indicators for the SAT dummy variables were obtained using SPSS. Next, the measurement model for Equation 2 was estimated using the (consistent) indicators for ALT and EXI, and single indicators for the five SAT dummy variables (i.e., Sat_Dummy1, Sat_Dummy2, etc.), with LISREL and maximum likelihood estimation. Then, the loadings and measurement error variances from this measurement model were used to adjust the SPSS variances and covariances of ALT, EXI and the SAT dummy variables using equations proposed by Ping (1996) such as

$$\mathrm{Var}(\xi_X) = (\mathrm{Var}(X) - \theta_X)/\Lambda_X^2$$

and

$$\mathrm{Cov}(\xi_X,\xi_Z) = \mathrm{Cov}(X,Z)/\Lambda_X\Lambda_Z\ ,$$

where $\mathrm{Var}(\xi_X)$ is the error adjusted variance of X, Var(X) is the error attenuated variance of X (available from SAS, SPSS, etc.), $\Lambda_X = \mathrm{avg}(\lambda_{X1} + \lambda_{X2} + ... + \lambda_{Xn})$, avg = average, and $\mathrm{avg}(\theta_X = \mathrm{Var}(\varepsilon_{X1}) + \mathrm{Var}(\varepsilon_{X2}) + ... + \mathrm{Var}(\varepsilon_{Xn}))$, ($\lambda$'s and ε_X's are the measurement model loadings and measurement error variances--1 and 0 respectively--for the SAT dummy variables, and n = the number of indicators of the latent variable X), $\mathrm{Cov}(\xi_X,\xi_Z)$ is the error adjusted covariance of X and Z, and Cov(X,Z) is the error-attenuated covariance of X and Z.[2]

[2] These equations make the classical factor analysis assumptions that the measurement errors are independent of each other, and the x_i's are independent of the measurement errors. The indicators for X and Z must be consistent in the Anderson and Gerbing (1988) sense.

The resulting error-adjusted variance-covariance matrix for Equation 2 was then input to SPSS' OLS regression procedure, and the results are shown in Table D.

DISCUSSION

The dummy variables for categories 4 and 5 (Sat_Dummy4 and Sat_Dummy5) in Table D were nonsignificant, while the other dummy variables were significant. Comparing the Table D results for ALT to those from Tables B and C, the statistics for the coefficient of ALT were practically unaffected by omitting a single dummy variable, or by using regression through the origin. Also note that the unstandardized coefficient for Sat_Dummy2 with Sat_dummy1 omitted in Table B was Sat_Dummy2 - Sat_Dummy1 (within rounding), in Table D. Similarly, the unstandardized coefficient for Sat_Dummy3 with Sat_dummy1 omitted in Table B was Sat_Dummy3 - Sat_Dummy1, in Table D within rounding. Similarly, the other Table B dummy variables were the difference within rounding between their Table D value and Sat_Dummy1. For this reason Sat_Dummy1 is sometimes referred to as a "reference variable." In Table C Sat_Dummy5 is the reference variable for the unstandardized coefficient values shown there.

However, the interpretation or "meaning" of the unstandardized coefficients of the dummy variables is slightly different from the unstandardized coefficient of ALT. For example, the signs on the Table D unstandardized coefficients of the dummy variables have no meaning. Specifically, Sat_Dummy1, for example, is not "positively" associated with EXI. The unstandardized coefficient of Sat_Dummy1 is the absolute value of the change in the mean of EXI "caused" (associated) with Sat_Dummy1 (1.51). In this case, the mean of EXI changed in absolute value for the "very dissatisfied" category by the amount 1.51. Similarly, the absolute value of the change in EXI for the "very satisfied" (Sat_Dummy5) was .15. (Nevertheless, the "direction" of the associations of the set of dummy variables with EXI can be inferred--see below).

THE SATISFACTION-EXITING HYPOTHESIS

Satisfaction was hypothesized to be associated with Exiting, but so far we have estimated only variables derived from Satisfaction, its dummy variables, and several of them were nonsignificant. To estimate the effect of Satisfaction on Exiting using the dummy variables, the coefficients of the dummies were aggregated using a weighted average,[3] and the results are shown in Table E. Interpreting this aggregated result, Satisfaction, estimated as a categorical variable, was significantly associated with Exiting as hypothesized.

The "direction" of this association can be inferred by linearly ordering the unstandardized coefficients of the dummy variables from low to high. In this case EXI (i.e., its means in absolute value) decreased as the category represented by the dummy variables increased, which is consistent with the Table A result that Satisfaction is negatively associated with Exiting. Specifically, in the lower satisfaction categories, Sat_Dummy1 and Sat_Dummy2, Exiting was higher (i.e., the means were higher), while in the higher satisfaction categories, Sat_Dummy4 and Sat_Dummy5, Exiting was lower.

COMMENTS

[3] An overall F test of the "effect" of the dummies (e.g., $F = [(R_2^2 - R_1^2)/(k_2 - k_1)] / [(1- R_2^2)/(N - k_2 - 1)]$ where Ri^2 is R Square (or Squared Multiple Correlation), i = 1 denotes the model with the dummies omitted, i = 2 denotes the model with the dummies included, k_i is the number of exogenous variables (predictors), and N is the number of cases--see for example Jaccard, Turrisi and Wan, 1990) is inappropriate because R^2's are not comparable between intercept and no intercept models (Hahn 1977). (In addition, R^2 is usually incorrectly calculated in no intercept models--see Gordon 1981).

Anecdotally, it is believed that regression through the origin (no intercept regression) is biased, perhaps because its R^2 appears to be biased (Gordon 1981)--especially when the intercept is likely to be non-zero. However, experience suggests that with dummy variables, coefficient estimates are consistent between intercept regression and no intercept regression. For example, the coefficient estimates from omitting a dummy variable (which used intercept regression) such as those in Tables B and C were the difference between the omitted dummy variable's coefficient and the other coefficients in the no intercept results shown in Table D, within rounding). Also, the Equation 2 coefficients for ALT were practically unaffected by regression through origin (e.g., see Table B and Table D).

However, these results do not "prove" anything. They merely hint that the suggested approach may be useful for a categorical exogenous variable[4] in theory (hypothesis) tests of latent variable models with survey data. Nevertheless, as an additional example, ALT was estimated as a categorical variable with the results shown in Table F. These results paralleled those from estimating Satisfaction. For example, the Equation 2 coefficients for SAT were practically unaffected by regression through origin. Similarly, the coefficient estimates from omitting a dummy variable (which used intercept regression) were the difference between the omitted dummy variable's coefficient and the other coefficients in the no intercept results in Table F (within rounding). The "direction" of this effect was inferred by linearly ordering the unstandardized coefficients for the Alt_Dummy variables. As these means increased (al4 increased), Exiting increased, which is consistent with Equation F in Table F. Finally, the

---

[4] Blalock (1979) points out that multiple categorical variables cannot be jointly estimated using regression through the origin.

aggregated coefficient for the Alt4 dummy variables, "confirmed" the Equation F results that Alternatives, estimated as a categorical variable, was positively associated with Exiting as hypothesized.

However, the proposed approach is tedious to use. The adjusted variance-covariance matrix for Latent Variable Regression must be manually calculated. Similarly, the standard errors for Latent Variable Regression must be manually calculated (measurement model covariances could be substituted for the calculated covariances), and aggregation of the dummy variable coefficient results must be manually performed. (EXCEL templates are available from the authors for these calculations.)

The sample size of each dummy variable was a fraction of the total sample. Although most of the dummy variables were significant in the examples, the typically small sample sizes of survey data tests (e.g., about 200 cases) can produce one or more nonsignificant associations in the dummy variables because of their small subsamples. Thus, aggregation of the dummy variables is desirable to judge the overall significance of the categorical variable from which they were created.[5]

The suggested aggregation approach for the dummy variables was uninvestigated for any bias and inefficiency. Thus, the significance threshold for the aggregated coefficient of a categorical variable probably should be higher than the customary |t-value| = 2.0, where "| |" indicates absolute value. For example, since the significances of the Table E and Table F aggregated coefficients were materially larger than t = 2.0, they were judged to be significant.

---

[5] Parenthetically, the significance of the Sa2-Exiting association in Table A, for example, was very different from that of the aggregation of its dummy variables shown in Table E because the associations themselves were estimated differently.

Obviously, Latent Variable Regression is limited to a estimating a single dependent or endogenous variable, and it provides Least Squares coefficient estimates, rather than the preferred Maximum Likelihood estimates. However, Castella (1983) has proposed adding a "leverage data point" that "forces" a regression intercept of zero (i.e., the intercept or constant estimated in no-origin regression is zero). This leverage data point also may permit Equation 2, for example, to be estimated using covariant structure analysis and all its dummy variables with maximum likelihood.

Unfortunately, the results for the dummy variables are sensitive to how the dummy variables are coded. Changing the assignment of 1 for category exclusion (e.g., Sat_Dummy5 = 1 if Sa2 = 5, and Sat_Dummy5 = 0 otherwise) to -1, for example, for category inclusion, reverses the signs on all the dummy variables in the tables. However, this sensitivity to coding provides further "meaning" for dummy variables. Specifically, the absolute value of the unstandardized coefficient of Sat_Dummy5, for example, in Table D is the change from zero, the (arbitrary) category exclusion value for the dummy variable coding, for the mean of EXI that is "caused" (associated) with Sat_Dummy5 (.15). In this case, the mean of EXI changed in absolute value (from zero) for the "very satisfied" category by the amount .15. Similarly, the absolute value of the change in EXI (from zero) for the "very dissatisfied" (Sat_Dummy1) was 1.51.

The "direction" of a (truly) categorical association across its categories can be nearly impossible to hypothesize in the customary manner. For example, it is not obvious how one would argue, a priori, the "directionality" of any association between the eight VALS Psychographic categories (SRI International 1989) and Exiting using customary hypotheses such as

H2: VALS increases Exiting

or

H2': VALS decreases Exiting.

However, an hypothesis involving VALS and Exiting might be stated without "directionality" as

H2": VALS is associated with (affects) Exiting.

Any "directionality" could be inferred later from linearly ordering the resulting means (even if they were difficult to explain). Such an approach of disconfirming an association stated without directionality, then observing or "discovering" directionality is within the "logic of discovery" (e.g., Hunt 1983), so long as this "discovery" of directionality is presented as potentially an artifact of the study that must be hypothesized, theoretically supported, then disconfirmed in an additional study.

SUMMARY

The paper suggested an approach to estimating an exogenous (truly) categorical variable (e.g., Gender) in theory (hypothesis) tests of latent variable models with survey data. The approach involved using dummy variables for the categories, and Latent Variable Regression (Ping 1996). The dummy variable estimation results were aggregated to gauge the disconfirmation of a categorical variable hypothesis. The paper also suggested that associations between exogenous categorical variables and endogenous latent variables might be hypothesized without the customary "directionality" statement (that can be difficult to predict in categorical variables).

REFERENCES

Anderson, James C. and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 103 (May), 411-23.

Blalock, Hubert M., Jr. (1979), *Social Statistics*, New York: McGraw Hill.

Castella, G. (1983), "Leverage and Regression Through the Origin, The American Statistician, 37 (2) (May), 147-152.

Gordon, H. A. (1981), "Errors in Computer Packages: Least Squares Regression through the Origin," *The Statistician*, 30, 23-29.

Hahn, G. (1971), "Fitting regression Models with no Intercept Term," *Journal of Quality Technology*, 9, 56-61.

Hunt, Shelby D. (1983), *Marketing Theory: The Philosophy of Marketing Science*, Homewood, IL: Richard D. Irwin.

Jaccard, James, Robert Turrisi and Choi K. Wan (1990), *Interaction Effects in Multiple Regression*, Newbury Park, Ca: Sage Publications.

Jöreskog, Karl G. and Dag Sörbom (1996), *Lisrel 8 User's Reference Guide*, Chicago: Scientific Software International, Inc.

Ping, R. (1996), "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients," *Multivariate Behavioral Research*, 31 (1), 95-120.

______ (2001), "A Suggested Standard Error for Interaction Coefficients in Latent Variable Regression," *2001 Academy of Marketing Science Conference Proceedings*, Miami: Academy of Marketing Science.

SRI International (1989), *The VALS 2 Segmentation System*, Menlo Park, CA: SRI International.

Table A--Abbreviated Equation 1[a] Estimation Results[b]

Equation 1 Estimation Results:

Endog	EXI		
Exog	Unstd		
Variable	Str Coef	SE[c]	t-value
ALT	0.63	0.11	-5.93
SAT	-0.58	0.10	-5.76

Equation 1 with Sa2[d] Estimation Results:

Exog	Unstd		
Variable	Str Coef	SE[c]	t-value
ALT	0.69	0.07	9.65
SA2	-0.42	0.05	-7.79

---

[a] $EXI = b_1 SAT + b_2 ALT + \zeta$.

[b] The structural models were judged to fit the data. Estimates involved LISREL and maximum likelihood.

[c] SE is Standard Error.

[d] $EXI = b_1' Sa2 + b_2' ALT + \zeta''$.

Table B--Abbreviated Equation 2 Results with Sat_Dummy1 Omitted[a b]

Exog Variable	Unstd Str Coef	SE[c]	t-value
ALT	0.68	0.07	9.92
Sat_Dummy3	-0.98	0.26	-3.79
Sat_Dummy4	-1.30	0.26	-5.05
Sat_Dummy2	-0.02	0.26	-0.06
Sat_Dummy5	-1.36	0.28	-4.87

____

[a] $EXI = \cancel{b_{11}Sat\_Dummy1} + b_{12}'Sat\_Dummy2 + b_{13}'Sat\_Dummy3 + b_{14}'Sat\_Dummy4 + b_{15}'Sat\_Dummy5 + b_2''ALT + \zeta'''$

[b] The structural model was judged to fit the data. Estimates involved LISREL and maximum likelihood.

[c] SE is Standard Error.

Table C--Abbreviated Equation 2 Results with Sat_Dummy5 Omitted[a b]

Exog Variable	Unstd Str Coef	SE[c]	t-value
ALT	0.68	0.07	9.92
Sat_Dummy3	0.38	0.15	2.48
Sat_Dummy1	1.36	0.28	4.87
Sat_Dummy4	0.06	0.12	0.51
Sat_Dummy2	1.34	0.18	7.53

---

[a] $EXI = b_{11}'\text{Sat\_Dummy1} + b_{12}''\text{Sat\_Dummy2} + b_{13}''\text{Sat\_Dummy3}$
$+ b_{14}''\text{Sat\_Dummy4} + \cancel{b_{15}\text{Sat\_Dummy5}} + b_{2}'''\text{ALT} + \zeta''''$

[b] The structural model was judged to fit the data. Estimates involved LISREL and maximum likelihood.

[c] SE is Standard Error.

Table D-- Abbreviated Equation 2 Results with all the Sa2 Dummy Variables[ab]

Exog Variable	Unstd Str Coef	SE[c]	t-value
ALT	0.68	0.07	9.92
Sat_Dummy1	1.51	0.36	4.20
Sat_Dummy5	0.15	0.18	0.85
Sat_Dummy3	0.54	0.24	2.21
Sat_Dummy2	1.50	0.28	5.37
Sat_Dummy4	0.22	0.20	1.09

____

[a] EXI = b_{11}"Sat_Dummy1 + b_{12}"'Sat_Dummy2 + b_{13}"'Sat_Dummy3
+ b_{14}"'Sat_Dummy4 + b_{15}'Sat_Dummy5 + b_2""ALT + ζ"""

[b] Estimation involved Latent Variable Regression with least squares.

[c] The standard error (SE) is from Ping (2001).

Table E--Aggregation Results[a b]

Case Weighted Average of the Unstandardized Structural Coefficients	Standard Error of the Case Weighted Average of the Unstandardized Structural Coefficients	T-value of Case Weighted Average of the Unstandardized Structural Coefficients
0.51	0.21	2.47

---

[a] The Case Weighted Average is $\Sigma w_i b_i$, where Σ is summation, i = 1 to 5, w_i is the weighted average (number of cases in category i divided by the total number of cases) of the unstandardized coefficient, b_i , of Sat_Dummyi, and i = 1 to 5.

[b] The aggregated Standard Error is the Square Root of the variance of the weighted sum of the individual standard errors (e.g., $sqrt(Var(w_1SE_1 + w_2SE_2 + w_3SE_3 + w_4SE_4 + w_5SE_5) = sqrt(\Sigma w_i^2 SE_i^2 + 2(\Sigma Cov(SE_i,SE_j)))$, where "sqrt" is the square root, Var is variance, w_i is the weighted average (number of cases in category i divided by the total number of cases) of the unstandardized coefficient of Sat_Dummyi, SE is standard error, Σ is summation, i = 1 to 5, j = 2 to 5, and i > j.

Table F--Abbreviated Estimation Results[a] for ALT as a Categorical Variable

a) Abbreviated Equation 1 with Al4[b] Estimation Results: [c]

Exog Variable	Unstd Str Coef	SE[d]	t-value
SAT	-0.57	0.07	-8.59
AL4	0.52	0.05	9.69

b) Abbreviated Equation 2 Results with Al4[b] and Alt_Dummy1 Omitted [e]

Exog Variable	Unstd Str Coef	SE[d]	t-value
SAT	-0.56	0.07	-8.56
Alt_Dummy3	0.67	0.18	3.85
Alt_Dummy5	1.80	0.26	7.02
Alt_Dummy4	1.56	0.20	7.72
Alt_Dummy2	0.27	0.17	1.61

c) Abbreviated Equation 2 Results with Al4[b] and Alt_Dummy5 Omitted [f]

Exog Variable	Unstd Str Coef	SE[d]	t-value
SAT	-0.56	0.07	-8.56
Alt_Dummy3	-1.12	0.19	-5.83
Alt_Dummy1	-1.80	0.26	-7.02
Alt_Dummy4	-0.23	0.19	-1.21
Alt_Dummy2	-1.53	0.21	-7.42

d) Abbreviated Equation 2 Results with All Al4[b] Dummy Variables[g]

Exog Variable	Unstd Str Coef	SE[d]	t-value
SAT	-0.56	0.07	-8.56
Alt_Dummy5	5.70	0.23	24.45
Alt_Dummy4	5.47	0.23	23.89
Alt_Dummy1	3.90	0.34	11.54
Alt_Dummy3	4.58	0.26	17.90
Alt_Dummy2	4.17	0.29	14.62

Table F (con't.)--Abbreviated Estimation Results[a] for ALT as a Categorical Variable

e) Aggregation Results [h i]

Case Weighted Average of the Unstandardized Structural Coefficients	Standard Error of the Case Weighted Average of the Unstandardized Structural Coefficients	T-value of Case Weighted Average of the Unstandardized Structural Coefficients
4.48	0.27	16.86

---

[a] The structural models were judged to fit the data. Exhibits a) through c) involved LISREL and maximum likelihood estimates, and Exhibit d) involved Latent Variable Regression and least squares estimates.

[b] Al4 was the heaviest loading indicator of ALT.

[c] $EXI = b_1''SAT + b_2''Al4 + \zeta''''$.

[d] SE is Standard Error.

[e] $EXI = b_1'''SAT + \cancel{b_{21}Alt\_Dummy1} + b_{22}Alt\_Dummy2 + b_{23}Alt\_Dummy3$
$+ b_{24}Alt\_Dummy4 + b_{25}Alt\_Dummy5 + \zeta'''''$.

[f] $EXI = b_1''''SAT + b_{21}Alt\_Dummy1 + b_{22}'Alt\_Dummy2 + b_{23}'Alt\_Dummy3$
$+ b_{24}'Alt\_Dummy4 + \cancel{b_{25}Alt\_Dummy5} + \zeta''''''$.

[g] $EXI = b_1'''''SAT + b_{21}'Alt\_Dummy1 + b_{22}''Alt\_Dummy2 + b_{23}''Alt\_Dummy3$
$+ b_{24}''Alt\_Dummy4 + b_{25}''Alt\_Dummy5 + \zeta'''''''$

[h] The Case Weighted Average is $\Sigma w_i b_i$, where Σ is summation, $i = 1$ to 5, w_i is the weighted average (number of cases in category i divided by the total number of cases) of the unstandardized coefficient, b_i , of Alt_Dummyi, and $i = 1$ to 5.

[i] The aggregated Standard Error is the Square Root of the variance of the weighted sum of the individual standard errors (e.g., $sqrt(Var(w_1SE_1 + w_2SE_2 + w_3SE_3 + w_4SE_4 + w_5SE_5) = sqrt(\Sigma w_i^2 SE_i^2 + 2(\Sigma Cov(SE_i,SE_j)))$, where "sqrt" is the square root, Var is variance, w_i is the weighted average (number of cases in category i divided by the total number of cases) of the unstandardized coefficient of Alt_Dummyi, SE is standard error, Σ is summation, $i = 1$ to 5, $j = 2$ to 5, and $i > j$.

INTERACTIONS MAY BE THE RULE RATHER THAN THE EXCEPTION, BUT . . . : A NOTE ON ISSUES IN ESTIMATING INTERACTIONS IN THEORETICAL MODEL TESTS

ABSTRACT

Authors have called for more frequent investigation of interactions in theoretical models involving survey data. However, there are competing proposals for interaction specification in structural equation models. And, there are other interaction issues that have received little or no attention in theoretical model testing. For example, what types of evidence suggest that an interaction should be hypothesized? Is an interaction a construct or a mathematical form, or both? There are theoretical and practical issues involving the existing structural equation estimation proposals for interactions. For example, specifying the interaction between X and Z, for example, as XZ is an insufficient disconfirmation test. This paper critically addresses these and other theory-testing matters in conceptualizing, estimating and interpreting interactions in survey data.

Authors have noted that in some disciplines interactions may be the rule rather than the exception in survey data models (e.g., Jaccard, Turrisi and Wan 1990). However, interactions in published theoretical model tests with survey data are comparatively rare (Aiken and West 1991). Perhaps as a result, recent conference keynote speakers have called for increased investigations of substantive interactions. While infrequent appearances of interactions in substantive structural equation papers may have been the result of the unavailability of suitable analysis tools until recently, substantive researchers may not be accustomed to conceptualizing and estimating interactions. Theoretical model building with an interaction involves conceptualizing at least three variables, developing theory to justify two variables' proposed associations with a target variable, then developing additional theory for the variability (form) of least one of these relationships. In ANOVA, interactions are usually estimated en masse after the main effects are estimated. In this event, theorizing about interactions occurs after they are found to be significant, if at all.

The "value" added by the increased effort required for theorizing and assessing interactions may seem comparatively low, especially for theoretical models involving "interesting" new constructs or "interesting" new paths among relationships. There is also the tedium of interaction specification using the existing structural equation estimation proposals.

Anecdotally, some authors believe interactions should not appear in theoretical models because they are not "proper" constructs. They are not "indicated" (pointed to) by observed variables.

Further, specifying an hypothesized interaction between X and Z, for example, as the product of X and Z, XZ is an insufficient test of the hypothesized interaction. XZ is but one of many interaction forms (Jaccard, Turrisi and Wan 1990).

THE PRESENT RESEARCH

This paper critically addresses several matters related to interactions in theoretical model tests involving survey data. For example, it discusses foreseeing the plausibility of an interaction at the model-building stage, then justifying these interactions theoretically. It also discusses how interaction hypotheses are phrased, and it discusses issues involving the existing estimation proposals for interactions such as a lack of an adequate interaction disconfirmation test. Along the way it discusses matters such as interpreting interactions in survey data, probing for interactions after the hypothesized model has been estimated, and the "trap" of hypothesizing interactions that are found post hoc as though they were hypothesized before the model was first estimated.

CONCEPTUALIZING INTERACTIONS

We will discuss each of these matters, beginning with conceptualizing or envisioning interactions--foreseeing their plausibility at the model-building stage.

Because interactions are usually characterized as "moderators," this might hinder their theory development. The term "moderator" might signal that X, for example, reduces the Z-Y association as X increases, and that interactions are restricted to this case. However, X actually might reduce the Z-Y association as X decreases. In this case X amplifies the Z-Y assoc. There are other forms of the interaction meaning of "moderation" as well. For example, X could reduce the Z-Y association at one end of the range of X in the study, and it could amplify the Z-Y association at the other end. These interactions are termed disordinal interactions.

Thus, it may be useful to resist thinking of interactions as moderators in the development of a model. Instead, it may be fruitful to think of what might happen to the Z-Y relationship, for example, when X was at a low level versus when X was high. For example, it is well known that relationship satisfaction reduces relationship exiting, and attractive alternatives increase exiting (e.g., Ping 1993). However, what would happen if a data set were split at the median of satisfaction, and the strength of the alternatives-exiting association were compared between the two split halves? Alternatives should increase exiting when satisfaction is lower, but when satisfaction is higher, the effect of alternatives on exiting should be lower or non significant (see Ping 1994).

While this split-halves approach has been disparaged for interaction estimation (e.g., Lubinski and Humphreys 1990) it may be a fruitful "though experiment" for conceptualizing interactions. Because of concerns about model parsimony, this thought experiment should probably be restricted to major constructs, but for the "most important" pair of variables that should be related, could this relationship plausibly change in strength or direction between low and high values of a third variable? Specifically, for low values of X, for example, should the Z-Y association somehow be different from the Z-Y association when X is high?

Obviously this though experiment could be conducted on new models. It also could be conducted on previously investigated models that did not consider interactions. Specifically, because a disordinal interaction could have caused a main effect to be non significant in a published study (see Aiken and West 1991), previous studies with non-significant hypothesized associations might be fruitfully considered for the above thought experiment. A disordinal interaction also could have caused a main effect to be positive in one study and negative in another (see Aiken and West 1991). Thus, previous studies with an association that is not consistently significant, or not consistent in sign might be fruitfully considered for the thought experiment. It even might be fruitful to consider the major associations in a previously investigated model for the thought experiment, even if they have been consistently significant and in the same direction.

There is an additional way to identify interactions for theory development (and estimation in a future study): post hoc probing for them as in ANOVA. This matter will be discussed later.

JUSTIFYING INTERACTIONS

However, an interaction can be challenging to justify theoretically. While existing theory might be available to directly support a plausible interaction, it is more likely that there has been little previous thought about the target interaction. This lack of previous though about a topic has been a hallmark of science, and researchers have used many strategies to construct explanations for the topic under study. While these can include deduction, induction and abduction (see Peirce 1931–1935, 1958), in general, researchers use any sort of evidence, including direct experience such as exploratory focus groups, to support a proposed interaction.

For example, and as previously mentioned, satisfaction and alternatives both affect exiting. However, Ping (1994) argued that satisfaction attenuates the alternative-exiting association. To justify this hypothesis he used prior arguments that at high-satisfaction subjects were not aware of alternatives (Dwyer, Schurr and Oh 1987) or they devalued them (Thibaut and Kelly 1959). He also noted that alternatives previously had been argued to reduce exiting, and that argument had been empirically "confirmed." To resolve this paradox, he proposed the interaction. In this case he used existing arguments about high satisfaction, existing theory and prior results about alternatives-exiting, and a proposal that the prior alternatives-increases-exiting results likely applied to lower satisfaction samples to justify a proposed interaction. These results also might have been found in focus groups of low and high satisfaction subjects.

However, it usually is insufficient to use "experience" and previous writings as justification; a "why" must be supplied. For example, at low satisfaction increasing alternatives are likely to increase exiting because with reduced satisfaction (reduced relationship rewards) the alternatives' rewards may appear more attractive than the current relationship's. At high satisfaction, alternatives are not likely to be associated with exiting because the effort (cost) to compare rewards is unnecessary, or the alternative's rewards are less than certain (a risk). In general, rewards and cost (see for example Shaw and Costanzo 1982), and risk (Kahneman and Tversky 1979) have been used to justify considerable research involving human behavior, and they might continue to be useful for interactions. Examples of interaction justification can be found in Aiken and West (1991), and the citations therein, Ajzen and Fishbein (1980), Kenny and Judd (1984) and Ping (1994, 1999).

HYPOTHESIZING INTERACTIONS

Given that X, for example, is argued to increase the Z-Y association, should the hypothesis be stated as "X moderates the Z-Y association?" Terms such as "interacts with," "modifies," "amplifies" or "increases" would be more precise. Specifically,

> H: X interacts with/modifies/amplifies/increases the Z-Y association

would be more precise. However, it still may be insufficient, especially if the low-high though experiment and justification approach suggested above is used. In this case

> H: At low X, the Z-Y association is comparatively weak, while at high X the Z-Y association is stronger,

would fit a low-high argument.

AN EXAMPLE As previously mentioned, attractive alternatives are likely to increase relationship exiting. Relationship dissatisfaction amplifies this alternatives-exiting relationship. Specifically, when dissatisfaction is low the positive alternatives-exiting association should be weak (small, possibly non significant). However, when dissatisfaction is high, the alternatives-exiting association should be stronger (larger compared to the low dissatisfaction coefficient).

Thus, in this case a "complete" interaction hypothesis would be

H1a: Dissatisfaction is positively associated with exiting,

H1b: Alternatives are positively associated with exiting, and

H1c: Dissatisfaction moderates/interacts with/attenuates/reduces the alternatives-exiting association.

Alternatively,

H1c': As dissatisfaction increases, the alternatives-exiting association becomes weaker.

Note that the interaction hypothesis is accompanied by two other hypotheses involving exiting with satisfaction and alternative, and that "moderates," meaning "to reduce' is appropriate.

Instead, one might hypothesize

H1c": When dissatisfaction is low the alternatives-exiting association is weaker than it is when dissatisfaction is higher.

Obviously, an equivalent interaction hypothesis statement would be "as dissatisfaction declines, the alternatives-exiting association becomes weaker." However, this may not match the above argument quite as well as H1c-H1c". It would match the above argument if the "direction" of the argument were reversed (i.e., "when dissatisfaction is high the positive alternatives-exiting association should be strong (large), but when dissatisfaction is lower, the alternatives-exiting association should be weaker (smaller, possibly non significant)).

A property of the dissatisfactionXalternatives interaction that is useful in justifying interactions and framing their hypotheses is their symmetry. To explain, an abbreviated structural equation involving dissatisfaction (DISSAT), alternatives (ALT), and exiting (EXIT) would be

$$\text{EXIT} = a\ \text{DISSAT} + b\ \text{ALT} + c\ \text{DISSATxALT}\ . \qquad (1$$

Factoring,

$$\text{EXIT} = a\ \text{DISSAT} + (b + c\ \text{DISSAT})\text{ALT}\ . \qquad (2$$

In words, since b and c are constants, as DISSAT changes from subject to subject in the study, the structural coefficient of ALT, b+c DISSAT, changes.

However, Equation 1 could be re factored into

$$\text{EXIT} = b\text{ALT} + (a + c\ \text{ALT})\ \text{DISSAT}\ . \qquad (3$$

Thus, as ALT changes, the structural coefficient of DISSAT, a+cALT, changes. Thus, ALT interacts with DISSAT.

In general, if X interacts with Z in the Z-Y association, then Z interacts with X in the X-Y association. In the DISSATxALT case it turns out that it is easier to argue that increasing alternatives increases the dissatisfaction-exiting association, so an interaction hypothesis such as

H1c''': Alternatives moderate/interact with/attenuate/reduce the dissatisfaction-exiting association,

H1c'''': As alternatives increase, the dissatisfaction-exiting association becomes stronger, or

H1c''''': When alternatives are low the dissatisfaction-exiting association is weaker than it is when alternatives is higher,

is appropriate.

For emphasis, two thought experiments are possible with DISSAT: what happens to the DISSAT-EXIT association as ALT changes, and what happens to the ALT-EXIT association as DISSAT changes?

INTERACTION COST-BENEFITS

The "value" of the increased effort required for theorizing and determining interaction effects may be comparatively low, especially for theoretical models involving "interesting" new constructs or new relationships among constructs.

Specifically, and as mentioned earlier, an interaction involves theory development for two variables' association with a target variable, and additional theory development for the variability (form) of least one of these relationships. In addition, interactions typically explain comparatively little additional variance (Cohen and Cohen 1983). Specifically, in Equation 1, adding DISSATxALT explains little additional variance in EXIT. However, in theory testing it is more important to know how an interaction affects a target relationship, and thus the behavior of its factored coefficient (e.g., b+cDISSAT in Equation 2), rather than to explain lots of additional variance. (Materially explaining additional variance is important in model building, e.g., epidemiology.) Parenthetically, experience suggests that even for an interaction that explains comparatively little additional variance, a factored coefficient such as (b+cDISSAT) can be quite large for some values of DISSAT.

Interactions reduce parsimony. Specifically, adding an interaction decreases degrees of freedom, and increases model collinerarity. However, if there is strong theoretical justification for an interaction, it is likely the interaction will be significant in any reasonably sized sample. In addition, an hypothesized interaction's collinerarity is an important part of a model's test.

Papers with interactions tend to be overly methods oriented, and the interactions tend to dominate the paper. One method to reduce the appearance of a "methods paper" is to place the interaction details in an appendix. Considering interaction(s) for major exogenous constructs only also should reduce their apparent "dominance" in a paper.

PROPOSED APPROACHES Unfortunately, issues with structural equation estimation approaches for interactions may further reduce their apparent value. Specifically, there are several proposals for specifying latent variable interactions including (1) Kenny and Judd 1984; (2) Bollen 1995; (3) Jöreskog and Yang 1996; (4) Ping 1995; (5) Ping 1996a; (6) Ping 1996b; (7) Jaccard and Wan 1995; (8) Jöreskog 2000; (9) Wall and Amemiya 2001; (10) Mathieu, Tannenbaum and Salas 1992; (11) Algina and Moulder 2001; (12) Marsh, Wen and Hau 2004; (13) Klein and Moosbrugger 2000/Schermelleh-Engle, Kein and Moosbrugger 1998/Klein and Muthén 2002; and (14) Moulder and Algina 2002.

They are all very tedious to use, most are inaccessible to substantive researchers (Cortina, Chen and Dunlap 2001), and some do not involve Maximum Likelihood estimation, or commercially available estimation software (proposals 2, 6 and 13).

Several of these proposals have not been formally evaluated for bias and inefficiency (i.e., proposals 8 and 10). In additional proposal 10 did not perform well in a comparison of interaction estimation approaches (see Cortina, Chen and Dunlap 2001).

Most of these proposals are based on the Kenny and Judd product indicators (for example, x_1z_1, x_1z_2, ... x_1z_m, x_2z_1, x_2z_2, ... x_2z_m, ... x_nz_m, where n and m are the number of indicators of X and Z respectively). However, specifying all the Kenny and Judd product indicators usually produces model-to-data fit problems (e.g., Jaccard and Wan 1995).

Several proposals use weeded subsets of the Kenny and Judd (1984) product indicators or indicator aggregation to avoid these inconsistency problems (proposals 3, 4, 5, 7, 9, 11, 12 and 14). Unfortunately, weeding the Kenny and Judd product indicators raises questions about the face or content validity of the resulting interaction (e.g., if all the indicators of X and Z are not represented in the indicators of XZ, for example, is XZ still the product of X and Z as they were operationalized in the study?) (proposals 3, 7, 9, 11, 12 and 14). In addition, the formula for the reliability of a weeded XZ is unknown. Specifically, the formula for the reliability of XZ is a function of (unweeded) X and unweeded Z, and thus it assumes XZ is operationally (unweeded) X times (unweeded) Z. Weeded Kenny and Judd product indicators also produce interpretation problems using factored coefficients because XZ is no longer (unweeded) X times (unweeded) Z operationally (see Equation 2).

Finally, although proposal 4 has none of these drawbacks except that it is tedious, and it assumes the loadings are tau equivalent. Proposal 4's tediousness may be reduced using the specification templates at http://home.att.net/~rpingjr/research1.htm. Although the tau equivalency assumption can be removed using weighting, experience with real-world data suggests that interaction significance is not particularly sensitive to this assumption.

INTERPRETING INTERACTIONS

Once an interaction in survey data is estimated, how should one interpret it? Two point graphical techniques that are used in ANOVA ignore much of the information available in survey data. For example, authors have noted that interaction significance varies (Aiken and West 1991, Jaccard, Turrisi and Wan 1990).

There have been several proposals for interpreting regression interactions (e.g., Aiken and West 1991; Darlington 1990; Denters and Van Puijenbroek 1989; Friedrich 1982; Hayduk 1987; Hayduk and Wonnacott 1980; Jaccard, Turissi and Wan 1990; Stolzenberg 1979). However, there is little guidance for interpreting latent variable interactions. The following presents an approach adapted from Friedrich's (1982) suggestions for interpreting interactions in regression (see Darlington 1990; Jaccard, Turrisi and Wan 1990).

AN EXAMPLE To explain this suggested interpretation approach, a real-world, but disguised, survey data set will be analyzed. The abbreviated results of a LISREL 8 Maximum Likelihood estimation of a structural model is shown in Table A. There the XZ interaction is large enough to warrant interpretation (i.e., its coefficient b_{XZ} is significant). Interpretation of this interaction relies on tables such as Table B that are constructed using factored coefficients such as the factored coefficient of Z, $b_Z + b_{XZ}X$, from Table A. Column 2 in Table B, for example, shows the factored coefficient of Z from Table A (.047 - .297X) at several Column 1 levels of X in the study. Column 3 shows the standard errors of these factored coefficients of Z at the Column 1 levels of X, and Column 4 shows the resulting t-values. Footnotes b) through d) in Table B further explain the Columns 1-4 entries. In particular, Footnote b) explains how values for the unobserved variable X are determined by the values of its indicator that is perfectly correlated with it (i.e., the indicator of X with a loading of 1). In addition, Footnote d) discusses the Standard Error of the variable Z coefficient. The variance of b is of course the square of the Standard Error of b, and Cov(b_Z,b_{XZ}), the covariance of b_Z and b_{XZ}, is equal to r(b_Z,b_{XZ})SE(b_Z)SE(b_{XZ}), where r is the "CORRELATIONS OF ESTIMATES" value for b_Z and b_{XZ} in LISREL 8, and SE indicates Standard Error.

Footnote a) of Table B provides a verbal interpretation of the moderated Z-Y association shown in Columns 2 and 4. Specifically, when the level of X was low in the study (i.e., 1.2 to above 3 in Column 1), small changes in Z at a particular value of low X were positively associated with Y (i.e., the coefficient of Z was 0.89-- see Column 2). However, as X increased in the study, Z was less strongly associated with Y (i.e., the column 2 Z coefficient declined), until near the study average for X, 4.05, the Z-Y association was non significant (e.g., at X = 4.05 the Z-Y association was 0.04, t = 0.59-- see Columns 1, 2 and 4). Then, when the level of X was above the study average, Z was again significantly associated with Y (i.e., the Z-Y association was -0.23, t = -2.48-- see Columns 1, 2 and 4- small changes in Z at a particular value of high X were negatively associated with Y).

Because there are always two factored-coefficients produced by a significant interaction (e.g., Equations 2 and 3), and the XZ interaction was large enough to warrant the Table B attention, Columns 5-8 are provided in Table B to help interpret the factored coefficient of X, -.849 - .297Z. Column 6 shows this factored coefficient at several Column 5 levels of Z. Column 7 shows the standard errors of this factored coefficient at these levels of Z, and Column 4 shows the resulting t-values. Again, additional information regarding Columns 5-8 is provided in Footnotes e) through i), and Footnote

e) provides a verbal summary of the moderated X-Y association produced by the significant XZ interaction in Table A.

DISCUSSION Notice that in Table A b_Z was non significant (t = 0.59), yet the factored coefficient of Z was significant at both ends of the range of X in the study (see Column 4 of Table B). Z had a negative association with Y when the existing level of X was high or above its study average in the sample, but its association with Y was positive when X was lower or below its study average. Similarly, b_X was significant (t= -5.32), yet the coefficient of X moderated by Z was non significant when Z was very low in the sample (see Column 8 of Table B).

Finally, notice that b_Z was not significant in Table A, yet it was included in the Table B, Columns 2 and 3, calculations. This was done because if Z is excluded, the t-value of the factored coefficient of Z is singular at the mean of X (i.e., it is undefined, and in a neighborhood of the mean of X the t-value of the factored coefficient approaches infinity).

INSUFFICIENT DISCONFIRMATION

Unfortunately, specifying an hypothesized interaction as XZ is an insufficient disconfirmation test of the interaction hypothesis. XZ is one of many interaction forms (see Jaccard, Turrisi and Wan 1995). Specifically, there are at least a countable infinity more possible mathematical forms an interaction can take besides XZ. Specifically, XZ^w, where W can be any (positive or negative) real number, is an interaction. This interaction form includes not only XZ (w = 1), it also includes X/Z (see Jaccard, Turrisi and Wan 1995) (w = -1). It also includes XZ^2, the interaction between X and the square of Z (see

Aiken and West 1991), and it curiously includes XX^w, where Z = X and X is moderated by itself (which is called a quadratic when w = 1) (see Lubinski and Humphreys 1990). Thus, an hypothesized population interaction may not have the form XZ. And, specifying it as XZ may produce non-significant results. In this case, an erroneous conclusion that there is no interaction between X and Z also may result.

Unfortunately, latent variable interaction specifications besides XZ are unknown at present. However, as a post hoc test it may be efficacious to use a median split (a subgroup analysis) of the data to test for the hypothesized interaction. While this test is fallible (Ping 1996c observed 8% false positives with subgroup analysis), if this produces a significant interaction, it suggests the hypothesized interaction may have been "confirmed" (i.e., in was confirmed in this test only), but its form is unknown at present.

INTERACTION EPISTEMOLOGY

Anecdotally, some authors believe interactions should not be included in model because they are not constructs--they are "indicated" (pointed to) by products of observed variables, and these product variables cannot be observed. Again anecdotally, interactions have implausible reliabilities, and the validity criteria do not apply. Thus, interactions are inappropriate for the structural equation models used to estimate theoretical models.

It is difficult to argue that interactions are constructs in the usual sense. However, they are "construct-like": they have reliability and aspects of validity.

RELIABILITY The reliability of interaction XZ is

$$\rho_{XZ} = \frac{\mathrm{Corr}^2_{X,Z} + \rho_X\rho_Z}{\mathrm{Corr}^2_{X,Z} + 1}$$

(Bohrnstedt and Marwell 1978, see Busemeyer and Jones 1983), which produces the implausible result that the reliability of XZ increases as the correlation between X and Z increases. However, this result parallels the result that the reliability of X is a function of the sum of the correlations of the indicators of X, and it increases as the correlation between the indicators increases.

VALIDITY It could be argued that most of the usual validities do apply to interactions. Authors disagree on what constitutes an adequate set of validity criteria (e.g., Bollen 1989, Campbell 1960, DeVellis 1991, Heeler and Ray 1972, Nunnally 1978, Peter 1981). Nevertheless, a minimal demonstration of validity in theory testing might include content or face validity (how well a latent variable's indicators tap its conceptual definition), construct validity (its correlations with other latent variables are theoretically sound), convergent validity (e.g., its average extracted variance is greater than 0.5--Fornell and Larker 1981), and discriminant validity (e.g., its correlations with other measures are less than 0.7, or) (e.g., Bollen 1989, DeVellis 1991, Fornell and Larker 1981, Nunnally 1978). The validity of a measure is then qualitatively assessed considering reliability and the measure's performance over this minimal set of validity criteria.

An interaction XZ is content or face valid if X and Z are content valid and the specification of XZ includes all the indicators of X and Z. (Without all the indicators of X accounted for in the itemization XZ, it does not specify the interaction between X and Z because the two X's are different constructs.) The formula for the Average Variance Extracted (AVE) of XZ is $\Sigma(\lambda_{xi}\lambda_{zj})^2\mathrm{Var}(XZ)/[\Sigma(\lambda_{xi}\lambda_{zj})^2\mathrm{Var}(XZ) + \Sigma\mathrm{Var}(\varepsilon_{xz})]$, where $\Sigma(\lambda_{xi}\lambda_{zj})^2$ is the sum of squares of $\lambda_{xi}\lambda_{zj}$, i = 1 to m, j = 1 to n, m is the number of indicators of X, n is the number of indicators of Z, and Var(XZ) is the error-dissattenuated variance of XZ (available in the structural model) (Fornell and Larker 1981). However, the construct (correlational) validity of a second-order is usually impossible to judge.

Thus, interactions may or may not be constructs as they are usually defined. However, structural models always contain one or more exogenous variables, structural disturbances, that are specified as a model variable yet they are not "indicated" or "pointed to" as a construct would be. Thus, structural models never contain just constructs as they are usually defined. Also, while not everyone would agree, Bollen (1989:268) states, " . . . *the model . . . helps us to understand the relations between variables . . . "* This suggests that the objective of theoretical model testing is to test hypothesized relations between variables. It also could be argued that constructs are important in structural equation analysis only to enable covariant structural analysis, and, given that they are sufficiently present to enable that end, that relationships, including moderated relationships, are the primary focus of theoretical model testing.

POST HOC PROBING

Obviously interactions can be "discovered" post hoc (i.e., after the hypothesized model has been estimated for the first time) as they are in ANOVA. Significant interactions could be used to improve interpretation of study results. Specifically, significant unhypothesized interactions may provide plausible explanations for hypothesized but non-significant first-order associations (main effects), which avoids casting a shadow on the relevant theory (a non-significant result suggests the relevant theory does not apply), and it improves the interpretation of significant associations that are actually conditional.

Unfortunately, survey researchers are discouraged from post-hoc probing for interactions (e.g., Aiken and West 1991, Cohen and Cohen 1983) on grounds this is unscientific because these variables were not hypothesized. However, the logic of model testing and its variables can easily be separated from the logic of discovery and its variables (e.g., interactions) (e.g., Hunt 1983) as long as any discovered interactions are clearly presented as "discovered." Specifically, any interactions discovered in post-hoc probing should be presented as potentially an artifact of the sample: Their existence in any population and thus in other samples/studies should be viewed as an empirical question to be answered in later studies.

Because of the potential for detecting spurious interactions (unhypothesized interactions that do not exist in the population and are significant by chance in the sample), an F-test is desirable to determine if any unhypothesized interactions are likely to be significant above the level of chance. To accomplish this, after the hypothesized structural model has been estimated, all possible interactions should be added to the hypothesized model.

AN F-TEST To reduce the likelihood of spurious (chance) interactions, the increase in R^2 (e.g., the "Squared Multiple Correlations for Structural Equations" in LISREL) due to adding all implied interactions to a model should be significant. A test statistic that assesses this increase is

$$F = [(R_2^2 - R_1^2)/(k_2 - k_1)] / [(1- R_2^2)/(N - k_2 - 1)]$$

where R_2^2 is the total explained variance (Squared Multiple Correlations for Structural Equations) in the structural model with the interactions added, R_1^2 is the total explained variance in the structural model with no interactions added, k_1 is the number of exogenous variables (predictors) in the structural model without the interactions, k_2 is the number of exogenous variables in the structural model plus the number of interactions added, and N is the number of cases (see for example Jaccard, Turrisi and Wan 1990). This F statistic has $k_2 - k_1$ and $N - k_2 - 1$ degrees of freedom.

Calculating F With a single endogenous or dependent variable (e.g., in the structural model $Y = b_1X + b_2Z + b_3W + \zeta$, Y is the endogenous or dependent variable and there are three exogenous variables or predictors of Y on the right-hand side of the equal sign) the F statistic is a straightforward calculation.

With multiple dependent or endogenous variables the suggested F-test is performed multiple times, once for each endogenous variable. (An overall F-test is discussed later.) First, the linear equations implied by the structural model are written out, the relevant interactions are added, and F is computed for each equation as above.

If F is significant, it means there is likely to be one or more non-spurious interactions in the population model (represented by the present sample). If F is not significant, it suggests it is unlikely there are any population interactions in the population model.

ESTIMATION Next, the interactions are estimated. However, adding all possible interactions to a model can produce few or no significant interactions. Experience suggests this is common in real-world data because interactions are typically highly correlated. Thus, a "search technique" is required.

In general, depending on the search technique, different search results can obtain. However, Lubinski and Humphreys' (1990) suggestion that an interaction, XZ for example, should be estimated with its relevant quadratics, XX and ZZ, suggests a post hoc search technique for interactions: Gauge each interaction with its relevant quadratics (Step 1); then estimate a final model containing only the significant interaction(s) from each of these tests (Step 2). This avoids mistaking an interaction for its related quadratic (see Lubinski and Humphreys 1990), and the number of interactions to be jointly tested in Step 2 is reduced, which should materially reduce their masking each other.

DISCUSSION The next step would be to develop theoretical justifications for surviving interactions to further reduce the likelihood of their being an artifact of the sample. Stated differently, if a post-hoc interaction cannot be theoretically justified, it should not be interpreted or used as an explanation for a non-significant association because this difficulty with theoretical justification may suggest that it is implausible. In fact, if an interaction cannot be theoretically justified, it probably should not be included in the Step 2 estimation.

It is possible that in Step 2, one or more interactions will be non significant. In that case, several approaches could be taken to further investigate the set of post hoc interactions. However, forward selection using LISREL's Modification Indices may be most appropriate because it is the most conservative. Specifically, the surviving Step 1 interactions are specified with their paths fixed at zero. Then, the interaction with the largest Modification Index (MI) is freed, the structural model is re-estimated, and if the interaction is significant, the next largest MI is freed, etc. until no more freed interactions are significant.

The suggested F-test can become an overall test with multiple structural equations using a Bonferroni approach (see Neter, Kunter, Nachtsheim and Wasserman 1996; however, also see Perenger 1998). Specifically, the confidence of multiple F-tests is greater than 1 minus the sum of the p-values of each test. Thus, if the confidence of the significant F-tests is at least 95%, for example, the overall confidence could be argued to be at least 95%. The amount of specification work could be reduced using the specification templates at http://home.att.net/~rpingjr/research1.htm.

A TRAP For emphasis, in theoretical model testing, the only situation where one should hunt for significant interactions after the hypothesized model has been first estimated is when one wishes to explain a non-significant first-order association (e.g., a or b in Equation 1), or to verify that a significant first-order association is not conditional. Otherwise, hunting for unhypothesized interactions can turn into "data snooping." Data snooping can tempt one to hypothesize significant interactions that are found, as though they were hypothesized before the model was first estimated. This is considered "unsound science" in hypothesis testing. It reverses the "hypothesis-before-first-test" logic of empiricism; producing a "test-before-hypothesis" study, which capitalizes on chance. In a follow-up study or replication, significant interactions could be hypothesized, then tested to investigate whether or not they were significant by chance in the previous study.

REPLICATION A replication or follow up study to test any significant interactions identified using post-hoc probing could be conducted using scenario analysis. Scenario analysis has been used elsewhere in the Social Sciences, and it could be used to provide a follow-up study or replication that could be reported along with the study in which interactions were discovered post hoc. A Scenario Analysis is a comparatively quick and inexpensive experiment in which subjects read written scenarios that portray a situation in which the study constructs are manipulated. Then they are asked to complete a questionnaire containing the study measures (see Ping 2004).

A Scenario Analysis using student subjects might provide a comparatively easily executed second study of the significant interactions using the existing questionnaire. Specifically, theoretical justifications and hypotheses could be added for significant post hoc interactions, and the 2nd study would investigate the resulting model with the significant interactions specified. The result could become a paper with two studies. ("Multiple study" papers are common in social science disciplines such as Social Psychology and Consumer Behavior, and it might be instructive to examine a few of them to determine how best to present two-study results (see recent issues of *The Journal of Consumer Research*, for example).).

The results of scenario analysis when compared with other research designs such as cross sectional surveys (see for example Rusbult, Farrell, Rogers and Mainous, 1988), have been reported to be similar enough to suggest that scenario analysis may be useful in "validating" post hoc interactions. Additional benefits of adding such a study to the original study include that it would provide a second study along with the original study (as routinely done in experimental studies in several branches of the Social Sciences).

SUMMARY AND CONCLUSION

This research critically addressed several matters related to interactions in theoretical models using survey data that have received comparatively little attention. For example, it discussed how to envision interactions (foresee them a priori at the model building stage) and suggested ways to justify and hypothesize them. It discussed the costs of estimating a-priori interactions, issues with the proposed interaction estimation proposals, and that current estimation proposals provide an insufficient interaction disconfirmation test. Along the way it discussed interpreting a latent variable interaction, conceptual difficulties with adding an interaction to a covariant structure model, post-hoc probing for interactions, and the temptation to hypothesize significant post-hoc interactions as though they were envisioned before the data was collected.

REFERENCES

Aiken, Leona S. and Stephen G. West (1991), *Multiple Regression: Testing and Interpreting Interactions*, Newbury Park, CA: SAGE Publications.

Ajzen, Icek and Martin Fishbein (1980), *Understanding Attitudes and Predicting Social Behavior*, Englewood Cliffs, NJ: Prentice Hall.

Algina, James and Bradley C. Moulder (2001), "A Note on Estimating the Jöreskog-Yang Model for Latent Variable Interaction Using LISREL 8.3," Structural Equation Modeling, 8 (1) 40-52.

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.

Bollen, Kenneth A. (1995), "Structural Equation Models that are Nonlinear in Latent Variables: A Least Squares Estimator," *Sociological Methodology*, 25, 223-251.

Bohrnstedt, G.W. and G. Marwell (1978), "The Reliability of Products of Two Random Variables," in Sociological Methodology, K. F. Schuessler ed., San Francisco: Jossy Bass, 254-273.

Busemeyer, Jerome R. and Lawrence E. Jones (1983), "Analysis of Multiplicative Combination Rules When the Causal Variables are Measured With Error," *Psychological Bulletin*, 93 (May), 549-62.

Campbell, Donald T. (1960), "Recommendations for APA Test Standards Regarding Construct, Trait and Discriminant Validity," *American Psychologist*, 15, 546-553.

Cohen, Jacob and Patricia Cohen (1983), *Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences*, Hillsdale, NJ: Lawrence Erlbaum.

Cortina, Jose M. Gilad Chen and William P. Dunlap (2001). Testing interaction effects in Lisrel: examination and illustration of available procedures. *Organizational Research Methods*, 4 (4), 324-360.

Darlington, R. B. (1990), *Regression and Linear Models*, New York: McGraw-Hill.

Denters, Bas and Rob A. G. Van Puijenbroek (1989), "Conditional Regression Analysis," *Quality and Quantity*, 23 (February), 83-108.

DeVellis, Robert F. (1991), *Scale Development: Theory and Applications*, Newbury Park, CA: SAGE Publications.

Dwyer, F. Robert, Paul H. Schurr, and Sejo Oh (1987), "Developing Buyer-Seller Relationships," *Journal of Marketing*, 51 (April), 11-27.

Fornell, Claes and David F. Larker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.

Friedrich, R. J. (1982), "In Defense of Multiplicative Terms in Multiple Regression Equations," *American Journal of Political Science*, 26, 797-833.

Hayduk, Leslie A. and Tom Wonnacut (1980), " 'Effect Equations' or Effect Coefficients:' A Note on the Visual and Verbal Presentation of Multiple Regression Interactions," *Canadian Journal of Sociology*, 5, 399-404.

Hayduk, Leslie A. (1987), *Structural Equation Modeling with LISREL: Essentials and Advances*, Baltimore, MD: Johns Hopkins Press.

Heeler, Roger M. and Michael L. Ray (1972), "Measure Validation in Marketing," *Journal of Marketing Research*, 9 (November), 361-70.

Hunt, Shelby D. (1983). *Marketing Theory: The Philosophy of Marketing Science.* Homewood, IL: Irwin.

Jaccard, James, Robert Turrisi and Choi K. Wan (1990), *Interaction Effects in Multiple Regression*, Newbury Park, CA: SAGE Publications.

Jaccard, James and C. K. Wan (1995), "Measurement Error in the Analysis of Interaction Effects Between Continuous Predictors Using Multiple Regression: Multiple Indicator and Structural Equation Approaches," *Psychological Bulletin*, 117 (2), 348-357.

Jöreskog, Karl G. and Fan Yang (1996), "Nonlinear Structural Equation Models: The Kenny and Judd Model with Interaction Effects," *Advances in Structural Equation Modeling Techniques*, G.A. Marcoulides, R.E. Schumacker, eds., Hillsdale, NJ: LEA.

Jöreskog, Karl G. (2000), "Latent Variable Scores and Their Uses," (on-line paper), http://www.ssicentral.com/lisrel/ techdocs/lvscores.pdf.

Kahneman, Daniel, and Amos Tversky (1979) "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, XLVII (1979), 263-291Kendall, M.G. and A. Stuart (1958), *The Advanced Theory of Statistics*, Vol. 1, London: Charles Griffith.

Kenny, D. and C. M. Judd (1984), "Estimating the Nonlinear and Interactive Effects of Latent Variables," *Psychological Bulletin*, 96, 201-10.

Klein, A.G. and H. Moosbrugger (2000), "Maximum Likelihood Estimation of Latent Interaction Effects with the LMS Method," *Psychometrika*, 65, 457-474.

Klein, A.G. and B.O. Muthén (2002), "Quasi Maximum Likelihood Estimation of Structural Equation Models with Multiple Interactions and Quadratic Effects," Unpublished ms., Graduate School of Education, UCLA.

Lubinski, D. and Humphreys, L.G. (1990), "Assessing Spurious Moderator Effects: Illustrated Substantively with the Hypothesized ("Synergistic") Relation Between Spatial and Mathematical Ability," *Psychological Bulletin*, 107, 385-393.

Marsh, Herbert W., Zhonglin Wen and Kit-Tai Hau (2004), "Structural Equation Models of Latent Interactions: Evaluation of Alternative Estimation Strategies and Indicator Construction," *Psychological Methods*, 9 (3), 275-300.

Mathieu, J.E., S.I. Tannenbaum and E. Salas (1992), "Influences of Individual and Situational Characteristics on Measuring of Training Effectiveness," *Academy of Management Journal*, 35, 828-847.

Molder, Bradley C. and James Algina (2002), "Comparison of Methods for Estimating and Testing Latent Variable Interactions," *Structural Equation Modeling*, 9 (1), 1-19.

Neter, John, Michael H. Kunter, Christopher J. Nachtsheim and William Wasserman (1996). *Applied linear statistical models*. Homewood, IL: Irwin.

Nunnally, Jum C. (1978), *Psychometric Theory*, 2nd Ed., New York: McGraw-Hill.

Peirce, C.S., (1931–1935, 1958), *Collected Papers of Charles Sanders Peirce*, vols. 1–6, Charles Hartshorne and Paul Weiss (eds.), vols. 7–8, Arthur W. Burks (ed.), Harvard University Press, Cambridge, MA.

Perenger, T. V. (1998), What's wrong with Bonferroni adjustments? *British Medical Journal*, 316 18 Apr), 1236-1238.Peter, J. Paul (1981), "Construct Validity: A Review of Basic Issues and Marketing Practices," *Journal of Marketing Research*, 18 (May), 133-45.

Peter, J. Paul (1981), "Construct Validity: A Review of Basic Issues and Marketing Practices," *Journal of Marketing Research*, 18 (May), 133-45.

Ping, R. (1993), "The Effects of Satisfaction and Structural Constraints on Retailer Exiting, Voice, Loyalty, Opportunism, and Neglect," *Journal of Retailing*, 69 (Fall), 320-352.

Ping, R. (1994). Does satisfaction moderate the association between alternative attractiveness and exit intention in a marketing channel?" *Journal of the Academy of Marketing Science*, 22 (1 Fall), 364-371.

Ping, R. (1995), "A Parsimonious Estimating Technique for Interaction and Quadratic Latent Variables," *The Journal of Marketing Research*, 32 (August), 336-347.

Ping, R. (1996a), "Latent Variable Interaction and Quadratic Effect Estimation: A Two-Step Technique Using Structural Equation Analysis," *Psychological Bulletin*, 119 (January), 166-175.

Ping, R. (1996b), "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients," *Multivariate Behavioral Research*, 31 (1), 95-120.

Ping, R. (1996c), "Improving the Detection of Interactions in Selling and Sales Management Research," *Journal of Personal Selling and Sales Management*, 16 (Winter), 53-64.
Ping, R. (1999). Unexplored antecedents of exiting in a marketing channel. *Journal of Retailing*, 75 (2), 218-241.
Ping, R. (2004) "On Assuring Valid Measures for Theoretical Models Using Survey Data," *Journal of Business Research*, 57 (2) (February), 125-141.
Rusbult, Caryl E., Dan Farrell, Glen Rogers and Arch G. Mainous III (1988), "Impact of Exchange Variables on Exit, Voice, Loyalty, and Neglect: An Integrative Model of Responses to Declining Job Satisfaction," *Academy of Management Journal*, 31 (September), 599-627.
Schermelleh-Engle, K., A. Kein and H. Moosbrugger (1998), "Estimating Nonlinear Effects using a Latent Moderated Structural Equations Approach," in *Interaction and Nonlinear Effects in Structural Equation Modeling*, R. E, Schumacker and G. A. Marcoulides, eds., Mahwah, NJ: Erlbaum.
Shaw, Marvin E. and Philip R. Costanzo (1982), *Theories of Social Psychology*, NY: McGraw-Hill
Stolzenberg, Ross M. (1979), "The Measurement and Decomposition of Causal Effects in Nonlinear and Nonadditive Models," in *Sociological Methodology*, Karl F. Schueller ed., San Francisco: Jossey-Bass, 459-488
Thibaut, John W., Harold H. Kelley (1959), *The Social Psychology of Groups*, New York: Wiley.
Wall, M.M. and Y. Amemiya (2001), "Generalized Appended Product Indicator Procedure for Nonlinear Structural Equation Analysis," *Journal of Educational and Behavioral Statistics*, 26, 1-29.

Table A- Equation 2 Structural Model Estimation Results

Y =	$b_X X$	+ $b_Z Z$	+ $b_{XZ} XZ$	+ $b_{XX} XX$	+ $b_{ZZ} ZZ$	
	-.849	.047	-.297	.001	.004	(= unstd. b)
	(-5.32)	(0.59)	(-4.00)	(0.10)	(0.09)	(= t-value)

Table B- Unstandardized Y Associations with Z and X Implied by the Table A Results

Z-Y Association Moderated by X[a]				X-Y Association Moderated by Z[e]			
X Level[b]	Z Coef-ficient[c]	SE of Z Coef-ficient[d]	t-value of Z Coef-ficient	Z Level[f]	X Coef-ficient[g]	SE of X Coef-ficient[h]	t-value of X Coef-ficient
5	-0.23	0.09	-2.48	5	-1.31	0.25	-5.19
4.05[i]	0.04	0.08	0.59	4	-1.01	0.19	-5.33
4	0.06	0.08	0.77	3.44[i]	-.84	0.16	-5.32
3	0.36	0.12	2.92	3	-.71	0.14	-5.15
2	0.65	0.18	3.52	2	-.42	0.11	-3.59
1.2	0.89	0.24	3.70	1	-.12	0.13	-.90
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8) (Col. Number)

[a] The Table displays the variable association of X and Z with Y produced by the significant XZ interaction. In Columns 1-4 when the level of X was low in Column 1, small changes in Z were positively associated with Y (see Column 2). At higher levels of X however, Z was less strongly associated with Y, until near the study average for X, the association was non significant (see Column 4). When X was above its study average, Z was negatively associated with Y.

[b] X is determined by the observed variable (indicator) with the loading of 1 on X (i.e., the indicator that provides the metric for X). This indicator, and therefore X ranged from 1.2 (= low X) to 5 in the study.

[c] The coefficient of Z was (.047-.297X)Z with X mean centered. E.g., when X = 1.2 the coefficient of Z was .047-.297*(1.2 - 4.05) = .89.

[d] The Standard Error of the Z coefficient is:

$$\sqrt{Var(b_Z+b_{XZ}X)} = \sqrt{Var(b_Z) + X^2Var(b_{XZ}) + 2XCov(b_Z,b_{XZ})},$$

where Var and Cov denote variance and covariance, and b denotes unstandardized structural coefficients from Table A.

[e] This portion of the Table displays the association of X and Y moderated by Z. When Z was low in Column 5, the X association with Y was not significant (see Column 8). However, as Z increased, X's association with Y quickly strengthened, until it was negatively associated with Y for most values of Z in the study.

[f] Z is determined by the observed variable (indicator) with the loading of 1 on Z (i.e., the indicator that provides the metric for Z). This indicator, and therefore Z ranged from 1 (= low Z) to 5 in the study.

[g] The unstandardized coefficient of X is (-.849-.297Z)X with Z mean centered. E.g., when Z = 1 the coefficient of X is -.849-.297*(1-3.44) = -.12.

[h] The Standard Error of the X coefficient is:

$$\sqrt{Var(b_X+b_{XZ}Z)} = \sqrt{Var(b_X) + Z^2Var(b_{XZ}) + 2ZCov(b_X,b_{XZ})} ,$$

where Var and Cov denote variance and covariance, and b denotes unstandardized structural coefficient from Table A.

[i] Mean value in the study.

ON THE MAXIMUM OF ABOUT SIX INDICATORS PER LATENT VARIABLE WITH REAL-WORLD DATA

Robert Ping
Associate Professor of Marketing
College of Business Administration
Wright State University
Dayton, OH 45435
(937) 775-3047   (FAX) -3545
rping@wright.edu

ON THE MAXIMUM OF ABOUT SIX INDICATORS
PER LATENT VARIABLE WITH REAL-WORLD DATA

ABSTRACT

Authors have noted that consistent latent variables have a maximum of about six indicators each. This paper discusses this perhaps surprising behavior and its implications, and an explanation is offered. Approaches to utilizing more than about six indicators in latent variables are also discussed, and several novel approaches are proposed. Each of these approaches is explored using real-world data.

Theoretical model tests (hypothesis testing) involving structural equation analysis combine unobserved or latent variables with proposed linkages among these variables (a model) and proposed (observed) measures of these unobserved variables. These model tests usually involve several steps including defining the model constructs, stating the relationships among these constructs, developing appropriate measures of the constructs, gathering data using these measures, validating these measures, and validating the proposed model.

Commenting on step three, developing appropriate measures, authors have noted that latent variables seem to have an upper limit of about six indicators each (Anderson & Gerbing, 1984; Gerbing & Anderson, 1993; Bagozzi & Heatherton, 1994; Ping, 2004). This apparent "maximum" for latent variable itemization has produced an unfortunate result. Cattell (1973) commented that measures used with structural equation analysis tend to be "bloated specific" (operationally narrow) instances of their target construct. Larger well-established measures developed before structural equation analysis became popular have virtually disappeared from published theoretical model tests involving latent variables. When they do appear in published studies involving structural equation analysis, they frequently are "shadows of their former selves" because of extensive item weeding (i.e., the deletion of items from a measure to attain model-to-data fit).

This paper explores the apparent ceiling of about six indicators per latent variable. An explanation for this result in real-world data is proposed, and approaches to avoiding this apparent limit in theoretical model testing are explored using real-world data.

The observed upper limit of about six indicators per latent variable in published model tests is apparently the result of persistent model-to-data fit difficulties (i.e., inconsistency,[1] see Anderson & Gerbing, 1982) with itemizations containing more than about six indicators per latent variable in real-world data. Gerbing and Anderson (1993) commented that "...fit indices indicated less fit as the...number of indicators per factor, increased..." They went on to propose that "Models with few indicators per factor...have fewer df (degrees of freedom), leaving more 'room to maneuver' the parameter estimates so as to minimize the fit function, which in turn is a function of the residuals."

An Additional Explanation

Intuitively, lack of model-to-data fit in a set of items is the result of unrelated items in that set of items--items that do not "cluster" well enough with the other measure items. Mechanically, the input correlation between an unrelated item in a measure and each of the other measure items cannot be satisfactorily accounted for by the model paths connecting them.[2] Gerbing and Anderson's (1993) comment above suggests that "unrelatedness" increases simply by specifying additional indicators.

The Footnote 2 equation for the model-implied covariance of two unidimensional items suggests an alternative explanation for increased "unrelatedness" when an additional indicator is added to a latent variable. Specifying indicators without accounting for correlations among measurement errors in real-world data (e.g., because of common method) may eventually ruin model-to-data fit. In different words, by ignoring the potential for correlated measurement errors in real-world data, and thus not specifying them, the sum of the residuals (i.e., the sum of the differences between the Footnote 2 computed covariances of the items without the correlated error terms, and the input covariances) eventually becomes unacceptably large.

Next, we will discuss several remedies for lack of model-to-data fit, which will subsequently be investigated using real-world data.

Classical Remedies for Lack of Fit

Classical remedies for lack of model-to-data fit include removing items (item weeding), and correlating indicator measurement errors. The pros and cons of each of these remedies are discussed next.

Item Weeding In published theoretical model tests involving structural equation analysis and real-world data, the about-six-indicators limit frequently produces "item weeding," the removal of items from a measure, to attain a set of indicators that fits the data. This approach has the benefit of producing a subset of items that "clusters" together (i.e., their single construct measurement model is consistent; it fits the data).

However, because the items to be deleted are usually unknown beforehand, item weeding usually capitalizes on chance. In addition, the process of weeding is tedious. As we will see, there may also be several subsets of items for a latent variable that will fit the data (i.e., item weeding may be indeterminate). Finally, structural coefficients, standard errors, and thus observed significances and their interpretation, can vary across these weeded itemizations (i.e., the interpretation of structural coefficients in a model involving weeded subsets can be equivocal).

Item weeding to attain model fit in valid and reliable measures has also been criticized because it impairs content or "face" validity[3] (e.g., Cattell, 1973, 1978; see Gerbing, Hamilton & Freeman, 1994). As mentioned earlier, Cattell (1973) remarked that the resulting weeded measures tend to be bloated specific (operationally narrow) instances of their target construct.

Correlated Measurement Errors It is well known that correlating measurement errors can improve model-to-data fit. This result becomes apparent by examining the Footnote 2 Equation. Including a non-zero correlated measurement error term can improve the model-implied (computed) covariance estimate, and thus it can reduce the corresponding residual. The use of correlated measurement errors presumably to improve fit has been reported (e.g., Bagozzi, 1981a; Byrne & Shavelson, 1986; Bearden & Mason, 1980; Duncan, Haller & Portes, 1971; Reilly, 1982), although this approach has become increasingly rare in recent published model tests. It has the benefit of producing a (sub)set of items that appears to "cluster" together (i.e., their single construct measurement model is consistent; it fits the data). However, as we will see, the indiscriminant use of correlated measurement errors can result in a set of items that appears to be consistent but is actually multidimensional (see Gerbing & Anderson, 1984).

Authors have criticized the use of correlated measurement errors to improve fit (e.g., Bagozzi, 1983; Fornell, 1983; Gerbing & Anderson, 1984) for several reasons. These include that it is a departure from the assumptions underlying classical test theory and factor analysis, and the correlated measurement errors that are specified are typically unhypothesized and thus discovered by capitalizing on chance. In addition, the process of identifying measurement errors that should be correlated is tedious, and, as we will see later, there may be several sets of correlated measurement errors that will produce model-to-data fit (i.e., the results of correlating measurement errors may be indeterminate).

Recent Remedies for Lack of Fit

Comparatively recent remedies for lack of model-to-data fit include using second-order constructs, and aggregating items. These remedies are discussed next.

Second-Order Constructs Gerbing and Anderson (1984) argued that a second-order construct is an alternative to using correlated measurement errors.[4] They suggested that a pair of items with correlated measurement errors could be re-specified as a factor (i.e., as a latent variable, and without using correlated measurement errors), and that a second factor containing the rest of the items, along with the first factor, could be specified as the "indicator" latent variables of a second-order construct. This approach has the benefit of producing a set of items that in their second-order specification fits the data.

However, because the items that should be specified in the first factor are unknown beforehand, the process of identifying these first-factor items could be viewed as capitalizing on chance. In addition, the process of identifying the first-factor items is tedious, and there may be several second-order constructs that will fit the data (i.e., the results of this approach may be indeterminate).

Aggregation Kenny (1979) is apparently credited with an approach that involves summing items in a measure to provide a single indicator of a latent variable. The approach uses reliabilities for loadings and measurement error variances, and variations of this approach have been used in the Social Sciences with structural equation analysis presumably to avoid item weeding. (e.g., Heise & Smith-Lovin, 1981; James, Mulaik & Brett, 1982; Williams & Hazer, 1986).[5]

This full or total aggregation (Bagozzi & Heatherton, 1994) alternative to item weeding has several merits including that it allows the use of older well-established measures having more than six items with structural equation analysis (e.g., Williams & Hazer, 1986).

An assumption in structural equation analysis is that the indicators are continuous. When it is averaged, a summed indicator produces a more-nearly-continuous indicator (e.g., averaged ordinal-scaled indicators then take on ratio-valued numbers) that better approximates this continuous data assumption, and thus an aggregated indicator can reduce the bias that attends the criticized use of structural equation analysis with ordinal (e.g., rating scale) data (e.g., Bollen, 1989; Jöreskog & Sörbom, 1996).

A summed indicator also reduces the size of the input covariance matrix (i.e., the input covariances of a summed indicator replace the input covariances of the several indicators comprising the sum), thus reducing the asymptotic incorrectness of the input covariance matrix for a given sample size. In different words, this helps enable the use of the methodological small samples typical in survey-model tests in the Social Sciences (e.g., 200-300) with larger structural models by improving the ratio of the sample size to the size of the covariance matrix.[6] The use of summed indicators also separates measurement issues from model structure issues in structural equation models. In different words, for an unsaturated structural model, lack of fit with a summed-indicators model unambiguously suggests structural model misspecification, rather than suggesting a combination of measurement model difficulties and structural model misspecification.

However, the indiscriminant use of summed indicators could produce a summed item that is composed of multidimensional items. Summed indicators are also non-traditional in structural equation analysis, and their use could be viewed as not particularly elegant when compared to multiple indicator specification. Further, it is believed that a reliability loading can underestimate the loading of a summed item.

Other Remedies

There are several other remedies for lack of model-to-data fit, including partial aggregation, gauging external consistency only, and using measure validation studies. These remedies are discussed next.

Partial Aggregation Bagozzi and Heatherton (1994) also used partial aggregation--items were grouped into subsets and each subset was summed. This approach avoids the use of reliability loadings used in full aggregation if three or more consistent subsets of items can be found. This approach also has all the benefits and drawbacks of full aggregation. However, because the items that should be aggregated are unknown beforehand, partial aggregation could be viewed as capitalizing on chance. The process of finding consistent subset of items is also tedious, and there may be several aggregations of items that will fit the data (i.e., the results of partial aggregation may be indeterminate).

External Consistency Only An additional alternative to item weeding would be to weed (unidimensional) measures jointly, instead of weeding them singly. Item weeding is typically performed one measure at a time (i.e., using single construct measurement models--see Jöreskog, 1993) to establish the internal consistency of each measure (i.e., each measure fits its single construct measurement model--see Anderson and Gerbing, 1988). Later, the resulting internally consistent (unidimensional) measures are jointly specified in a full measurement model (i.e., a measurement model that contains all the measures) to assess the external consistency of the (unidimensional) measures (i.e., the measures jointly fit a unidimensionally specified full measurement model--again see Anderson and Gerbing 1988). However, it could be argued that the ultimate objective of item weeding is for a full (unidimensionally specified) measurement model to fit the data (to isolate any structural model fit problems to the structural paths among the latent variables). Thus, it should be possible to accomplish full measurement model fit by using measures that are unidimensional in the exploratory common factor sense, omitting the internal consistency evaluation step, and item-weeding using a full measurement model only.

Although this remedy has not been used as far as we know, it should have the benefit of producing measures with fewer items weeded out. However, because the items that should be weeded are typically unknown beforehand, this alternative could be viewed as capitalizing on chance. In addition, the process of weeding is tedious, and there may be several sets of the resulting items that will fit the data (i.e., the results of this weeding may be indeterminate). Further, skipping the internal consistency step violates the current received view in theoretical model testing using survey data: Anderson and Gerbing's (1988) "Two-Step" approach to model respecification in order to attain model-to-data fit (i.e., first verify internal consistency, then verify external consistency).

Measure Validation An approach that might reduce some of the drawback of the above approaches would be to conduct a measure validation study. Ideally, measure validation uses several data sets, one to show measure adequacy (i.e., acceptable psychometrics), and one more to validate (i.e., disprove) the adequacy of the measure.

A measure validation approach might allow the "discovery" of "the" (content valid) weeded subset of items for each measure, "the" (acceptable) second-order construct structure, "the" partial aggregation structure, "the" correlated measurement error structure, or "the" external-consistency-only structure of a measure in study one, and the (dis)confirmation of that structure could be attempted in study two. Thus, this approach might permit the use of weeded subsets, second-order constructs, etc. with less criticism because capitalizing on chance would be removed in the second study.

EXAMPLES

To investigate their efficacy, the above approaches were used with a real-world data set. A mailed-out survey used in a theoretical model test produced more than 200 usable responses. Among the variables in the hypothesized model was the construct N that was measured using a new 18-item measure.[7] While the measure for N was judged to be content or face valid, it was multidimensional (i.e., it had three dimensions using maximum likelihood exploratory common factor analysis). The items in the first factor were subsequently judged to be valid and reliable (the coefficient alpha for Factor 1 was .963), but the single construct measurement model for the Factor 1 items was inconsistent (i.e., it was judged to not fit the data using a single construct measurement model--chi square/df/p-value/RMSEA/GFI/AGFI = 270/35/0.0/.227/.670/.481).[8]

Appendix A provides an example of *item weeding* to produce a subset of consistent items for N. In summary, a total of 20 consistent but different weeded subsets of the items of N were found using a procedure suggested by Ping (1998) (see Ping, 2004). The search for additional weeded subsets was discontinued after it became difficult to determine which weeded subset had the "best" content validity.

Because the weeded items were unknown beforehand, the resulting consistent subsets of items all capitalized on chance. In addition, the process of weeding was very tedious, and because the weeding produced multiple itemizations, the resulting subsets of weeded items were indeterminate.[9] Finally, in a simple structural model of the antecedents of N, one of the structural coefficients, its standard error, and thus its significance, became non significant as alternative weeded itemizations of N were specified. Thus, the interpretation of the structural model involving weeded itemizations of N was equivocal.

Appendix B provides an example of the use of *correlated measurement errors* to produce a set of items for N that fits the data. In summary, two sets of correlated measurement errors were found that resulted in all of the Factor 1 items fitting a single construct measurement model for N. An efficient procedure for finding these correlated measurement errors was discovered, and this procedure was used to find a third set of correlated measurement errors that permitted the full 18 item set to fit a single construct measurement model for N.

Because the measurement errors that were correlated were unknown beforehand, these correlated measurement errors capitalized on chance. Further, the process of identifying measurement errors that should be correlated was tedious. There were also several sets of correlated measurement errors, and thus the resulting sets of correlating measurement errors were indeterminate.

Appendix C probed the use of *second-order constructs* to enable model-to-data fit using the Factor 1 items from the measure for N. In summary, no second order specification of Factor 1 could be found that fit the data without resorting to correlated measurement errors. This suggests that with real-world data a second-order specification for inconsistent items may not always be readily apparent. Specifically, these results suggest that in real-world data logically grouping inconsistent items (e.g., Hunter and Gerbing, 1982; Gerbing, Hamilton and Freeman, 1994) and combining weeded and weeded-out items (e.g., Gerbing & Anderson, 1984) in second-order constructs may not always result in a second-order construct that fits the data.

Appendix D provides examples of the use of *full aggregation*. Using factor scores the full 18-item measure for N was aggregated and used to estimate a structural model containing N.

Aggregation was also accomplished with a single averaged indicator for N composed of its Factor 1 items, then a single averaged indicator for N composed of weeded Factor 1 items. In summary, three interpretationally equivalent (i.e., the directions and significances of their structural coefficients were equivalent) full aggregation approaches were reported in addition to the use of factor scores; averaged indicators with averaged LISREL 8 loadings and measurement errors, with averaged maximum likelihood EFA loadings and measurement errors, and with reliability loadings and measurement errors.

Appendix E presents the results of investigating the other suggestions: Partial Aggregation, gauging External Consistency Only to achieve measurement model fit, and the use of Measure Validation.

Using *Partial Aggregation*, several partitionings of the items were investigated. These included logical groupings of all 18 items in the measure for N (i.e., groupings of items that appeared to tap the same facet of N as Bagozzi and Heatherton, 1994 suggested), creating two summed indicators for N from its Factor 1 items (i.e., an indicator that was the average of the weeded Factor 1 items, and another indicator that was the average of the Factor 1 items that were weeded out), creating 3 summed indicators for N from its Factors 1, 2 and 3 items as three averaged indicators for N, subsets of the 18 items of N using maximum likelihood exploratory common factor analysis (EFA) with forced 7, 6, etc. factor solutions, and subsets of the Factor 1 items of N using EFA with forced 7, 6, etc. factor solutions. However, none of these partial aggregations of the items of N fit the data.

This suggests that the specification of a measure with a large number of items may not always be readily apparent using partial aggregation, even when the items cluster together unidimensionally in an exploratory factor analysis (i.e., an acceptable partial aggregation of the Factor 1 items of N could not be found).

Investigating the omission of the internal consistency verification step and achieving model-to-data fit using *External Consistency Only*, we itemized the 9 latent variables in the study with their Factor 1 items. Then we estimated a full measurement model containing all the model latent variables with each set of items specified unidimensionally (each item was specified with only one underlying latent variable). This full measurement model was judged to fit the data without deleting any additional items to attain model-to-data fit.

This suggests that in real-world data omitting the internal consistency verification step for unidimensional items in the maximum likelihood exploratory factor analysis sense may produce a full unidimensionally specified measurement model that fits the data, thus separating measurement from structure as Anderson and Gerbing (1988) and others have stressed.

In order investigate the use of a *Measure Validation* study to avoid some of the above criticisms of item weeding, correlated measurement errors, etc., we conducted a Scenario Analysis. A Scenario Analysis is an experiment in which subjects read written scenarios that portray a situation or scenario in which the study constructs are manipulated. Then they are asked to complete a questionnaire containing the study measures. Unfortunately the protocol used for the scenario analysis produced missing treatments. As a result, while the resulting scenario analysis was useful for assessing reliability and facets of validity, its results were not appropriate for finding "the" (content valid) weeded subset of items for N, "the" correlated measurement error structure, etc. in order to permit the use of weeded subsets, second-order constructs, etc. with fewer of the criticisms mentioned earlier.

Discussion

These results suggest that several of the proposed alternatives to item weeding may not always be useful in real-world data. Second-order constructs failed to perform in the example.

The example also suggested that partial aggregation of a multidimensional measure may not always be an alternative to item weeding in real-world data. Similarly, the example use of a measure validation study with Scenario Analysis did not perform as expected.

Of the alternatives to item weeding discussed, only full aggregation, external consistency only, and correlated measurement errors performed in the example.

Because correlated measurement errors are comparatively rare in recent published model tests, it seems almost pointless to discuss them further. In addition, the example illustrated how they are found by chance. Because there were several sets of correlated measurement errors, the results of correlating measurement errors can be indeterminate, and an unexplored issue is the effect of changes in correlated measurement errors on structural coefficients.

The example suggested that full aggregation might be used to specify a multidimensional measure as a 2nd order construct. Nevertheless, item weeding will probably continue as the preferred approach to attaining measurement model-to-data fit in survey data model tests even though its use has been criticized, and as the examples suggested, the results of these tests and their interpretation may change materially when items are omitted simply to attain measurement model to data fit.

However, the example also suggested an improved approach to item weeding: for each weeded measure find several item weedings that fit the data, then re-convene the item-judging panel to determine which set of items that best taps the conceptual definition for the measure's construct. A through weeding would include the results of weeding the full measure (e.g., the 18 item measure for N), along with "jacknifed" weedings of the full measure (i.e., remove the first item, then weed the rest; replace the first item then remove the second item and weed the rest; etc.). It would also include weedings from pairwise combinations of any factors (e.g., for the items of N, F1 and F2, F1 and F3, and F2 and F3), along with their "jacknives," and weedings from F1 and its "jacknives." A through presentation of the results of weeding to an item-judging panel would include the full measure and the Factor one items, along with the weeded subsets.

If the full measure or its F1 items are judged to be more content valid than any of the weeded submeasures, External Validity Only and full aggregation could be used for that (sub)measure[10] (i.e., other weeded measures might be combined with Externally Valid Only measures or fully aggregated measures).

Several comments may be of interest. Appendix A illustrated the alternative explanation proposed earlier for the apparent ceiling of about six internally consistent indicators: item weeding reduced the number of unspecified but significant measurement error intercorrelations that contributed to the residuals in a single construct measurement model (specified without correlated measurement errors). Specifically, before weeding there were 25 significant modification indices for the correlations between the measurement errors in the Factor 1 items (not reported), and the sum of these modification indices without regard to sign was 474. As each item was weeded (removed), the number of these significant modification indices declined, and so did their sum without regard to sign. Perhaps surprisingly, the resulting consistent weeded subset, Subset 2, had three significant modification indices for the correlations between the remaining measurement errors.

REFERENCES

Anderson, James C. and David W. Gerbing (1982), "Some Methods for Respecifying Measurement Models to Obtain Unidimensional Construct Measurement," *Journal of Marketing Research*, 19 (November), 453-60.

Anderson, James C. and David W. Gerbing (1984), "The Effect of Sampling Error on Convergence, Improper Solutions, and Goodness of Fit Indices for Maximum Likelihood Confirmatory Factor Analysis," *Psychometrika*, 49, 155-73.

Anderson, James C. and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 103 (May), 411-23.

Bagozzi, Richard P. (1981a), "Attitudes, Intentions, and Behavior: A Test of Some Key Hypotheses," *Journal of Personality and Social Psychology*, 41 (4), 607-27.

Bagozzi, Richard P. (1981b), "An Examination of the Validity of Two Models of Attitude," *Multivariate Behavioral Research*, 16 (July), 323-59.

Bagozzi, Richard P. (1983), "Issues in the Application of Covariant Structure Analysis: A Further Comment," *Journal of Consumer Research*, 9 (March), 449-50.

Bagozzi, Richard P. and Todd F. Heatherton (1994), "A General Approach to Representing Multifaceted Personality Constructs: Application to Self Esteem," *Structural Equation Modeling*, 1 (1), 35-67.

Bearden, William O. and J. Barry Mason (1980), "Determinants of Physician and Pharmacist Support of Generic Drugs," *Journal of Consumer Research*, 7 (September), 121-30

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.

Browne, Michael W. and Robert Cudeck (1993), "Alternative Ways of Assessing Model Fit," in *Testing Structural Equation Models*, K. A. Bollen et al. eds, Newbury Park CA: Sage.

Byrne, B. M. and R. J. Shavelson (1986), "Adolescent Self-Concept: Testing the Assumption of Equivalent Structure Across Gender," *American Educational Research Journal*, 12, 365-85.

Cattell, R. B. (1973), *Personality and Mood by Questionnaire*, San Francisco: Jossey-Bass.

Cattell, R. B. (1978), *The Scientific use of Factor Analysis in Behavioral and Life Sciences*, New Your: Plenum.

Duncan, O. D., A. O. Haller and A. Portes (1971), "Peer Influences on Aspirations: A Reinterpretation," in *Causal Models in the Social Sciences*, H. M. Blalock, Jr. ed., Chicago: Aldane.

Dwyer, F. Robert and Sejo Oh (1987), "Output Sector Munificence Effects on the Internal Political Economy of Marketing Channels," *Journal of Marketing Research*, 24 (November), 347-358.

Fornell, Claes (1983), "Issues in the Application of Covariant Structure Analysis: A Comment," *Journal of Consumer Research*, 9 (March), 443-47.

Fornell, Claes and David F. Larker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.

Gerbing, David W. and James C. Anderson (1984), "On the Meaning of Within-Factor Correlated Measurement Errors," *Journal of Consumer Research*, 11 (June), 572-80.

Gerbing, David W. and James C. Anderson (1993), "Monte Carlo Evaluations of Goodness-of-Fit Indices for Structural Equation Models," in *Testing Structural Equation Models*, K. A. Bollen and J. S. Long, eds., Newbury Park, CA: SAGE Publications.

Gerbing, David W., Janet G. Hamilton and Elizabeth B. Freeman (1994), "A Large-scale Second-order Structural Equation Model of the Influence of Management Participation on Organizational Planning Benefits," *Journal of Management*, 20, 859-85.

Heise, D. R. and L. Smith-Lovin (1981), "Impressions of Goodness, Powerfulness, and Liveliness from Discerned Social Events," *Social Psychology Quarterly*, 44, 93-106.

Hunter, John Edward and David W. Gerbing (1982), "Unidimensional Measurement, Second-Order Factor Analysis and Causal Models," in *Research in Organizational Behavior*, Vol. IV, Barry M. Staw and L. L. Cummings eds., Greenwich CT: JAI Press, 267-320.

James, J. R., S. S. Mulaik, and J. M. Brett (1982), *Causal Analysis*, Beverly Hills: SAGE.

Jöreskog, Karl G. (1970), "A General Method for Analysis of Covariance Structures," *Biometrika*, 57, 239-251.

Jöreskog, Karl G. (1993), "Testing Structural Equation Models," in *Testing Structural Equation Models*, Kenneth A. Bollen and J. Scott Long eds., Newbury Park, CA: Sage.

Jöreskog, Karl G. and Dag Sörbom (1996), *LISREL 8 User's Reference Guide*, Chicago: Scientific Software International, Inc.

Kenny, David (1979), Correlation and Causality, New York: Wiley.

Ping, R. A. (1998), "Some Suggestions for Validating Measures Involving Unobserved Variables and Survey Data," *1998 Winter American Marketing Association Educators' Conference Proceedings*, Chicago: American Marketing Association.

Ping, R. A. (2004), "On Assuring Valid Measures for Theoretical Models Using Survey Data," *Journal of Business Research*, 57 (2), 125-41.

Reilly, Michael D. (1982), "Working Wives and Convenience Consumption," *Journal of Consumer Research*, 8 (March), 407-18.

Rindskopf, David and Tedd Rose (1988), "Some Theory and Applications of Confirmatory Second-order Factor Analysis," *Multivariate Behavioral Research*, 23 (January), 51-67.

Sörbom, D. (1975), "Detection of Correlated Errors in Longitudinal Data," *British Journal of Mathematical and Statistical Psychology*, 28, 138-51.

Steiger, J.H. (1990), "Structural Model Evaluation and Modification: An Interval Estimation Approach," *Multivariate Behavioral Research*, 25, 173-180.

Werts, C. E., R. L. Linn and K. G. Jöreskog (1974), "Intraclass Reliability Estimates: Testing Structural Assumptions," *Educational and Psychological Measurement*, 34, 25-33.

Williams, Larry J. and John T. Hazer (1986), "Antecedents and Consequences of Satisfaction and Commitment in Turnover Models: A Reanalysis Using Latent Variable Structural Equation Methods," *Journal of Applied Psychology*, 71 (May), 219-231.

Wright, Sewell (1934), “The Method of Path Coefficients, *Annals of Mathematical Statistics*, 5, 161-215.

APPENDICES

APPENDIX A--Item Weeding

To investigate item weeding, the full 18-item measure for N was subjected to a procedure for item weeding suggested by Ping (1998) (see Ping, 2004). This procedure uses partial derivatives of the likelihood function with respect to the measurement error terms (Modification Indices for Theta Epsilon in LISREL, LMTEST in EQS). Specifically, a single construct measurement model for the full 18 item (multidimensional) measure of N was specified unidimensionally and with the correlations among the measurement errors fixed at zero. This produced a matrix of modification indices for the fixed correlated measurement errors which was then examined, and the item in that matrix with the largest summed modification index without regard to sign (i.e., the sum of the item's column of modification indices without regard to sign) was deleted. Next, the single construct measurement model without this item was re-estimated, and the item with the largest summed modification index without regard to sign in the resulting modification indices for the correlations among the measurement errors was deleted. This process was repeated, deleting an item at each step, until a subset of the 18 items was found that fit the data.

The resulting 5-item subset (Subset 1--containing the items n_1, n_{12}, n_{14}, n_{15} and n_{18}) was consistent (it fit the data--chi square/df/p-value/RMSEA/GFI/AGFI = 2.84/5/.723/0/.991/.974) (see Footnote 8 for a discussion of model fit), and it contained items from Factor 1 (n_{12}, n_{14}, n_{15} and n_{18}) and an item from Factor 3. There was another consistent 5-item subset (Subset 1a-- n_4, n_{12}, n_{14}, n_{15} and n_{18}--chi square/df/RMSEA/GFI/AGFI = 9.09/5/.105/.079/.973/.919), that contained the Factor 1 items and a Factor 2 item. However, we could not find a consistent subset with items from all three factors. Nevertheless, it was possible to find consistent subsets with items from several factors.

Using this derivative procedure on the 10 items of Factor 1, a different consistent subset obtained (Subset 2--n_9, n_{13}, n_{15}, n_{16}, n_{17} and n_{18}--chi square/df/p-value/RMSEA/GFI/AGFI = 15/9/.086/.072/.961/.910). This 6-item subset was judged to be slightly more content or face valid than Subsets 1 or 1a, and its items clustered together using maximum likelihood exploratory common factor analysis slightly better than Subset 1 (the percent of the variance explained for Subset 2 was 75.7%, versus 59.8% for Subset 1 and 66.8% for Subset 1a).

Several more consistent subsets were then obtained. Obviously any subset of Subset 2 would fit the data, and there were 41 of these (= the total number of combinations of 6 things taken 5, 4 then 3 at a time). In addition, arbitrarily omitting an item from the Factor 1 set of 10 items produced two more consistent subsets, Subset 3 (n_9, n_{10}, n_{11} and n_{17}--chi square/df/p-value/RMSEA/GFI/AGFI = 3.34/2/.187/.072/.987/.939), and Subset 4 (n_{11}, n_{13}, n_{16}, n_{17} and n_{18}--chi square/df/p-value/RMSEA/GFI/AGFI = 2.64/5/.754/0.0/.992/.976). These subsets clustered together about as well as Subset 2 (71.6% and 75.1% explained variance respectively) and judging which had the "best" content validity became impossible without resorting to an item-judging panel. However, we judged each of the weeded measures to be less content valid than either the original 18-item measure, or the 10 item Factor 1 measure.

We then discontinued the search.[11] In summary, we identified 20 consistent subsets of the 18-item measure for N using the derivative procedure.

Finally, we gauged the sensitivity of structural coefficients to changing the itemization of N. In a simple saturated structural model with N as the single endogenous variable, the structural coefficients and standard errors were judged to vary unpredictably across these different itemizations of N. For example, the t-value for one of the 4 significant structural coefficients changed from t = 2.66 to t = 0.91 by changing the itemization of N from weeded Subset 1a to Subset 4.[12]

APPENDIX B--Correlated Measurement Errors

In order to investigate correlated measurement errors, the full 18-item measure for N was subjected to a procedure involving modification indices. A single construct measurement model for the full (multidimensional) measure of N was specified unidimensionally and with the correlations among the measurement errors fixed at zero to produce a matrix of modification indices for the fixed correlated measurement errors. Then, the measurement error correlation corresponding to the largest of these modification indices was freed (i.e., the corresponding measurement errors were allowed to correlate) (a modification index of 3.8 is significant at p = .05 with 1 degree of freedom, see the second part of Footnote 14). Next, the single construct measurement model was estimated with this measurement error correlation freed, and the largest of the resulting modification indices for the remaining fixed correlations among the measurement errors was found and freed. This process was repeated, freeing a measurement error correlation at each step, a total 90 times before we decided to abandon this "forward selection" process of identifying correlated measurement errors.[13]

However, we used the above "forward selection" approach on a smaller subset of items, the Factor 1 items (10 items-- n_9 through n_{18}), until the set of Factor 1 items was judged to fit the data. The Factor 1 items with correlated measurement errors was judged to be consistent (i.e., it fit the data--chi square/df/p-value/RMSEA/GFI/AGFI = 34/20/.022/.074/.949/.861) (see Footnote 8 for comments on model fit). The procedure required 27 estimations and produced 15 significant correlated measurement errors (9:10,11; 10:12,13,16,17; 11:--; 12:13,17; 13:14,15,16;

14:17,18; 15:16,18; 16:--; 17:--; where for example 9:10,11 denotes the correlations between ε_9 and ε_{10}, and ε_9 and ε_{11}, where ε denotes a measurement error term, and 11:-- for example indicates that ε_{11} was not correlated with its higher-ordinality measurement errors, ε_{12} through ε_{18}).

To find another set of correlated measurement errors for the Factor 1 items, we specified a single construct measurement model for the Factor 1 items with all the measurement error correlations fixed at zero, except for the ε_9 correlations which were freed.[14] Estimating this model we recorded the significant measurement error correlations between ε_9 and ε_{10} through ε_{18}. Next we re-fixed the measurement error correlations with ε_9 to zero, and freed the ε_{10} measurement error correlations with its higher-ordinality measurement errors, ε_{11} through ε_{18} (i.e., all measurement error correlations were fixed at zero except for those between ε_{10} and ε_{11}, ε_{10} and ε_{12}, ... , and ε_{10} and ε_{18}). After estimating this model, we recorded the significant measurement error correlations between ε_{10} and ε_{11} through ε_{18}. Repeating this process of re-fixing the previously freed measurement error correlations, and freeing and estimating the higher-ordinality measurement error correlations for ε_{11}, then ε_{12}, ... , then ε_{18} (e.g., the ε_{11} correlations with ε_{12} through ε_{18}, the ε_{12} with ε_{13} through ε_{18}, etc.), the result was a set of significant measurement error correlations for ε_9 through ε_{18}.

Next, we re-specified the single construct measurement model for Factor 1 with the all measurement error correlations again fixed at zero. Then, we freed the significant modification indices just recorded for ε_9, ε_{10}, etc. (i.e., based on their recorded modification indices, the significant correlations for ε_9 were freed, the significant correlations for ε_{10} were freed, etc.). This single construct measurement model was estimated (chi square/df/p-value/RMSEA/GFI/AGFI = 31/14/.004/.098/.956/.828) and the nonsignificant measurement error correlations were trimmed (i.e., fixed at zero--for example 10:11,15 were trimmed because they were nonsignificant when estimated in the presence of the other specified measurement error correlations). This trimmed model was then estimated, and because it did not yet fit the data (chi square/df/p-value/RMSEA/GFI/AGFI = 53/24/.0004/.097/.931/.841) the modification indices (MI's) for the remaining non-freed measurement error correlations were examined to find the largest significant MI, $MI_{9,12}$ (= the modification index for the ε_9-ε_{12} correlation = 10.30). The ε_9-ε_{12} correlation was then freed and the resulting single construct measurement model was estimated. This measurement model was judged to fit the data (chi square/df/p-value/RMSEA/GFI/AGFI = 41/21/.009/.079/.945/.870).

Several comments may be of interest. There were two sets of correlated measurement errors that would permit Factor 1 items to fit the data in a single construct measurement model. Stated differently, there was more than one set of correlated measurement errors that would make the Factor 1 items consistent in a single construct measurement model. While the correlations from the second or "column-wise" selection approach were more parsimonious (i.e., there were fewer of them) and they were found using half as many estimations (12--1 for each item, one more to trim the nonsignificant intercorrelations, plus one to add the additional intercorrelation $MI_{9,12}$--versus 27 for forward selection), their comparative statistics are trivially different (AIC/CAIC/EVCI for the column-wise selection = 105/229/.815 versus 104/240/.803 for forward selection).

Since the second or column-wise selection approach required considerably fewer estimations, we tried it on the full set of 18 items. Obtaining a set of measurement error correlations that were judged to make the full 18-item measure consistent required 19 estimations, one for each item, one to trim the resulting nonsignificant correlations, and 9 more estimations to add enough additional correlations to obtain consistency (chi square/df/RMSEA/GFI/AGFI = 122/69/.00007/.077/.906/.768).

Thus, the original multidimensional set of 18 items could be specified so that it fit the data in a single construct measurement model using correlated measurement errors. Stated differently, correlated measurement errors masked a multidimensional measure.

Finally, the trimming step in the column-wise selection approach (i.e., to remove nonsignificant correlated measurement errors) suggests that some measurement error correlations were collinear. Stated differently, freeing a correlation affected the significance or lack thereof in other correlations. This may explain the apparent indeterminancy in measurement error correlations. Specifically, the starting point (i.e., the measurement errors that were initially allowed to correlate) determined the remaining significant measurement error correlations.

APPENDIX C--Second-Order Constructs

To investigate the use of second-order constructs we tried several approaches. The initial objective was to find a second-order construct for the Factor 1 items that would fit its single construct measurement model. To this end authors have suggested grouping items into subsets using their content or face validity (i.e., grouping items that seem to be related based on their wording--see Hunter and Gerbing, 1982; Gerbing, Hamilton and Freeman, 1994). For example, we grouped the Factor 1 items into two subsets based on wording (i.e., the subset 1 items were used to indicate latent variable 1, the subset 2 items were used to indicate latent variable 2, and latent variables 1 and 2 were specified as the "indicators" of the now second-order Factor 1), then three subsets. However, we were unable to find a grouping of the Factor 1 items based on item wording that fit their single construct measurement model (i.e., the measurement model containing only the second-order construct) without resorting to correlated measurement errors.

Authors have also suggested that weeded-out items might be specified as a second "indicator" factor in a two-factor second-order construct (i.e., a second-order construct with the weeded items as one "indicator" latent variable, and the items that were not weeded out as the other indicator latent variable--see Gerbing and Anderson, 1984). To this end we specified a second-order construct with the weeded-out items as one "indicator" latent variable, and the surviving items as another "indicator" latent variable. Again, we were unable to find a second-order construct that would fit their single construct measurement model without resorting to correlated measurement errors.

We also tried a second-order construct with the two factors that resulted from a forced two-factor solution in maximum likelihood exploratory common factor analysis of the Factor 1 items, then a forced three-factor solution. These second-order constructs also would not fit their single construct measurement model without resorting to correlated measurement errors.

Finally, we tried specifying a second-order construct with three consistent subsets of the Factor 1 items (i.e., two consistent four-item subsets, and one three item subset that fit the data exactly). This second-order construct also did not fit the data without resorting to correlated measurement errors.

Several comments may be of interest. These results suggest that with real-world data the use of a second-order construct may not always easily improve model-to-data fit.

APPENDIX D--Full Aggregation

In full aggregation, a set of items is summed to form a single indicator. Because the resulting latent variable is underdetermined with only one indicator, it requires that two of its three estimated parameters, its loading, its measurement error variance or its latent variable's variance, be fixed for identification.

It is easy to show that the loading of a summed indicator is the sum of its individual indicator loadings, and that its measurement error variance is the sum of the individual indicator measurement error variances.[15]

A reliability loading and the well-known measurement error estimate Variance∗(1-reliability) has been used in the social sciences. It is easy to show that these estimates for an aggregated indicator are exact when the variance of its latent variable is 1 and latent variable reliability is available (see Appendix F).

Ping (2004) suggested using maximum likelihood exploratory factor analysis (EFA) loadings and reliability measurement error variances for a fully aggregated indicator. There is an additional estimate of the measurement error variance available using EFA results (see Equation F6 in Appendix F).

An additional approach would be to replace indicators with their fully aggregated EFA factor scores.

Each of these aggregation alternatives will be explored next.

While it is obviously possible to aggregate a multidimensional measure, none of the above estimates for the resulting single indicator's loading and measurement error variance would be appropriate (because each requires or assumes unidimensionality), with the exception of factor scores. Pursuing that option we produced factor scores using maximum likelihood exploratory factor analysis and the full 18 item measure of N. Specifically, maximum likelihood EFA of N produced three factors, and the factor score for each of these factors was summed then averaged. Next, a full structural model (i.e., with all the latent variables and N as the single endogenous variable) were estimated with N specified using this single aggregated factor score indicator (with a loading of 1 and a measurement error of 0). The resulting structural model was judged to fit the data (chi square/df/RMSEA/GFI/AGFI = 2349/1449/0/.066/.650/.614) (see Footnote 8 for comments about assessing model-to-data fit).

Next, we estimated a series of structural models involving N and the other model latent variables with a full aggregation (average) of the F1 items, or a full aggregation (average) of a weeded subset of the F1 items, Subset 2 from Appendix A (items n_9, n_{13}, n_{15}, n_{16}, n_{17} and n_{18}). The resulting single indicator used the estimates for loadings and measurement error variances mentioned above, averaged LISREL 8 loadings and measurement errors, averaged maximum likelihood EFA loadings and measurement errors, reliability loadings and measurement errors (see Appendix F), and factor scores.

The structural coefficients on the paths to N from the other 8 latent variables were compared to the structural coefficients produced by the equivalent structural model (i.e., containing N and the other 8 latent variables) that used either the 10 items in F1 or Subset 2, with N specified with multiple indicators that were the individual items of F1 or Subset 2.

The Subset 2 full measurement model fit the data (chi square/df/RMSEA/GFI/AGFI = 2789/1733/0/.065/.630/.596), as did its (saturated) structural model (chi square/df/RMSEA/GFI/AGFI = 2789/1733/0/.065/ .630/.596). The structural coefficients that resulted were used as a basis for assessing the efficacy of the various alternative loadings and measurement error variances mentioned above.

In summary, the factor score indicator produced by a maximum likelihood EFA with just the Subset 2 items (i.e., with no other items present in the EFA) was judged to produce the smallest differences between the corresponding t-values of the Sunset 2 (baseline) structural model and the factor score indicator structural model. The root mean square (RMS) (pairwise) difference of t-values across the 8 structural coefficients on the paths to N was .002 and the average difference without regard to sign (MAD) was .005 (with a range of between .000 for the nonsignificant structural coefficients to .012 for the significant structural coefficients) (structural coefficient RMS = .011, MAD = .024, range = [.004, .066]).

The smallest structural coefficient differences were produced by the averaged LISREL 8 loadings and measurement errors--the RMS difference of structural coefficients across the 8 structural coefficients on the paths to N was .007 and the MAD was .015 (the range was .001 to .041) (t-value RMS = .065, MAD = .149, range = [.001, .353]).

The t-value and structural coefficient differences for the other aggregation approaches were nearly identical to those produced by the averaged LISREL 8 loadings and measurement errors. For example, the average of the maximum likelihood EFA loadings and an Equation F6 measurement error variance produced a t-value RMS, MAD and range of .062, .149 and [.001, .353], respectively. Its structural coefficient RMS, MAD and range were .005, .015 and [.000, .024], respectively. The reliability loading and measurement error was similar (t-value RMS = .065, MAD = .149, range = [.001, .353]) (structural coefficient RMS = .006, MAD = .015, range = [.001, .031]).

These results were then used to predict the ranking of the performance of N specified using the fully aggregated F1 items and the above approaches. As a baseline structural model the External Consistency Only (see Appendix E) (full) measurement model for N using the F1 items was re-specified as a structural model. An External Consistency Only full measurement model uses unidimensional sets of indicators that are not necessarily internally consistent (i.e., their single construct confirmatory measurement model may not fit the data). In the present case, each latent variable in the External Consistency Only full measurement model had a unidimensional itemization (in a maximum likelihood EFA sense), but none of these itemizations was consistent (in the confirmatory factor analysis sense). However, the resulting External Consistency Only full measurement model measurement model fit the data (chi square/df/RMSEA/GFI/AGFI = 3480/1979/0/.073/.590/.556) (see Footnote 8 for comments about assessing model fit), as did the corresponding structural model (chi square/df/RMSEA/GFI/AGFI = 3480/1979/0/.073/.590/.556).

As with the weeded Subset 2 of the measure for N, the factor score indicator was judged to have produced the smallest t-value differences (t-value RMS = .032, MAD = .070, range = [.001, .167]). However, it also produced the smallest structural coefficient differences (structural coefficient RMS = .013, MAD = .022, range = [.000, .085]).

The other approaches investigated produced nearly identical results. For example, averaged LISREL 8 loadings and measurement errors produced t-value RMS, MAD and range of .048, .124, [.066, .236], respectively (structural coefficient RMS = .050, MAD = .093, range = [.005, .300]). Averaged maximum likelihood EFA loadings and measurement errors were similar (t-value RMS = .048, MAD = .122, range = [.066, .236]) (structural coefficient RMS = .050, MAD = .092, range = [.004, .300]), as were reliability loadings and measurement errors (see Appendix F) (t-value RMS = .047, MAD = .120, range = [.066, .236]) (structural coefficient RMS = .050, MAD = .093, range = [.000, .300]).

The details of these estimations were as follows. For the factor score approach, a maximum likelihood EFA of the 10 F1 items (i.e., with all other items absent) was performed, and the resulting factor scores were added to the data set (i.e., the resulting factor scores were saved and given the variable name FS). Next, N was specified using the single indicator FS with a fixed loading of 1 and a fixed measurement error variance of 0 (i.e., the variance of N, PHI_N, was free). The resulting structural model was judged to fit the data (chi square/df/RMSEA/GFI/AGFI = 2327/1449/0/.065/.652/.616) (see Footnote 8 for comments about assessing model-to-data fit).

In the LISREL 8 loadings and measurement errors approach, a single construct measurement model of the 10 F1 items (i.e., with all other latent variables absent) was estimated with the variance of N free (i.e., one indicator loading was fixed at 1 to provide a metric for N). Although this single construct measurement model did not fit the data (chi square/df/RMSEA/GFI/AGFI = 273/35/0/.219/.687/.509), the F1 items were unidimensional using maximum likelihood EFA, so the resulting loadings were averaged and the resulting measurement error variances were divided by 10^2 (using expectation algebra, the variance of an average of independent measurement errors, a constant times the sum of the measurement errors, is the constant squared times the sum of the variances of the measurement errors). Next, the F1 indicators were averaged, then N was specified using the resulting single indicator with a fixed loading and measurement error variance equal to the averaged LISREL 8 loadings and the "averaged" measurement error variances just described, respectively (i.e., the variance of N, PHI_N was free). The resulting structural model fit the data with the same fit statistics as the factor score model.

For the maximum likelihood EFA loadings and its Equation F6 measurement error variance estimation, a maximum likelihood EFA of the 10 F1 items (i.e., with all other items absent) was performed. The resulting loadings were re-scaled then averaged,[16] and the measurement error variance of this single was calculated using Equation F2 in Appendix F. Next, the F1 items were averaged, and N was specified using the resulting single indicator with a fixed loading and measurement error variance equal to the averaged EFA loadings and the Equation F6 measurement error variance just described, respectively (i.e., the variance of N, PHI_N again was free). The resulting structural model also fit the data with the same fit statistics as the factor score model.

In the reliability loading and measurement error approach, the coefficient alpha reliability (α) and the error-attenuated variance (V) (i.e., from SPSS, SAS, etc.) of the F1 items was determined. The F1 items were again averaged to form a single indicator of N, and this single indicator's the measurement error variance was fixed at the Equation F2 value of = V*(1- α), the loading of the single averaged indicator of N was fixed at the square root of the coefficient alpha reliability, and the variance of N, PHI_N was freed. Again the resulting structural model fit the data with the same fit statistics as the factor score model.

Several comments may be of interest. Specification of the weeded and "weeded-out" items of F1 was accomplished using several aggregation approaches.

The successful estimation of a structural model for a multidimensional N specified with aggregated factor scores, suggests that aggregated factor scores might be an alternative to a 2nd order factor or partial aggregation (see Appendix E) for specifying a multidimensional measure in structural equation analysis. In the present case an 18 item measure which formed three factors using maximum likelihood EFA was fully aggregated using (averaged) maximum likelihood EFA factor scores.

Other full aggregation indicators, loadings and measurement error terms were investigated (e.g., normed indicators) but not reported because they were judged to have performed worse than the reported approaches.

The results from reliability and LISREL 8 loadings were a surprise. Fixing the loading to the square root of coefficient alpha overstates the loading (see Equation F3 in Appendix F) which should not have performed well. Similarly, the LISREL 8 single construct measurement model for the items of F1 did not fit the data, yet the resulting loadings and measurement error variances were useful in this investigation. However, the External Consistency Only measurement model for N with the F1 items did fit the data, and the F1 loadings and measurement error variances were trivially different from those produced by single construct measurement model for the items of F1. This suggests an additional full aggregation approach which is identical to the LISREL 8 approach (a), except that it uses loadings and measurement error variances from an External Consistency Only measurement model for N.

Comparing the structural coefficients and t-values for the 18-item measure for N with those from F1 and the weeded Subset 2 (not reported), structural coefficients and significances changed materially when items were removed from an aggregated indicator for N (2 structural coefficients that were significant using the 18 item measure became nonsignificant when items were dropped, and 1 structural coefficient became significant when items were removed). In different words, this also suggests that changes in content or face validity of a measure (i.e., the items included or excluded in a measure) may change the construct validity of that measure (i.e., the correlations among the other latent variables in the model).

Finally, to investigate the sensitivity of structural coefficients and their significances to small changes in itemization with full aggregation, we specified N with fully aggregated weeded Subsets 1 through 4 and 1a (see Appendix A). The resulting sensitivity to different weeded subsets of items observed were similar to those reported in Appendix A. This suggests not only that item changes may change the study results and their interpretation, but that full aggregation may not mask these changes.

APPENDIX E--Other Approaches

Partial Aggregation To investigate partial aggregation, we grouped the items of N into subsets of items based on similar face or content validity (i.e., we grouped items that appeared to tap the same facet of N together) as Bagozzi and Heatherton (1994) suggested. We then specified N with the summed items (i.e., N was specified with 2 or 3 summed indicators). However, neither of these groupings fit a single construct measurement model for N.

Next we created two indicators for N from the results of weeding. We summed the weeded Factor 1 items, then did the same for the Factor 1 items that were weeded out. Using a fixed reliability measurement error variance for the summed weeded-out items because the latent variable was underdetermined, the measurement model was judged to not fit the data (chi square/df/RMSEA/GFI/AGFI = 16/8/.047/.082/.979/.871) (the model fit is close to being acceptable, but N was not unidimensional--modification indices suggested that the "weeded out" indicator loaded significantly on several latent variables).

Next, we created three indicators for N using the summed items from Factor 1 for one indicator, the summed items from Factor 2 for another, and the summed items from Factor 3 for the last indicator. However, N was again not unidimensional in the full measurement model containing N and the other latent variables, and it did not fit the data.

Finally, we tried obtaining subsets of the items of N using maximum likelihood exploratory common factor with forced 7, 6, etc. factor solutions, and each factor's items were then summed. While the 7, 6, etc. forced factor varimax and oblimin exploratory factorings converged with the 18 item measure of N, and the 6, 4 and 3 factor exploratory factorings also converged with the Factor 1 items of N, the forced 7 and 5 factors with the Factor 1 items did not converge. In addition, none of the successful forced factorings fit the data (the 3-factor solution was not unidimensional in the full measurement model).

These results suggest that partial aggregation may not always allow the specification of a large number of items (i.e., more than about 6). Specifically, in the present case none of the partial aggregating approaches produced a full measurement model that was external consistent.

External Consistency Only To investigate omitting the internal consistency step for unidimensional measures and achieving full measurement model-to-data fit using external consistency only, we itemized each of the 9 latent variables with their Factor 1 items. Each latent variable thus had a unidimensional itemization (in a maximum likelihood exploratory factor analysis sense), but none of these itemizations was consistent (i.e., none fit their single construct measurement model). However, a full measurement model containing the 9 latent variables specified unidimensionally with their respective Factor 1 items (i.e., each item was "pointed to" by only one latent variable) was judged to fit the data (chi square/df/RMSEA/GFI/AGFI = 3480/1979/0/.073/.590/.556) (see Footnote 8 for comments on assessing model-to-data fit).

To probe the limits of this externally consistent measurement model we weeded each of the multidimensional measures until they became unidimensional in a maximum likelihood exploratory factor analysis (EFA) of the un-weeded items. Specifically, each multidimensional measure was specified in a single factor measurement model as though it were unidimensional. For each measure the partial derivative technique used in Appendix A was used to weed the first item from that measure. The un-weeded items in that measure were factored using maximum likelihood EFA to check their factor structure. If the un-weeded items were multidimensional another item was weeded using the partial derivative technique, and the factor structure of the resulting un-weeded items was again checked using maximum likelihood EFA. This process was repeated until the measure was unidimensional using maximum likelihood EFA.

A full measurement model containing these larger but unidimensional measures was also judged to fit the data, but LISREL produced a warning message that the sample size was smaller than the number of parameters to be estimated, and that the parameter estimates were thus unreliable.

Measure Validation Measure validation (i.e., the determination of the adequacy--reliability and validity--of a measure) in survey models can take several approaches. These include a separate large scale study(s) aimed solely at validating the study measures. However, presumably because of budget and time constraints, large scale measure validation studies are sometimes bypassed, and measure adequacy is gauged in a small scale pretest survey(s) (e.g., 100 cases). These small pretest surveys are used to preliminarily assess the measures, and to determine response rates. Obviously, weeded subsets, second-order constructs, etc. may be difficult if not impossible to investigate with the resulting small data sets.

Again perhaps because of budget and time constraints, the final- (model-) test data set is sometimes used for measure validation. In this case the final-test data are used for two separate purposes: to assess the measures, and to validate or test the hypothesized model that uses these measures. In this case weeded subsets, second-order constructs, etc. could obtain, but all the earlier criticisms (capitalizing on chance, etc.) then apply.

To investigate the use of a separate large-scale measure validation study that takes less time and is less expensive than to a mailed-out survey, we conducted a Scenario Analysis. A Scenario Analysis is an experiment in which the subjects (usually students) read written scenarios that portray a situation in which the study constructs are verbally manipulated (i.e., high for one experimental subject--"you are *very* satisfied with..."--low for another--"you are *very dissatisfied* with..."). Then the subjects are asked to complete the study questionnaire which contains the measures to be validated. Compared with other research designs such as cross sectional surveys, the results of Scenario Analyses have been reported to be similar enough that they might be useful in measure development and validation (Ping, 2004).

A Scenario Analysis designed to assess N and the other study measures was conducted using students. An audit of the resulting completed questionnaires suggested, however, that many scenarios were incomplete or were not administered. Specifically, while there were nine variables in the proposed model, eight were exogenous, and thus the scenario required $2^8 = 256$ completed questionnaires to produce one questionnaire for each treatment (i.e., treatment_1 = high exogenous variable_1, high exogenous variable_2, ... , high exogenous variable_8; treatment_2 = high exogenous variable_1, high exogenous variable_2, ... , low exogenous variable_8; treatment_3 = high exogenous variable_1, high exogenous variable_2, ... , low exogenous variable_7, high exogenous variable_8; treatment_4 = high exogenous variable_1, high exogenous variable_2, ... , low exogenous variable_7, low exogenous variable_8; ... ; treatment_{256} = low exogenous variable_1, low exogenous variable_2, ... , low exogenous variable_7, low exogenous variable_8). However, substantially fewer than 256 usable questionnaires were obtained, and re-administering the missing scenarios was judged to be out of the question because it was nearly impossible to determine which scenarios were missing.[17] The effect of these missing treatments was subsequently judged to be unknown.

However, comparing the results of single construct exploratory common factor analysis of each measure, of the 9 measures, the behavior of 7 measures was same between the scenario analysis and the final test: 5 measures were unidimensional in both studies and 2 measures were multidimensional in the two studies, while 2 measures were unidimensional in scenario analysis and multidimensional in the final test, and no measures were multidimensional unidimensional in scenario analysis and unidimensional in the final test (not reported). However, loadings in the unidimensional measures were different between the data sets (i.e., an item that loaded high in the scenario data loaded lower in the final test data, and vice versa). Further, the multidimensional factor structure was not constant across the data sets--the number of factors were usually different between the data sets, but items in the scenario factor 1's were contained in (i.e., a subset of the) final test factor 1's in all but 1 case.

Further, reliabilities were within a few points of each other between the two data sets. While Average Extracted Variances (AVE's) (see Fornell & Larker, 1981) varied more widely (1 to 19 points), when the scenario AVE's were above .5 so were the final test AVE's.

Encouraged by these sanguine results despite the missing treatments, we weeded the measure for N using the scenario data. As discussed in Appendix A, the first weeded subset of the Factor 1 N items in final test data was Subset 2 (n_9, n_{13}, n_{15}, n_{16}, n_{17} and n_{18}). However, n_{16} was the first item weeded out of Factor 1 in the scenario data (not reported). Thus, for the focal construct N, weeded subsets would not (all) be the same across data sets, and finding "the" weeded subset of items for the Factor 1 items of N using this scenario data set was judged unlikely.

Similarly, the first set of correlated measurement errors found for the Factor 1 items of N in the final test data was (9:10,11; 10:12,13,16,17; 11:--; 12:13,17; 13:14,15,16; 14:17,18; 15:16,18; 16:--; 17:--; where for example 9:10,11 indicates the correlations between ε_9 and ε_{10}, and ε_9 and ε_{11}--ε is a measurement error term, and 11:-- for example indicates that ε_{11} was not correlated with its higher-ordinality measurement errors, ε_{12} through ε_{18}). However, 17:18 was the first correlated measurement error identified in the Factor 1 items of N using the scenario data (ε_{17} and ε_{18} were not correlated in the Factor 1 items in the final test data). Thus, the correlated measurement errors for N would not all be the same across data sets, and finding "the" correlated measurement errors for the Factor 1 items of N using this scenario data set was also judged to be unlikely.

We did not investigate 2nd order constructs or partial aggregation because they did not perform in the final test data. Similarly, because the itemizations of Factor 1 in nearly half of the study latent variables were different between the two data sets, weeding using external consistency only was also not investigated in the scenario data.

Since the reliabilities were similar between the two data sets, however, we did investigate full aggregation. However, the correlations were not same between data sets. For example, N had a correlated with S, an important study variable, of .10 in the scenario data, but this correlation was -.54 in full test data. Thus, the proposed structural model was not investigated in the scenario data because structural coefficients are related to partial correlations (and the scenario data was intended for measure validation rather than model validation).

In summary, while this scenario analysis may have been useful for assessing N and the other study measures' reliability and facets of validity (even with missing treatments), its results were not appropriate for finding "the" (i.e., a content valid) weeded subset of items for N, "the" correlated measurement error structure, etc. in order to permit the use of weeded subsets, second-order constructs, etc. with fewer of the criticisms mentioned earlier because capitalizing on chance would be removed in the second study. However, it remains an open question whether or not a "proper" scenario analysis (i.e., one in which all the treatments were administered) would have produced the same conclusion regarding weeded subsets, etc.

Unfortunately, since one was not performed for this analysis (or the original model test, due to budget and time constraints), it is also an open question whether a large scale measure validation study could have been used to find "the" weeded subset of items for N, "the" correlated measurement error structure, etc.

APPENDIX F--Derivation of Single Indicator Loadings and Measurement Error Variances

Werts, Linn and Jöreskog (1974) proposed that the latent variable reliability (ρ_X) of a of a unidimensional measure X (i.e., the measure has only one underlying latent variable) is given by

F1) $$\rho_X = \frac{L^2_X \, Var(X)}{L^2_X \, Var(X) + E_X} ,$$

where L_X is the sum of the loadings of the items in the measure X on their latent variable *X*, Var(*X*) is the (error disattenuated) variance of *X* (i.e., from a measurement model of *X*), and E_X is the sum of the measurement error variances of the items in the measure X as they load on their latent variable *X*. It is also well known that E_X is given by

F2) $$E_X = Var(X)\,(1 - \rho_X) ,$$

where Var(X) is the (error attenuated) variance of X (e.g., obtained using SAS, SPSS, etc.). By solving Equation (F1) for L_X and substituting Equation (F2) into the result,

F3) $$L_X = [Var(X)\,\rho_X \,/\, Var(X)]^{1/2}$$

which becomes

F4) $$L_X = [\rho_X]^{1/2}$$

when Var(X) equals Var(*X*) (e.g., if *X*, and thus X, is standardized and its variance is equal to 1), or

F5) $$L_X \approx [\rho_X]^{1/2}$$

otherwise, where ≈ indicates "approximately equal to."

Finally, Anderson and Gerbing (1988) pointed out that for a unidimensional measure there is little practical difference between coefficient alpha (α) and latent variable reliability ρ.

Thus, for a single indicator specification of a standardized latent variable *X* (i.e., its variance is fixed at 1), its loading, L_X, is the square root of its latent variable reliability ρ (see Equation F1), and its measurement error variance, E_X, is 1 - ρ_X (see Equation F2).

These parameters can be estimated for a standardized latent variable *X* by substituting coefficient alpha reliability, α, into Equations (F2) and (F4), and for an unstandardized latent variable *X* its single indicator loading can be approximated using Equation (F5) and α , and its measurement error variance can be estimated using Equation F2.

Ping (2004) suggested summing maximum likelihood exploratory factor analysis (EFA) loadings for L_X, and using Equation F2 for the measurement error variance of a fully aggregated measure. Because EFA also produces an estimate of the Average Extracted Variance (AVE), the explained variance for a factor, the equation for the AVE of *X* (see Fornell & Larker, 1981)

$$AVE_X = \frac{\Sigma(l_{xj\ (j=1,p)})^2\, Var(X)}{\Sigma(l_{xj\ (j=1,p)})^2\, Var(X) + E_X}$$

$$= \frac{\Sigma(l_{xj\ (j=1,p)})^2}{\Sigma(l_{xj\ (j=1,p)})^2 + E_X} \ ,$$

where l_{xj} is a loading, Σ is a sum (of squared loadings--Equation F1 involves the square of the sum), and Var(*X*) = 1, can be solved for the measurement error variance of a sum of indicators, E_X

$$\text{F6)} \qquad E_X = \frac{\Sigma(l_{xj\ (j=1,p)})^2\, (1 - AVE)}{AVE} \ ,$$

where AVE is the explained variance of (unidimensional) factor containing the items in X.

Thus, an additional estimate of the loading of a summed indicator composed of unidimensional items is the sum of its EFA loadings, and an additional estimate of its measurement error variance is given by Equation (F6).

ENDNOTES

[1] In this case inconsistency is actually unacceptable consistency: the observed or input correlation between two indicators of the same latent variable is not acceptably numerically similar to the path analytic (Wright, 1934) product of the coefficients on the loading paths from their common latent variable.

[2] Using path analysis (Wright, 1934) the covariance of two unidimensional items x_1 and x_2 (i.e., items with only one underlying latent variable) implied by a model is the product of the path coefficients on their paths from their common latent variable (i.e., the product of their loadings), *plus* the product of the path coefficients due to correlated measurement error (i.e., $\lambda_1 * \lambda_2$ + $1*1*\mathrm{Cov}(\varepsilon_1,\varepsilon_2)$, where λ denotes loading, ε denotes measurement error, 1 is the implied path coefficient on the path between each measurement error and their respective x, and Cov denotes covariance). Because correlations between measurement errors are usually assumed to be zero, the covariance term is usually ignored.

[3] Content or face validity is usually established by qualitatively judging how well items match the conceptual definition of the target construct.

[4] A second-order construct has other constructs as its "indicators." For example in Dwyer and Oh's (1987) study of Environmental Munificence and Relationship Quality, the second-order construct Relationship Quality had the first-order constructs Satisfaction, Trust, and Minimal Opportunism as indicators (see Bagozzi, 1981b; Bagozzi and Heatherton, 1994; Gerbing and Anderson, 1984; Gerbing, Hamilton and Freeman, 1994; Hunter and Gerbing, 1982; Jöreskog, 1970; and Rindskopf and Rose, 1988 for accessible discussions of second-order constructs).

[5] Bagozzi and Heatherton (1994) also used a variation of this approach that did not use reliability loadings.

[6] For example, a model with just-identified latent variables (3 items per latent variable), and 5 latent variables requires 240 cases to produce at least two cases per input covariance matrix element. The same model with 5 summed indicators and 240 cases would have 16 cases available to compute each input covariance matrix element.

[7] The study details have been omitted to skirt matters such as conceptual definitions, hypotheses, etc. which were judged to be of minimal importance to the present purposes.

[8] Anderson and Gerbing (1984) suggested that GFI and AGFI may not be appropriate gauges of model-to-data fit in larger models. An RMSEA (Steiger, 1990) of .05 suggests close fit and values through .08 suggest acceptable fit--see Browne and Cudeck (1993); Jöreskog (1993).

[9] However, this indeterminacy could be remedied by reconvening an item-judging panel to judge the resulting measures and thus identify a weeded subset of items that best taps the conceptual definition of N.

[10] The variations of full aggregation (e.g., factor scores, LISREL 8 parameters, etc.) have not been formally investigated for structural coefficient bias and inefficiency, as far as we know, and the structural coefficient results from External Consistency Only and full aggregation should be compared for validation. A disagreement in nonsignificance (e.g., a structural coefficient varies between nonsignificance and significance with External Consistency Only and factor scores) should probably be judged nonsignificant.

[11] Other omissions were possible--e.g., omitting a Subset 2 item from the set of 18, etc., and there were 18 of these, some of which may have duplicated Subset 1.

[12] While these two itemizations had no items in common, equivalent behavior was observed with Subset 2 and Subset 1 or Subset 1a that did have common items. Parenthetically, the reliability of the antecedent latent variable was .86. However, similar behavior was observed for a latent variable with much lower reliability. Finally, there were other structural coefficients that were completely unaffected by changes in the itemizations of N.

[13] This process was actually repeated more than 180 times to check for errors because the results appeared to be cycling. Whether or not this process would have converged with this number of potential correlated measurement errors (171) is unknown.

[14] Correlating all measurement errors with ε_9 was not identified, so the correlation between ε_9 and ε_{13} was fixed at zero because its modification index (MI) was .0001, suggesting the correlation was nonsignificant. A MI in this case is approximately a Chi-Square statistic for freeing the correlation between ε_9 and ε_{13}--a MI of .0001 suggests the path coefficient on the correlation between ε_9 and ε_{13} would have a Chi-Square difference (from 0) of .0001 which is nonsignificant with a p-value of .992 and 1 degree of freedom.

[15] Using expectation algebra and the usual assumptions regarding latent variables and their errors of measurement, the variance of a sum of indicators x_i, $\mathrm{Var}(x_1+x_2+...+x_p) = \mathrm{Var}(\lambda_{x1}X + \varepsilon_{x1} + \lambda_{x2}X + \varepsilon_{x2} + ... + \lambda_{xp}X + \varepsilon_{xp}) = (\Sigma\lambda_{xj\ (j=1,p)})^2\mathrm{Var}(X) + \Sigma\mathrm{Var}(\varepsilon_{xj\ (j=1,p)})$, where λ is a loading, X is a latent variable, and ε is a measurement error.

[16] Because exploratory factor analysis assumes variances of 1, the loadings were re-scaled by dividing each loading by the maximum loading to allow for latent variable variances other than 1. This produces one loading equal to one and the other loadings in the customary .6 to .9 range.
[17] In retrospect, we should have at least numbered the scenarios by treatment.

SECOND-ORDER
LATENT VARIABLE INTERACTIONS,
AND SECOND-ORDER LATENT VARIABLES

Robert Ping, Assoc. Prof.
Department of Marketing
266 Rike Hall
Wright State University
Dayton, OH 45435
937-775-3047
rping@wright.edu

SECOND-ORDER
LATENT VARIABLE INTERACTIONS,
AND SECOND-ORDER LATENT VARIABLES

ABSTRACT

Because there is little guidance for estimating interactions involving a second-order latent variable (a latent variable with other latent variables as its "indicators"), the paper explores second-order latent variables and the estimation of a second-order interaction. It suggests a specification for an interaction between a second-order latent variable and a first-order latent variable (a latent variable with observed indicators). It also illustrates the estimation of these interactions/quadratics using real-world data.

FIRST-ORDER-BY-SECOND-ORDER LATENT VARIABLE INTERACTIONS, AND SECOND-ORDER LATENT VARIABLES

Tests of hypothesized latent variable interactions (e.g., XZ in

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \zeta_Y \, , \qquad (1$$

where X and X are latent variables, β_1 through β_3 are unstandardized structural coefficients, β_0 is an intercept, and ζ_Y is the structural disturbance term) in theoretical (hypothesis testing) models have appeared in substantive articles (e.g., Bansal & Taylor, 2002; Bhuian, Menguc & Borsboom, 2005; Bisbe & Otley, 2004; Cadogan, Cui & Li, 2003; Capaldi & Stoolmiller, 1999; De Ruyter & Wetzels, 2000; Featherman & Pavlou, 2003; Fullerton & Taylor, 2002; Gustafsson, 1997; Harris, Mowen & Brown, 2005; Huang, Lu & Wong, 2003; Iglesias, 2004; Jensen & Szulanski, 2004; Kuklinski & Weinstein, 2001; Lam, 1999; Lee & Ganesh, 1999; Lusch & Brown, 1996; Masterson, 2001; Matsuno, Mentzer & Ozsomer, 2002; Ozsomer & Simonin, 2004; Ping, 1999; Porter & Lilly, 1996; Singh, 1998; Taylor, 1997; Torsheim, Aaroe & Wold, 2001; and Zhou, Yim & Tse, 2005).

But suppose Z is a second-order latent variable (it has "indicators" that are latent variables). How should XZ be specified? Extant interaction specification guidance is exclusively for first-order-by-first-order interactions.

Second-order (confirmatory) latent variables (LV's) were proposed by Jöreskog (1970) (see Thurstone, 1947). These LV's have other LV's as their "indicators" (e.g., Z in Figure 1 and 1a). Each of these "indicator" LV's has observed indicators as usual. For example Dwyer and Oh (1987) proposed that the second-order LV Relationship Quality had as its "indicators" the first-order LV's Satisfaction, Trust, and Minimal Opportunism (guileful self-interest seeking) (see Bagozzi, 1981a; Bagozzi & Heatherton, 1994; Gerbing & Anderson, 1984; Gerbing, Hamilton & Freeman, 1994;

Hunter & Gerbing, 1982; Jöreskog, 1971a, 1971b; and Rindskopf & Rose, 1988 for discussions of second-order LV's).

Although comparatively rare, second-order latent variables in the Jöreskog-Thurstone sense have been reported in substantive articles (e.g., Bagozzi, 1981a, 1981b; Bagozzi & Heatherton, 1994; Dwyer & Oh, 1987; Gerbing, Hamilton & Freeman, 1994; Ping, 1994, 1997, 1999; Weeks, 1980). Rarer still is a second-order interaction, perhaps because there is little guidance for its estimation.

Because some authors believe interactions are more likely than their reported occurrence suggests (e.g., Jaccard, Turrisi & Wan, 1990; see the citations in Aiken & West, 1991), this paper explores an interaction between a second-order latent variable and a first-order latent variable. The paper suggests a specification for this interaction. It begins with a discussion of second-order latent variables, which leads to a suggested specification. Along the way the paper touches on latent variable interactions in general. It concludes with a pedagogical example using real-world data.

SECOND-ORDER LATENT VARIABLES

A second-order latent variable (LV) involves first-order LV's. A first-order LV can be conceptualized as a unidimensional set of items in an exploratory factor analysis. The relationship between indicators and their unobserved LV typically assumes the unobserved LV "drives" the indicators (i.e., these indicators are assumed to be observable instances or manifestations of their unobserved LV, and thus changes in the unobserved LV are "indicated" by observable changes in the items in its measure--see Bagozzi, 1984).

A second-order LV can be conceptualized as multiple factors in an exploratory factor analysis (e.g., in Figure 1: 1a, Factor 3 = Z_3 , etc.) that are not particularly orthogonal. Ideally, when the items in each of these factors are summed, an exploratory factor analysis of the resulting summed items is unidimensional. A second-order LV also "drives" its factors, and these "indicator" first-order LV's in turn "drive" their observed indicators. A diagram of a second-order LV shows it specified or connected to its "indicator" first-order LV's with paths or arrows from the second-order LV to its "indicator" first-order LV's, then arrows from the first-order "indicator" LV's to their observed indicators (see Z in Figure 1 and 1a). Because the relationships between a second-order LV and its "indicator" LV's is imperfect, the "indicators" have "measurement errors" (ζ's) that are the structural disturbances in their relationships with their second-order LV (see Figure 1:1a).

As the Dwyer and Oh (1987) example suggests, a second-order LV can be used to in effect combine several related (first-order) LV's into a single higher-order LV to simplify a structural equation model. Although other uses for second-order latent variables have been proposed (e.g., Gerbing & Anderson, 1984; Rindskopf & Rose, 1988), their use in published substantive articles has been primarily to simplify a structural equation model by combining first-order LV's into a single (higher order) LV.

Unidimensionality

In theoretical model (hypothesis) tests, the unidimensionality, reliability and validity of the LV's representing the model constructs are important to evaluating the operationalization of the model constructs. Thus, a second-order LV should be unidimensional, reliable and valid in a theoretical model test.

In substantive articles the unidimensionality of a first-order LV is typically demonstrated in several steps. Its internal consistency is established by a single-construct measurement model (MM) (a MM containing only the LV and its observed indicators) fitting the data (see Anderson & Gerbing, 1988). Next, the external consistency of pairs of these internally consistent LV's are evaluated using model-to-data fit in a series of two-construct MM's (MM's containing only two LV's and their observed indicators, with each LV specified unidimensionally). Then, trios of these LV's are evaluated in similar three-construct MM's, then sets of four LV's, etc. (see Jöreskog, 1993). Finally, the external consistency of all the LV's together is evaluated using their model-to-data fit in a full MM of all the study LV's specified unidimensionally (see Anderson & Gerbing, 1988).

Anecdotally, after the internal consistency of the model LV's has been established, the consistency of pairs, trios, etc. of LV's are sometimes skipped to save time, and the full MM is evaluated next because internally consistency LV's may be sufficient for the full MM to fit the data acceptably in real-world data. However, if there are indicators that are unacceptably externally inconsistent (i.e., if "properly" specified they would load significantly on more than one LV), they could be found using the modification indices of the fixed loadings (i.e., a large modification index for a fixed loading would suggest an externally inconsistent indicator), and these indicators could be excized.

Consistency for a second-order LV exists at several levels. Gerbing, Hamilton and Freeman (1994) argued that for a second-order LV internal consistency should be established at the second-order level. They suggested in effect that first-order "indicator" LV's should be unidimensional in the exploratory (common) factor analysis sense rather than internally consistent. Then, the second-order LV should be (internally) consistent in its single factor measurement model (i.e., one with only the second-order LV, its "indicator" first-order LV's, and their observed indicators specified unidimensionally).

Other approaches to establishing second-order consistency are possible. The internal consistency of the "indicator" LV's could be established (i.e., each "indicator's" single construct measurement model (MM) fits the data). Then, a single factor second-order MM (i.e., one with only the second-order LV, its "indicator" LV's, and their observed indicators) could be estimated to establish the internal consistency of the second-order LV. Next, MM's with pairs of these internally consistent LV's could be estimated treating the second-order LV as just another LV (i.e., it does not have to appear in every pair of LV's). Then, trios of LV's could be evaluated, etc. culminating in a full MM containing the second-order LV and all the other LV's. A shortcut would be to skip the estimation of pairs, etc. of LV's, and estimate the full MM, checking for unacceptable multiple loaders using the modification indices of fixed loadings.

Discussion In effect, correlated measurement errors could be used with "indicator" LV's by correlating their structural disturbance terms. Assuming the resulting model is identified, this approach would improve model-to-data fit in a second-order LV. However, most of the criticisms of correlated measurement errors among observed indicators would also apply in this case (e.g., Bagozzi, 1983; Fornell, 1983; Gerbing & Anderson, 1984).

Internal consistency in first-order LV's is a stricter form of unidimensionality that usually limits the number of indicators per latent variable to about six in real-world data (see Anderson & Gerbing, 1984; Bagozzi & Heatherton, 1994; Gerbing & Anderson, 1993). The resulting measures tend to be operationally narrow instances of their target construct (i.e., content or "face" validity is sacrificed for consistency--see Cattell, 1973, 1978; see Gerbing, Hamilton & Freeman, 1994). Thus Gerbing, Hamilton and Freeman's (1994) suggestion of using unidimensional "indicators" in the exploratory factor analysis sense, instead of internally consistent "indicators," is substantively appealing because it should result in fewer items deleted from "indicator" latent variables (Gerbing, Hamilton and Freeman 1994 reported one second-order LV with a 16-item "indicator").

If item deletion is required to attain consistency, it should be done with concern for the content or "face" validity of the resulting LV. Specifically, in establishing external consistency, the item with the largest significant modification index (MI) of the fixed loadings should not be deleted without weighing the deletion's impact on content validity. Similarly, in establishing internal consistency, an item should not be deleted without weighing the deletion's impact on content validity.

Reliability

One formula for LV reliability of X, with indicators $x_1, x_2, \ldots, x_n$ is

$$\text{Latent Variable Reliability} = \frac{(\Sigma\lambda_i)^2 \mathrm{Var}(X)}{(\Sigma\lambda_i)^2 \mathrm{Var}(X) + \Sigma \mathrm{Var}(\varepsilon_i)}\ , \tag{2}$$

where λ_i is the loading of x_i on X, ε_i is the measurement error for x_i, and Σ denotes a sum (Werts, Linn and Jöreskog, 1974; see Bagozzi, 1980; Bollen, 1989; Dillon & Goldstein, 1984; Fornell & Larker, 1981).

However, Gerbing and Anderson (1988) pointed out for unidimensional LV's there is little practical difference between coefficient alpha (Cronbach, 1951) and Latent Variable Reliability.

Rindskopf and Rose (1988) distinguished between second-order- and first-order reliability (i.e., "indicator" LV's each have a reliability, and a second-order LV has a reliability due to its "indicators"). Thus, to gauge the reliability of a second-order LV, it may be sufficient to gauge reliability at the second-order level (acceptably reliable first-order LV's can have unacceptably reliable indicators--ones with loadings, that when squared reveal a measure of item reliability, that are less than the square root of the customary "cut-off" for acceptable reliability, .7). This is accomplished using Equation 2 by substituting the "indicator" loadings, γ's, for λ's, and "indicator" structural disturbance terms, ζ's, for ε's.

Discussion "Indicator" LV's could be weeded (i.e., delete items with due care for content validity) to improve their coefficient alpha.

Experience with second-order LV's suggests their reliabilities can be low. (e.g., Rindskopf and Rose 1988 reported second-order reliabilities as low as .309 presumably using Equation 2).[1]

Validity

Authors in the Social Sciences disagree on what constitutes an adequate demonstration of validity (e.g., Bollen, 1989; Campbell, 1960; DeVellis, 1991; Heeler and Ray, 1972; Nunnally, 1978; Peter, 1981), and this is reflected in the published second-order substantive articles. Nevertheless, a minimal demonstration of the validity of a second-order LV should probably include the following: its content or face validity (how well its "indicators" tap into the conceptual definition of the second-order construct), its construct validity, and its convergent and discriminant validity (e.g., Bollen, 1989; DeVellis, 1991; Nunnally, 1978). The "validity" of the second-order LV would then be qualitatively assessed considering its reliability and its performance over the above minimal set of validity criteria.

Construct validity is concerned in part with an LV's correspondence (i.e., correlation) with other LV's. To begin to suggest construct validity, the other LV's in the study should be valid and reliable, and their correlations with the target LV (e.g., significance, direction and magnitude) should be theoretically sound.

Convergent and discriminant validity are Campbell and Fiske's (1959) proposals involving the measurement of multiple constructs with multiple methods, and they are frequently considered to be additional facets of construct validity. Convergent measures are highly correspondent (e.g., correlated) across different methods. Discriminant measures are internally convergent. However, convergent and discriminant validity are frequently not assessed in substantive articles as Campbell and Fiske (1959) intended (i.e., using multiple traits and multiple methods-- see Bollen, 1989; Heeler & Ray, 1972). Perhaps because constructs are frequently measured with a single method (i.e., the study at hand), reliability is frequently substituted for convergent validity, and LV correlational distinctness (e.g., correlations with other measures less than .7) is substituted for discriminant validity.

Discussion Fornell and Larker (1981) suggested that adequately convergent LV's should have measures that contain more than 50% explained variance in the factor analytic sense (also see Dillon and Goldstein, 1984). Their Average Variance Extracted (AVE)

$$AVE = \frac{\Sigma\lambda_i^2 Var(X)}{\Sigma\lambda_i^2 Var(X)+\Sigma Var(\varepsilon_i)} \quad , \qquad (3$$

where λ, ε and Σ are as in Equation (2), can be used to gauge percent explained variance in an LV. Unfortunately, acceptably reliable LV's can have less than 50% explained variance (AVE). Thus, a convincing demonstration of convergent validity would be an AVE of .5 or above.

Although there is no firm rule for demonstrating discriminant validity, correlations with other LV's less than |.7| are frequently accepted as evidence of discriminant validity. A larger correlation can be tested by examining its confidence interval to see if it includes 1 (see Anderson and Gerbing, 1988). It can also be tested by using a single-degree-of-freedom test that compares two measurement models, one with the target correlation fixed at 1, and a second with this correlation free (see Bagozzi and Phillips, 1982). If the difference in resulting chi-squares is significant, this suggests the correlation is not 1, and this suggests the LV's are correlationally distinct.

AVE can also be used to gauge discriminant validity (Fornell and Larker 1981). If the squared (error-disattenuated) correlation between two LV's is less than either of their individual AVE's, this suggests the LV's each have more internal (i.e., extracted) variance than variance shared with other LV's. This in turn suggests discriminant validity.

Thus, in addition to being reliable at least at the second-order level, a second-order LV should be at least content or "face" valid, its correlations with other LV's should be theoretically sound, and it should be convergent and discriminant valid (e.g., using AVE). Second-order AVE could be computed using Equation 3 by substituting the "indicator" loadings on their second-order LV, γ's, and "indicator" structural disturbance terms, ζ's, for ε's.

Second-order AVE's are seldom published. However, AVE is much less than reliability (see Equations 2 and 3), and experience suggests a rough estimate of AVE is the cube of reliability. Thus, reliability should probably be above .8 to avoid an AVE below .5. Because second-order reliabilities in several published articles were less than .8, it is likely that several second-order LV's in published articles were not convergent valid in the Fornell and Larker (1981) sense.

SECOND-ORDER INTERACTIONS

First Orders

The amount of interaction between X and Z in their association with Y (also termed X's moderation of the Z-Y association, or Z's moderation of the X-Y association) is the strength (i.e., the magnitude) of the coefficient of XZ, β_3, in an equation such as Equation 1.

Specification There have been several proposals for specifying a latent variable interaction (e.g., Hayduk, 1987; Jaccard & Wan, 1995; Kenny & Judd, 1984; etc.). The most frequently encountered specification in substantive articles was suggested by Ping (1995). This specification uses a single indicator for an interaction, XZ, that is the product of the sum or an average of the indicators of X and the sum or an average of the indicators of Z. Under the Kenny and Judd (1984) normality assumptions and expectation algebra,[2] the loading, λ_{xz}, and error variance, $Var(\varepsilon_{xz})$, for this single indicator of XZ with averaged indicators (so the variance of XZ is commensurate with X and

Z) are

$$\lambda_{xz} = \Lambda_X\Lambda_Z/mn, \tag{4}$$

and

$$Var(\varepsilon_{xz}) = [\Lambda_X^2 Var(X)Var(\varepsilon_z) + \Lambda_Z^2 Var(Z)Var(\varepsilon_X) + Var(\varepsilon_X)Var(\varepsilon_z)]/(mn)^2 , \tag{4a}$$

where m and n are the number of indicators of X and Z respectively, Var(X) and Var(Z) are error-dissattenuated variances, λ's are loadings, $\Lambda_X = \lambda_{x1}+\lambda_{x2}+...+\lambda_{xm}$, $\Lambda_z = \lambda_{z1}+\lambda_{z2}+...+\lambda_{zn}$, ε's are measurement errors, $Var(\varepsilon_X) = Var(\varepsilon_{x1})+Var(\varepsilon_{x2})+...+Var(\varepsilon_{xm})$, and $Var(\varepsilon_X) = Var(\varepsilon_{z1})+Var(\varepsilon_{z2})+...+Var(\varepsilon_{zn})$. Equations (4) and (4a) can be estimated directly using some structural equation software (e.g., LISREL 8 or CALIS), or in several steps using measurement model estimates of the parameters in Equations (4) and (4a) (see Ping 1998a).

Factored Coefficients Equation 1 can be factored to produce a coefficient of Z due to the interaction XZ (i.e.,

$$Y = \beta_0 + \beta_1 X + (\beta_2 + \beta_3 X)Z + \zeta_Y) \tag{5}$$

(see Aiken & West 1991). Similarly Equation 1 can be re-factored to produce a coefficient of X due to the interaction XZ (i.e., $(\beta_1 + \beta_3 Z)X$). These factored coefficients are instrumental in interpreting interactions in survey data. When β_3 is significant, depending on the signs and magnitudes of β_2 and β_3, the (factored) coefficient of Z, $(\beta_2+ \beta_3 X)$, can be of one sign (e.g., positive) for X at one end of the range of X in a study, zero near the middle of the range of X, and another sign (e.g., negative) at the other end of the range of X in the study.

The standard error of the factored coefficient of Z also varies over the range of X in a study. Determined by the square root of $Var(\beta_2+ \beta_3 X)$ and using expectation algebra, the standard error of the factored coefficient of Z is

$$[Var(\beta_2) + X^2 Var(\beta_3) + 2XCov(\beta_2,\beta_3)]^{1/2}, \tag{6}$$

where Var(β) is the square of the standard error of β, and Cov indicates covariance (e.g., Jaccard, Turrisi & Wan, 1990). The standard error of the factored coefficient of X is similar. Thus, the factored coefficient of Z, for example, can be significant for some X in a study but nonsignificant for other values of X in the study.

Second-Orders

The possibilities for specifying a second-order by first-order interaction are numerous, but most of them are impractical. For example in Figure 1, specifications involving all possible products of individual indicators (e.g., Kenny & Judd, 1984) (e.g., for Z_3, $x_1z_{3,1}$, $x_1z_{3,2}$, ... , $x_1z_{3,n}$, $x_2z_{3,1}$, $x_2z_{3,2}$, ... , $x_mz_{3,n}$, where x_j are the indicators of X, the first-order LV, and $z_{i,k}$ are the indicators of Z_i, the second-order LV (see Figure 1) is rarely consistent enough to avoid spoiling structural model fit (see Jaccard and Wan, 1995 for evidence of this difficulty).

Thus, we will explore specifications that reduce the number of second-order interaction indicators. For example Z, a second-order LV with three "indicators" could be respecified as a first-order LV by replacing the "indicator" Z_1 by the sum of its indicators, and doing the same for Z_2 and Z_3 (see Figure 1:1c). This respecification of a second-order LV has been reported (e.g., Dwyer & Oh, 1987; Ping, 1999). XZ_i in Figure 1 could then be specified with the indicators $x{\bullet}z_1 = (x_1+x_2+...+x_m)(z_{1,1}+z_{1,2}+...+z_{1,p})$, $x{\bullet}z_2 = (x_1+x_2+...+x_m)(z_{2,1}+z_{2,2}+...+z_{2,q})$, and $x{\bullet}z_3 = (x_1+x_2+...+x_m)(z_{3,1}+z_{3,2}+...+z_{3,m})$ (Specification 1), where the sums in the parentheses represent the result of summing variables in each case. These indicators are comparatively few in number (in this case three), they have Equation 4 and 4a loadings and measurement error variances, and their observed values, $x{\bullet}z_i = (x_1+x_2+...+x_m)(z_{i,1}+z_{i,2}+...+z_{i,n})$ can be computed for each case in a data set.

Alternatively, XZ in Figure 1 could be specified with a single indicator $x{\bullet}z = (x_1+x_2+...+x_m)(\Sigma z_{1,i}+\Sigma z_{2,i}+\Sigma z_{3,i})$, where the sums in the parentheses represent the result of summing variables in each case, and $\Sigma z_{j,i}$ is the sum of the indicators of Z_j (i.e., a sum of sums) (Specification 2). This indicator has Equation 4 and 4a loadings and measurement error variances, and its observed values (i.e., $(x_1+x_2+...+x_m)(\Sigma z_{1,i}+\Sigma z_{2,i}+\Sigma z_{3,i})$) can be computed for each case in a data set.

A variation of these approaches would be to specify Z in Figure 1 as a first-order LV by replacing Z_1 by its (confirmatory) factor scores, and doing the same for Z_2 and Z_3 (see Jöreskog & Sörbom, 1996, and Kim & Mueller, 1978). The resulting XZ interaction would then have in this case three indicators, $x{\bullet}z_i = (x_1+x_2+...+x_m)f_i$ (i=1,3) (Specification 3), or a single indicator $x{\bullet}z = (x_1+x_2+...+x_m)(f_1+f_2+f_3)$ (Specification 4), where f_i is the factor score for F_i. These indicators are comparatively few in number, they have Equation 4 and 4a loadings and measurement errors, and their observed values (i.e., $(x_1+x_2+...+x_m)f_i$ or $(x_1+x_2+...+x_m)(f_1+f_2+f_3)$) can be computed in each case once factor scores are available.

Reliability and Validity The reliability, ρ_{XZ}, of a second-order (or first-order by first-order) interaction, XZ, is approximately the product of the reliabilities of X and Z:

$$\rho_{XZ} = \frac{r_{XZ}^2 + \rho_X \rho_Z}{r_{XZ}^2 + 1}, \qquad (7$$

where ρ denotes reliability and r_{XZ}^2 is the error-disattenuated correlation of X and Z (Bohrnstedt & Marwell, 1978). Thus, a second-order (or first-order) interaction composed of reliable LV's could be unreliable.

The demonstration of the validity of a second-order interaction, XZ, is less tedious than for X or Z. XZ is content or face valid if X and Z are content valid and the specification of XZ includes all the indicators of X and Z. The construct (correlational) validity of any interaction is usually impossible to evaluate. In addition, the formula for the Average Variance Extracted (AVE) of an interaction is unknown. Nevertheless, the experience-based cube-of-reliability rough approximation of AVE could be used to gauge the convergent validity of XZ. Discriminant validity could be gauged using the "correlations less than |.7|" criterion.

AN EXAMPLE

For pedagogical purposes a real-world data set will be reanalyzed. A survey involving the Figure 2 model and the first-order LV's U, V and W, T the second-order LV, with "indicators" A, I and C, and the second-order by first-order interaction UxT, produced more than 200 usable responses.[3]

Unidimensionality

First, the internal consistency of T the second-order LV was established at the first-order level. Specifically, the internal consistency of the first-order "indicator" A, for example, was established by estimating a single construct measurement model (MM) for A (i.e., a measurement model containing only the indicators of A) and omitting the indicator with the largest sum of Modificaton Indices (MI's) without regard to sign (Ping, 1998b; see Ping, 2004). The single construct MM with the remaining indicators of A was then estimated, and the indicator with the resulting largest sum of MI's without regard to sign (S-MI's) was omitted. This process of omitting, re-estimating, and then omitting the indicator with the resulting largest S-MI in each re-estimation was repeated until the MM's RMSEA (Steiger 1990, see Brown and Cudeck, 1993; Jöreskog, 1993) was .08 or less (.08 or less suggests adequate model-to-data fit). This process was repeated for I and C.

Then, the internal consistency of T the second-order LV was verified using a MM that excluded all the model variables except T, its "indicators," and the observed indicators of these "indicators" specified unidimensionally (see Figure 2:2a). This MM was judged to fit the data (χ^2/df/p-value = 110/51/0, GFI = .92, AGFI = .88, CFI = .97, RMSEA = .07) (GFI and AGFI may be inadequate for fit assessment in larger models--see Anderson and Gerbing, 1984), and thus T was judged to be internally consistent.

Next, the internal consistency of U, V and W was established using the S-MI procedure just used on A, I and C. Then, the external consistency of all the model variables, except the interaction UxT was established using a full MM for T, U, V and W with all LV's specified unidimensionally (χ^2/df/p-value = 449/266/0, GFI = .86, AGFI = .83, CFI = .96, RMSEA = .05) (.05 or less suggests close model-to-data fit--see Brown and Cudeck, 1993).

Reliability

Then, the reliability of the LV's was gauged. Coefficient alpha was calculated for each first-order variable (see Anderson and Gerbing, 1988), and these variables were judged to be reliable (U had a coefficient alpha of .943, and the other LV's had coefficient alpha's of .85 or above). The reliability of T (.709) was calculated using Equation 2 (with the "loadings," γ"s, of A, I and C on T, and the "measurement errors," ζ's, of A, I and C) and it was judged to be reliable.

Validity

The first-order LV's were judged to be content or "face" valid. Using a full MM (excluding UxT), they were also judged to be construct valid. Next, using Equation 3, and full MM (excluding UxT) parameter estimates, the convergent validities of U, V and W were judged to be adequate (the AVE of U was .770 and the other AVE's were .5 or above). The convergent validity of T was low (AVE = .474) but sufficient for these pedagogical purposes. Then, using the (error-dissattenuated) correlations among T, U, V and W from the same MM, the discriminant validities of T, U, V and W were judged to be adequate.

Discussion

Indicator weeding was required for A, I, C, U, V and W, and changes in content or "face" validity was a major concern.

The consistency of T the second-order LV was established at the first-order level because T was not consist at the second-order level otherwise (i.e., T's MM with unidimensional "indicators" in the exploratory factor analysis sense, rather than consistent "indicators," did not fit the data).

Pairs, trios, etc. of the first-order LV's were not examined for external consistency because the full MM for T, U, V and W fit the data. However, this MM produced significant MI's for the fixed loadings. Thus, there were LV's, in this case two LV's with three indicators, that were not "perfectly" externally consistent. Because the expected changes in the loadings if these significant loadings were freed were comparatively small (i.e., their unstandardized loadings on "alien" LV's were .36 or less), deleting the offending indicators was judged to materially impair the content or face validity of their LV (for one LV the indicators would be reduced from 4 to two), and the full MM model already fit the data with these externally inconsistent indicators, the offending indicators were not excised. Experience suggests that "perfect" external consistency is frequently impossible to attain in real-world data without severely affecting content or face validity, and sometimes producing just- or under-determined LV's. While there is no hard-and-fast rule, a few fixed loadings with an expected change if they were freed less than about .4 in unidimansional LV's could probably be ignored in real-world data if the full MM already fits the data (i.e., the full MM is acceptably externally consistent).

Respecification of T

To investigate the respecification of T as a first-order LV, each "indicator" of T was summed then averaged, and a MM corresponding to Figure 2:2b was estimated. To gauge the equivalence of this alternative specification of T to its second-order specification, its reproduced covariance matrix was compared to that produced by the second-order specification. However, this respecification of T substantially over-estimated the variance of T produced by second-order specification (see Table G portions (1) and (2)) (weighted averages of the indicators performed similarly), so a summed indicator specification of T was not pursued further.

Next, each "indicator" of T was replaced by its factor score. These factor scores were produced in a MM corresponding to Figure 2 without the XZ interaction present, and with the Figure 2:2a second-order specification of T. Specifically, they were computed in each case for the "indicator" A, for example, by averaging the factor score (regression) of A on u_1, u_2, ... , u_5, a_1, a_2, ... , a_4, i_1, ... , i_4, c_1, ... , c_4, v_1, ... , v_4, w_1, ... , w_4 (i.e., the factor score of A, f_A, was $(\omega_{A,1}x_1+\omega_{A,2}x_2+ ... +\omega_{A,25}x_{25})/ (\omega_{A,1}+\omega_{A,2}+ ... +\omega_{A,25})$ where the ω's are the factor weights and the x's are the indicators u_1, u_2, ... , w_4) (averages were used to produce the same metric as the indicators). For the "indicator" I, f_I was $(\omega_{I,1}x_1+\omega_{I,2}x_2+ ... +\omega_{I,25}x_{25})/(\omega_{I,1}+\omega_{I,2}+ ... +\omega_{I,25})$, and the factor scores for the "indicator" C were computed similarly. The MM corresponding to Figure 2 without the XZ interaction, and with the Figure 2:2c factor-score specification of T (A= f_A, etc. in Figure 2:2c) was then estimated, and the reproduced covariance matrix of this respecification of T was equivalent to that produced by the second-order specification of T (see Table G portions (1) and (3)). This respecification of T was also (trivially) unidimensional (i.e., it fit the data exactly).

Centering

At this point each indicator should have been mean centered by subtracting the indicator's average from its value in each case. Mean centering exogenous LV's is important to reduce collinearity between an interaction, XZ, and its constituent LV's, X and Z, and centering endogenous LV's is important to compensate for not estimating intercepts (see Jöreskog and Yang, 1996).

However, because factor scores were computed earlier, the indicators were mean centered before the factor scores were estimated. Parenthetically, because the factor scores were determined using indicators that had means of zero, the factor scores had means of zero and thus they were also mean centered.

Interaction Specification

Then, the Figure 2:2c first-orders-only MM (i.e., without UxT and with the first-order/factor-scored T) was examined for external consistency (model-to-data fit) in order to use its parameters in Equation 2 and 2a for the specification of UxT. This MM was judged to fit the data (χ^2/df = 168/98, GFI = .91, AGFI = .88, CFI = .96, RMSEA = .05).

Next, UxT was specified using Specification 3 ($u \bullet a = (u_1+u_2+...+u_5)f_A$, etc.) and Equations 4 and 4a with parameter estimates from the above first-orders-only MM. Then, the (full) measurement model corresponding to Figure 2:2c (with UxT) was estimated to verify the external consistency of the latent variables T, U, V, W and UxT. To accomplish this, starting values for the MM parameters, especially the covariances of the LV's, are sometimes required, and parameter estimates from the first-orders-only MM just estimated were used along with error-attenuated variance and covariance estimates for UxT. This measurement model was judged to fit the data (χ^2/df = 186/111, GFI = .91, AGFI = .88, CFI = .96, RMSEA = .05).

Reliability and Validity

Then, the reliability of UxT was calculated using Equation 7 with T in its first-order/factor-score specification, correlations from the (full) measurement model corresponding to Figure 2:2c (with UxT), and coefficient alphas in place of latent variable reliabilities, and UxT was judged to be reliable.[4]

Next, using this same model, reliabilities and average extracted variances (Fornell & Larker, 1981), the convergent validity UxT may have been low, but sufficient for these purposes.[5] Then, using the (error-dissattenuated) correlations among T, U, V and W from the same MM and a "correlations with other LV's less than |.7| criterion," the discriminant validity of UxT was judged to be acceptable.

Interaction Estimation

Then, despite the low convergent validities of T and possibly of UxT,[6] the Figure 2:2c structural model was specified. As with the (full) measurement model corresponding to Figure 2:2c just discussed, starting values for the structural model parameters were specified using a combination of the parameter estimates from the (full) measurement model corresponding to Figure 2:2c just discussed, plus structural coefficient estimates (i.e., the β's in Equation 1) and structural disturbance terms (e.g., ζ in Equation 1) from OLS regression (ζ_V and ζ_W are estimated by $1\text{-}R_V^2$ and $1\text{-}R_W^2$ respectively). The results using LISREL, Maximum Likelihood and 2-step estimation[7] are summarized in Table A.[8]

For completeness three additional estimations are reported. Table B summarizes the results of a specification that was identical to the Table A specification except that direct estimation of the UxT interaction using LISREL 8's constraint equations was used (see Ping, 1998a for sample LISREL 8 direct estimation code). The results were trivially different from the 2 step estimates shown in Table A.

Table D summarizes the results of T and U specified as before, but with UxT specified using the Kenny and Judd (1984) approach of using all possible unique products of the indicators of U and T specified as a first-order LV. This specification involved products of each of the 5 indicators of U with the 3 factor-scored indicators of T, for a total of 15 cross-product indicators, and the resulting measurement and structural models did not fit the data.[9] Parenthetically, no difficulty was encountered in estimating the structural model, and the results were approximately those shown in Tables A and B.

Table C summarizes the results of U specified as before, but with T specified as a second-order LV and UxT specified the Kenny and Judd (1984) approach of specifying UxT with all possible unique cross-products of the indicators of U and T. Because there were 5 indicators for U, and T had three first-order LV's each with four indicators, this produced 60 cross-product indicators for UxT. The resulting model also did not fit the data. Nevertheless, no difficulty was encountered in estimating this model, and the results were are also approximately those in Tables A and B.

The interpretation of the Table A UxT interaction results is presented in Table H and its Footnotes a and f.

DISCUSSION

The above example hints that a second-order LV could be adequately specified by replacing the second-order LV's "indicators" by their factor-scores. However, because factor scores are known to be approximate, simulations involving combinations of data conditions that are encountered in real-world surveys (e.g., reliability, coefficient size, correlations, sample size, etc.) would be required to demonstrate that factor scores (asymptotically) reproduce T. (Nevertheless, it is widely believed among applied social science researchers that factor scores can be used to adequately represent LV's.). Simulations would also be required to demonstrate that a first-order by second-order interaction specified using factor scores for T produces unbiased and consistent estimates (although this is likely because factor score indicators do not violate the assumptions underlying the Ping 1995 technique any more or less than observed indicators).

Kenny and Judd (1984) suggested constraining the variance of UxT, for example, to its Kendall and Stewart (1958) expectation algebra result of $\mathrm{Var(T)Var(U)+Cov}^2\mathrm{(T,U)}$. This constraint is reasonable because it is used to derive the UxT interaction loadings and measurement errors. However, its use in real-world data can produce measurement and structural models that do not converge. When this happens, the constraint is typically relaxed (i.e., Var(UxT) is freed). However, this was not necessary in the example (i.e., the measurement and structural models converged with the variance of UxT constrained to Var(T)Var(U)+Cov^2(T,U)). Nevertheless, the actual procedure used was to estimate each model with Var(UxT) free to obtain convergence, then constrain Var(UxT) and reestimate to see if the model still converged. Parenthetically, the Table A and B results with Var(UxT) unconstrained were trivially different from the Table A and B results (see Table F).

Factor scores for A, for example, were computed for each case using the factor weights for A, $\omega_{A,i}$, and the equation $f_A = \omega_{A,1}x_1+\omega_{A,2}x_2+ \ldots +\omega_{A,25}x_{25}$, where the ω's are the factor weights and the x's are the indicators u_1, u_2, ... , w_4. T had three "indicator" LVs (and thus three factor score equations, one for each "indicator" LV) and the model had twenty-five indicators (without the interaction). As a result, because writing the code in PRELIS, SPSS, SAS, etc. for the three factor scores in each case was tedious, three shortcuts might be attractive. First, just the indicators of A, for example (instead of all the model indicators) could have been used to compute its factor score and reduce the number of terms in f_A . However, it is easy to show that the resulting factor-scores would not adequately reproduce the covariance matrix of T specified as a second-order.

Second, the factor weights (ω's) could have been used "as is" (i.e., unaveraged). However, this would create difficulties later during interaction interpretation because the factor scores would not have the same metric as their indicators (i.e., LIKERT-scaled).

Third, UxT might have been specified with a sum of indicators for A, I and C, instead of factor scores. This specification was estimated (see Table E), and it produced structural coefficient estimates that were equivalent to the Tables A and B factor-scores results. Specifically, T was specified as a second-order LV and u•t was specified with a single indicator, u•t = $(u_1+u_2+...+u_5)(\Sigma a_j+\Sigma i_j+\Sigma c_j)$ (i.e., Specification 2).[10] However this result may have been circumstantial, and simulations involving combinations of data conditions that are encountered in real-world surveys are required to demonstrate that this indicator for a second-order interaction is unbiased and consistent (although this too is likely because this single indicator does not violate the assumptions underlying the Ping 1995 technique any more or less than any other observed indicators).

An improved shortcut might have been to estimate a first-order by second-order interaction model such as UxT by specifying T as a second-order LV (e.g., Figure 2:2a) and specifying u•t with sums of indicators (i.e., u•t = $[u_1+u_2+...+u_5][\Sigma a_j+\Sigma i_j+\Sigma c_j]$) to reduce the estimation effort involved with factor-scores. Then if any of the Table E t-values of the structural coefficients involved in the moderation (i.e., UxT-V, U-V and T-V) were in a neighborhood of 2 (i.e., ±□.10), the factor-score version of u•t would be preferred.

REFERENCES

Aiken, Leona S. and Stephen G. West (1991), *Multiple Regression: Testing and Interpreting Interactions*, Newbury Park, CA: SAGE Publications.

Anderson, James C. and David W. Gerbing (1984), "The Effect of Sampling Error on Convergence, Improper Solutions, and Goodness of Fit Indices for Maximum Likelihood Confirmatory Factor Analysis," *Psychometrika*, 49, 155-173.

_____ and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 103 (May), 411-23.

Bagozzi, Richard P. (1980), *Causal Models in Marketing*, New York: Wiley.

_____ (1981a), "An Examination of the Validity of Two Models of Attitude," *Multivariate Behavioral Research*, 16 (July), 323-359.

_____ (1981b), "Attitudes, Intentions and Behavior: A test of Some Key Hypotheses," *Journal of Personality and Social Psychology*, 41, 607-627.

_____ and Lynn W. Phillips (1982), "Representing and Testing Organizational Theories: A Holistic Construal," *Administrative Science Quarterly*, 27 (September), 459-489.

Bagozzi, Richard P. (1983), "Issues in the Application of Covariant Structure Analysis: A Further Comment," *Journal of Consumer Research*, 9 (March), 449-50.

_____ (1984), "A Prospectus for Theory Construction in Marketing," *Journal of Marketing*, 48 (Winter), 11-29

_____ and Todd F. Heatherton (1994), "A General Approach to Representing Multifaceted Personality Constructs: Application to Self Esteem," *Structural Equation Modeling*, 1 (1), 35-67.

Bansal, H. S. And S. F. Taylor (2002), "Investigating Interactive Effects in the Theory of Planned Behavior in a Service-provider Switching Context," *Psychology & Marketing*, 19 (5), 407-425.

Bhuian, S. N., B. Menguc and R. Borsboom (2005), "Stressors and Job Outcomes in Sales: a Triphasic Model Versus a Linear-quadratic-interactive Model," *Journal of Business Research*, 58 (2), 141-150.

Bisbe, J. and D. Otley (2004), "The Effects of the Interactive Use of Management Control Systems on Product Innovation," *Accounting Organizations and Society*, 29 (8), 709-737.

Bohrnstedt, George W. and Gerald Marwell (1978), "The Reliability of products of two random variables," in *Sociological Methodology*, K.F. Schuesser ed., San Francisco: Jossy-Bass.

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.

Browne, Michael W. and Robert Cudeck (1993), "Alternative Ways of Assessing Model Fit," in *Testing Structural Equation Models*, K. A. Bollen et al. eds, Newbury Park CA: SAGE Publications.

Cadogan, J. W., C. C. Cui, and E. Y. K. Li (2003), "Export Market-oriented Behavior and Export Performance - the Moderating Roles of Competitive Intensity and Technological Turbulence," *International Marketing Review*, 20 (5), 493-513.

Campbell, Donald T. and Donald W. Fiske (1959), "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin*, 56, 81-105.

Campbell, Donald T. (1960), "Recommendations for APA Test Standards Regarding Construct, Trait and Discriminant Validity," *American Psychologist*, 15, 546-553.

Capaldi, D. M. and M. Stoolmiller (1999), "Co-occurrence of Conduct Problems and Depressive

Symptoms in Early Adolescent Boys: III. Prediction to Young-adult Adjustment," *Development and Psychopathology*, 11, 59-84.
Cattell, R. B. (1973), *Personality and Mood by Questionnaire*, San Francisco: Jossey-Bass.
_____ (1978), *The Scientific use of Factor Analysis in Behavioral and Life Sciences*, New Your: Plenum.
Cronbach, Lee J. (1951), "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika*, 16 (September), 297-334.
De Ruyter, K. And M. Wetzels (2000), "With a Little Help from My Fans - Extending Models of Pro-social Behaviour to Explain Supporters' Intentions to Buy Soccer Club Shares," *Journal of Economic Psychology*, 21 (4), 387-409.
DeVellis, Robert F. (1991), *Scale Development: Theory and Applications*, Newbury Park, CA: SAGE Publications.
Dillon, William R. and Matthew Goldstein (1984), *Multivariate Analysis: Methods and Applications*, New York: Wiley.
Dwyer, F. Robert and Sejo Oh (1987), "Output Sector Munificence Effects on the Internal Political Economy of Marketing Channels," *Journal of Marketing Research*, 24 (November), 347-358.
Featherman, M. S. and P. A. Pavlou (2003), "Predicting E-services Adoption: a Perceived Risk Facets Perspective," *International Journal of Human-computer Studies*, 59, 451-474.
Fornell, Claes and David F. Larker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.
Fornell, Claes (1983), "Issues in the Application of Covariant Structure Analysis: A Comment," *Journal of Consumer Research*, 9 (March), 443-47.
Fullerton, G. and S. Taylor (2002), "Mediating, Interactive, and Non-linear Effects in Service Quality and Satisfaction with Services Research," *Canadian Journal of Administrative Sciences- Revue Canadienne Des Sciences De L Administration,* 19 (2), 124-136.
Gerbing, David W. and James C. Anderson (1984), "On the Meaning of Within-Factor Correlated Measurement Errors," *Journal of Consumer Research*, 11 (June), 572-580.
_____ (1988), "An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment," *Journal of Marketing Research*, 25 (May), 186-92.
_____ (1993), "Monte Carlo Evaluations of Goodness-of-Fit Indices for Structural Equation Models," in *Testing Structural Equation Models*, K. A. Bollen and J. S. Long, eds., Newbury Park, CA: SAGE Publications.
Gerbing, David W., Janet G. Hamilton and Elizabeth B. Freeman (1994), "A Large-scale Second-order Structural Equation Model of the Influence of Management Participation on Organizational Planning Benefits," *Journal of Management*, 20, 859-885.
Gustafsson, J. E. (1997), "Measurement Characteristics of the IEA Reading Literacy Scales for 9- and 10-year-olds at Country and Individual Levels," *Journal of Educational Measurement*," 34, 233-251.
Harris, E. G., J. C. Mowen and T. J. Brown (2005), "Re-examining Salesperson Goal Orientations: Personality Influencers, Customer Orientation, and Work Satisfaction," *Journal of the Academy of Marketing Science*, 33 (1), 19-35.
Hayduk, Leslie A. (1987), *Structural Equation Modeling with LISREL: Essential and Advances*, Baltimore, MD: Johns Hopkins Press.
Heeler, Roger M. and Michael L. Ray (1972), "Measure Validation in Marketing," *Journal of Marketing Research*, 9 (November), 361-70.
Huang, L. J., M. T. Lu, B. K. Wong (2003), "The Impact of Power Distance on Email Acceptance:

Evidence from the PRC," *Journal of Computer Information Systems,* 44 (1), 93-101.Johnson, J. L. And U. N. Umesh (2002), "The Interplay of Task Allocation Patterns and Governance Mechanisms in Industrial Distribution Channels," *Industrial Marketing Management,* 31 (8), 665-678.

Hunter, John Edward and David W. Gerbing (1982), "Unidimensional Measurement, Second-Order Factor Analysis and Causal Models," in *Research in Organizational Behavior*, Vol. IV, Barry M. Staw and L. L. Cummings eds., Greenwich CT: JAI Press, 267-320.

Iglesias, V. (2004), "Preconceptions about Service - How Much Do They Influence Quality Evaluations?" *Journal of Service Research*, 7 (1) 90-103.

Jaccard, James, Robert Turrisi and Choi K. Wan (1990), *Interaction Effects in Multiple Regression*, Newbury Park, CA: SAGE Publications.

Jaccard, James and C. K. Wan (1995), "Measurement Error in the Analysis of Interaction Effects Between Continuous Predictors Using Multiple Regression: Multiple Indicator and Structural Equation Approaches," *Psychological Bulletin*, 117 (2), 348-357.

Jensen, R. and G. Szulanski (2004), "Stickiness and the Adaptation of Organizational Practices in Cross-border Knowledge Transfers," *Journal of International Business Studies*, 35 (6), 508-523.

Jöreskog, Karl G. (1970), "A General Method for Analysis of Covariance Structures," *Biometrika*, 57, 239-251.

_____ (1971a), "Simultaneous Factor Analysis in Several Populations, *Psychometrika*, 57, 409-426.

_____ (1971b), "Statistical Analysis of Sets of Congeneric Tests," *Psychometrika*, 36, 109-133.

_____ (1993), "Testing Structural Equation Models," In *Testing Structural Equation Models*, Kenneth A. Bollen and J. Scott Long eds., Newbury Park, CA: SAGE Publications.

_____ and Dag Sörbom (1996), *Lisrel 8 User's Reference Guide*, Chicago: Scientific Software International, Inc.

Jöreskog, Karl G. and F. Yang (1996), "Nonlinear Structural Equation Models: The Kenny and Judd Model with Interaction Effects," *Advances in Structural Equation Modeling Techniques*, G.A. Marcoulides, R.E. Schumacker, eds., Hillsdale, NJ: LEA.

Kendall, M.G. and A. Stuart (1958), *The Advanced Theory of Statistics*, Vol. 1, London: Charles Griffith.

Kenny, David and Charles M. Judd (1984), "Estimating the Nonlinear and Interactive Effects of Latent Variables," *Psychological Bulletin*, 96 (July), 201-10.

Kim, Jae-On and Charles W. Mueller (1978), *Factor Analysis: Statistical Methods and Practical Issues*, Beverly Hills, CA: SAGE.

Kuklinski, M. R. and R. S. Weinstein (2001), "Classroom and Developmental Differences in a Path Model of Teacher Expectancy Effects," *Child Development*, 72, 1554-1578.

Lam, S. P. (1999), "Predicting Intentions to Conserve Water from the Theory of Planned Behavior, Perceived Moral Obligation, and Perceived Water Right," *Journal of Applied Social Psychology*, 29, 1058-1071.

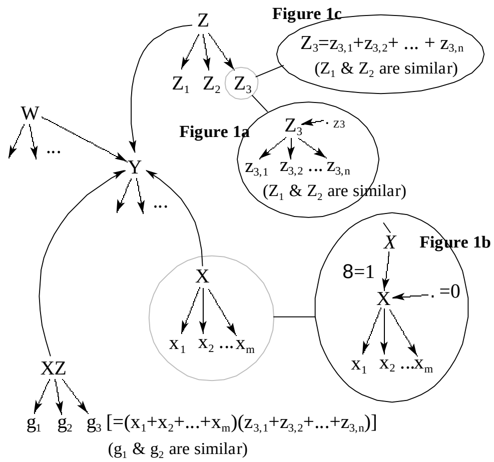
Lee, D. And G. Ganesh (1999), "Effects of Partitioned Country Image in the Context of Brand Image and Familiarity - a Categorization Theory Perspective," *International Marketing Review,* 16 (1), 18-39.

Lusch, R. F. And J. R. Brown (1996), "Interdependency, Contracting, and Relational Behavior in Marketing Channels," *Journal of Marketing*, 60 (4), 19-38.

Masterson, S. S. (2001), "A Trickle-down Model of Organizational Justice: Relating Employees' and Customers' Perceptions of and Reactions to Fairness," *Journal of Applied Psychology*, 86, 594-604.

Matsuno, K., J. T. Mentzer and A. Ozsomer (2002), "The Effects of Entrepreneurial Proclivity and Market Orientation on Business Performance," *Journal of Marketing*, 66, 18-32.
Nunnally, Jum C. (1978), *Psychometric Theory*, 2nd Ed., New York: McGraw-Hill.
Ozsomer, A. and B. L. Simonin (2004), "Marketing Program Standardization: a Cross-country Exploration," *International Journal of Research in Marketing*, 21 (4), 397-419.
Peter, J. Paul (1981), "Construct Validity: A Review of Basic Issues and Marketing Practices," *Journal of Marketing Research*, 18 (May), 133-45.
Ping, R. (1994), "Does Satisfaction Moderate the Association between Alternative Attractiveness and Exit Intention in a Marketing Channel?" *Journal of the Academy of Marketing Science*, 22, 364-371.
_____ (1995), "A Parsimonious Estimating Technique for Interaction and Quadratic Latent Variables," *The Journal of Marketing Research*, 32 (August), 336-347.
_____ (1997), "Voice in Business-to-Business Relationships: Cost-of-Exit and Demographic Antecedents, " *Journal of Retailing*, 73, 261-281.
_____ (1998a), "EQS and LISREL Examples Using Survey Data," in *Interactions and Nonlinear Effects in Structural Equation Modeling*, R. E. Schumacker and G. A. Marcoulides, eds., Mahwah, NJ:Erlbaum.
_____ (1998b), "Some Suggestions for Validating Measures Involving Unobserved Variables and Survey Data," *1998 Winter American Marketing Association Educators Conference*, Chicago: AMA.
_____ (1999), "Unexplored Antecedents of Exiting in a Marketing Channel," *Journal of Retailing*, 75 (2), 218-241.
_____ (2004), "On Assuring Valid Measures for Theoretica Models Using Survey Data," *Journal of Business Research*, 57 (2), 125-141.
Porter, T. W. And B. S. Lilly (1996), "The Effects of Conflict, Trust, and Task Commitment on Project Team Performance International," *Journal of Conflict Management,* 7 (4), 361-376.
Rindskopf, David and Tedd Rose (1988), "Some Theory and Applications of Confirmatory Second-order Factor Analysis," *Multivariate Behavioral Research*, 23 (January), 51-67.
Singh, J. (1998), "Striking a Balance in Boundary-spanning Positions: an Investigation of Some Unconventional Influences of Role Stressors and Job Characteristics on Job Outcomes of Salespeople," *Journal of Marketing*, 62 (3), 69-86.
Steiger, J.H. (1990), "Structural Model Evaluation and Modification: An Interval Estimation Approach," *Multivariate Behavioral Research*, 25, 173-180.
Taylor, S. A. (1997), "Assessing Regression-based Importance Weights for Quality Perceptions and Satisfaction Judgements in the Presence of Higher Order And/or Interaction Effects," *Journal of Retailing*, 73 (1), 135-159.
Thurstone, L. L. (1947), "Multiple-Factor Analysis," Psychological Review, 38, 406-427.
Torsheim, T., L. E. Aaroe and B. Wold (2001), "Sense of Coherence and School-related Stress as Predictors of Subjective Health Complaints in Early Adolescence: Interactive, Indirect or Direct Relationships?" *Social Science and Medicine*, 53, 603-614.
Weeks, David G. (1980), "A Second-Order Longtudinal Model of Ability Structure," *Multivariate Behavioral Research*, 15 (July), 353-365.
Werts, C.E., R.L. Linn and K.G. Jöreskog (1974), "Intraclass Reliability Estimates: Testing Structural Assumptions," *Educational and Psychological Measurement*, 34, 25-33.
Zhou, K. Z., C. K. Yim and D. K. Tse (2005), "The Effects of Strategic Orientations on Technology- and Market-based Breakthrough Innovations," *Journal of Marketing*, 69 (2), 42-60.

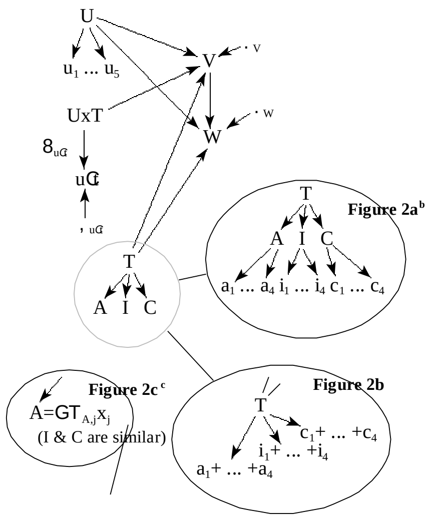
Figure 1- (Abbreviated) Structural Model with First-Order by Second-Order Interaction, XZ [a,b]



---

[a] Z is a second-order LV with first-order "indicator" LV's, Z_1, Z_2 and Z_3. W, X and Y are first-order LV's, and XZ is a second-order by first-order interaction. The circles show alternative specifications (e.g., Figure 1a shows Z_3 with observed indicators, $z_{3,1}$, $z_{3,2}$, etc., while Figure 1c shows Z_3 with a single summed indicator).
[a] X, Z, W, and XZ are correlated, and indicator error terms are uncorrelated.

Figure 2- Pedagogical Example (Abbreviated) Structural Model with First-Order by Second-Order Interaction T[a]



---

[a] T, U, and UxT were correlated, indicator error terms were uncorrelated, and the ζ's were uncorrelated.
[b] The loadings of T on A, I and C are not shown, and the measurement errors (structural disturbances) of A, I and C are also not shown.
[c] The ω's are the factor weights, and the x's are all the indicators in the model (except for UxT), u_1, u_2, ... , u_5, a_1, a_2, ... , a_4, i_1, ... , i_4, c_1, ... , c_4, v_1, ... , v_4, w_1, ... , w_4 .

Table A- Estimation Results for First-Order/Factor Scored T, Figure 2:2c (three indicator with factor scores) UxT and 2-Step Estimation (χ^2/df = 189/112, GFI = .91, AGFI = .88, CFI = .96, RMSEA = .05) (t-values are shown in parentheses)

	V	W	U	T	UxT
V			-.1625	.1897	-.3677
			(-1.323)	(1.414)	(-2.860)
W	-.1743		.1322	.1117	
	(-4.550)		(2.484)	(1.742)	

Table B- Estimation Results for First-Order/Factor Scored T, Figure 2:2c (three indicator with factor scores) UxT and Direct Estimation (χ^2/df = 187/112, GFI = .91, AGFI = .88, CFI = .96, RMSEA = .05) (t-values are shown in parentheses)

	V	W	U	T	UxT
V			-.1626	.1897	-.3674
			(-1.327)	(1.413)	(-2.861)
W	-.1743		.1321	.1118	
	(-4.550)		(2.488)	(1.741)	

Table C- Estimation Results for T as a Second-Order and UxT with 60 Kenny and Judd Indicators (χ^2/df = 28606/3593, GFI = .14, AGFI = .13, CFI = .29, RMSEA = .17) (t-values are shown in parentheses)

	V	W	U	T	UxT
V			-.089	.228	-.275
			(-1.30)	(1.68)	(-2.85)
W	-.179		.155	.119	
	(-4.63)		(5.58)	(1.81)	

Table D- Estimation Results for First-Order/Factor Scored T and UxT with 15 Kenny and Judd Indicators (χ^2/df = 3402/455, GFI = .44, AGFI = .39, CFI = .56, RMSEA = .17) (t-values are shown in parentheses)

	V	W	U	T	UxT
V			-.067	.158	-.211
			(-0.71)	(1.72)	(-2.32)
W	-.175		.158	.075	
	(-4.56)		(3.60)	(1.71)	

Table E- Estimation Results for T as a Second-Order, UxT as a Single Indicator with Summed Indicators Instead of Factor Scores, and 2-Step Estimation (χ^2/df = 479/288, GFI = .86, AGFI = .83, CFI = .96, RMSEA = .05) (t-values are shown in parentheses)

	V	W	U	T	UxT
V			-.1609	.1915	-.3817
			(-1.257)	(1.328)	(-2.775)
W	-.1739		.1328	.1107	
	(-4.537)		(2.415)	(1.617)	

Table F- Estimation Results from Table A Estimation with Var(UxT) Unconstrained (t-values are shown in parentheses)

	V	W	U	T	UxT
V			-.1609	.1925	-.3680
			(-1.295)	(1.364)	(-2.866)
W	-.1741		.1322	.1153	
	(-4.548)		(2.447)	(1.702)	

Table G- Reproduced Covariance Matrices for Alternative Specifications of T

(1) T as a Second-Order LV (see Figure 2a)

	U	T	V	W
U	0.51895			
T	0.28361	0.40877		
V	0.02618	0.04791	0.64251	
W	0.09564	0.07442	-0.10336	0.19173

(2) T as a Summed Indicator First-Order LV (see Figure 2b)

	U	T	V	W
U	0.51843			
T	0.26490	0.50086		
V	0.02614	0.05119	0.64177	
W	0.09562	0.08351	-0.10335	0.19175

(3) T as a Factor-Scored First-Order LV (see Figure 2c)

	U	T	V	W
U	0.51896			
T	0.27560	0.40562		
V	0.02618	0.04867	0.64238	
W	0.09564	0.07493	-0.10335	0.19172

Table H- Moderated V Associations with T and U Due to the Table A UxT Interaction

T-V Association Moderated by U[a]					U-V Association Moderated by T[f]				
U Level[b]	Centered U[c]	T Coefficient[d]	SE of T Coefficient[e]	t-value of T Coefficient	T Level[g]	Centered T[h]	U Coefficient[i]	SE of U Coefficient[j]	t-value of U Coefficient
1	-3.10	1.139	0.355	3.74	1.07	-2.10	0.609	0.238	2.56
2	-2.10	0.962	0.238	4.04	2	-1.17	0.267	0.142	1.88
3	-1.10	0.594	0.142	4.19	3	-0.17	-0.100	0.114	-0.88
4	-0.10	0.227	0.129	1.76	3.17	0.00	-0.163	0.123	-1.33
4.10	0.00	0.190	0.134	1.41	4	0.83	-0.468	0.197	-2.38
5	0.90	-0.141	0.215	-0.66	4.88	1.71	-0.791	0.298	-2.66
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

(Column Number)

[a] This portion of the table displays the unstandardized associations of T with V produced by the significant UxT interaction in Table A. In Columns 1-5 when the existing level of U was low in Column 1, small changes in T were positively and significantly associated with V (see Columns 3 and 5). At higher levels of U however, T was less strongly associated with V, and near the study average for U and higher, T was not associated with V.

[b] U is determined by the observed variable (indicator) with a loading of 1 on U (i.e., the indicator that provides the metric for U). The value of this indicator of U ranged from 1 (= low U) to 5 in the study.

[c] Column 1 minus the mean of U in the study, 4.10.

[d] The factored (unstandardized) coefficient of T was (.189-.367U)T with U mean centered. E.g., when U = 1 the coefficient of T was (.189-.367*(1 - 4.10) ∇ 1.329.

[e] The Standard Error of the T coefficient was:

$$\sqrt{Var(\beta_T+\beta_{UxT}U)} = \sqrt{Var(\beta_T) + U^2Var(\beta_{UxT}) + 2UCov(\beta_T,\beta_{UxT})} = \sqrt{SE(\beta_T)^2 + U^2SE(\beta_{UxT})^2 + 2UCov(\beta_T,\beta_{UxT})} ,$$

where Var and Cov denote variance and covariance, SE is standard error, and β denotes unstandardized structural coefficients from Table A.

[f] This portion of the table displays the unstandardized associations of U and V moderated by T. When T was low in Column 6, U was positively associated with V (see Column 10). As T increased, U=s association with V weakened and became non significant, then above the study average it strengthened again and was significant but negative (see columns 8 and 10).

[g] T is determined by the indicator with a loading of 1 on T (i.e., the indicator that provides the metric for T). The factor-scores for this indicator of T ranged from 1.07 (= low T) to 4.88 in the study.

[h] Column 6 minus 3.17, the mean of T in the study.

[i] The factored (unstandardized) coefficient of U was (0-.367T)U with T mean centered. E.g., when T = 1.07 the coefficient of U was -1.62-.367*(1.07-3.17) ∇ .609.

[j] The Standard Error of the U coefficient was:

$$\sqrt{Var(\beta_U+\beta_{UxT}T)} = \sqrt{Var(\beta_U) + T^2Var(\beta_{UxT}) + 2TCov(\beta_U,\beta_{UxT})} = \sqrt{SE(\beta_U)^2 + T^2SE(\beta_{UxT})^2 + 2TCov(\beta_U,\beta_{UxT})} ,$$

where Var and Cov denote variance and covariance, SE is standard error, and β denotes unstandardized structural coefficients from Table A.

ENDNOTES

1. However, Ping 1997 reported a second-order LV with a reliability of .80. Although Gerbing, Hamilton & Freeman 1994 did not report second-order reliabilities, estimating them using Equation 2 with Var(X) = 1, Var(ε) = 1 - the reported coefficient alphas, and the reported γ's, they may have been as high .90.

2. I.e., each of the latent variables X and Z is independent of its measurement errors, the measurement errors are mutually independent, and the indicators and the measurement errors are multivariate normal.

3. The variable names have been disguised and the study details have been omitted to skirt non-pedagogical matters such as the theory behind the model, hypotheses, etc.

4. The reliability of U was .943, and the reliability of T was .709. Because the correlation between T and U was .615, the reliability of UxT was .76 (= $[.615^2+.943*.709]/[.615^2+1]$).

5. As previously noted, the formula for the average variance extracted (AVE) of an interaction is not known, and experience suggests that AVE is roughly the cube of reliability. Thus, the AVE of UxT is roughly $.76^3 = .44$. If the formula for an interaction's AVE turns out to be similar to the formula for its reliability, the AVE of UxT would be .53 (= $[.615^2+.770*.474]/[.615^2+1]$).

6. Strictly speaking T (and possibly UxT) would be judged borderline unsuitable for a proper test of a theoretical model. However, impaired convergent validity does not affect the pedagogical purposes of this example.

7. In the so called 2-step version of Ping's (1995) LV interaction specification, the structural model is estimated with fixed Equation 4 and 4a values. If the resulting structural model estimates of the measurement parameters for T and U are not similar to those from the (full) measurement model corresponding to Figure 2:2c just discussed (i.e., equal in the first two decimal places) the Equation 4 and 4a loading and error variance are recomputed using the structural model parameter estimates, and the structural model is re-estimated. Experience with real world data suggests that with consistent LV's zero to one of these iterations are usually sufficient to produce exact estimates in real-world data (i.e., equal to direct estimation-- see Ping 1995).

8. UxT specified with a single summed factor score indicator, $u{\bullet}t = (u_1+u_2+...+u_5)(f_A+f_I+f_C)$ (i.e., Specification 4), and T specified with factor scores produced results similar to Table A.

9. Again, six indicators seems to be about the maximum with real-world data-- see Anderson and Gerbing, 1984; Bagozzi and Heatherton, 1994; Gerbing and Anderson, 1993. Also see Jaccard and Wan, 1995 for evidence of this difficulty with Kenny and Judd indicators.

10. UxT specified with three indicators, $u{\bullet}a = [u_1+u_2+...+u_5][\Sigma a_j]$, $u{\bullet}i = [u_1+u_2+...+u_5][\Sigma i_j]$, and $u{\bullet}c = [u_1+u_2+...+u_5][\Sigma c_j]$ (i.e., Specification 1), and T specified as a second-order construct produced results similar to Table E.

NOTES ON ESTIMATING CUBICS
AND OTHER "POWERED" LATENT VARIABLES

Robert A. Ping, Jr., Assoc. Prof.
Department of Marketing
266 Rike Hall
Wright State University
Dayton, OH 45435
937-775-3047
rping@wright.edu

NOTES ON ESTIMATING CUBICS AND OTHER "POWERED" LATENT VARIABLES

(An earlier version of Ping 2007, *Am. Mktng. Assoc. (Winter) Educators' Conf. Proc.*) (Revised Apr. 2007).

ABSTRACT

There is little guidance for specifying and estimating cubics (e.g., X^3 or XXX) in structural equation models. The paper explores these latent variables, and suggests a specification for them. A pedagogical example of their estimation is also provided. Along the way, the "factored coefficients" created by a significant cubic (e.g., the factored coefficient of X in the expression $\beta_1X + \beta_2XZ + \beta_3XXX = (\beta_1 + \beta_2X + \beta_3XX)X$ is $\beta_1 + \beta_2X + \beta_3XX$) are discussed, and standard errors of these factored coefficients are derived. The reliabilities of quadratics and cubics are also derived, an approach to interpretation of these "powered variables" is discussed, and a sequential procedure for testing hypothesized "satiation" or "diminishing returns" in theoretical model tests is suggested.

Interactions in survey data such as XZ have received attention recently (see Ping 2006 for a summary). However, related non-linear or powered variables such as quadratics, XX and ZZ, and their cubic relatives, XXX and ZZZ, in

$$Y = \beta_0 + \beta_1X + \beta_2Z + \beta_3XX + \beta_4XZ + \beta_5ZZ + \beta_6XXX + \beta_7ZZZ + \zeta_Y , \qquad (1$$

where β_1 through β_7 are unstandardized "regression" or structural coefficients (also termed associations or, occasionally, effects), β_0 is an intercept, and ζ_Y is the estimation or prediction error, also termed the structural disturbance term, have received comparatively little attention. Quadratics have been proposed and investigated in other social science literatures (e.g., Bandura 1966, Homans 1974, Howard 1989, Laroche and Howard 1980, Wheaton 1985, Yerkes and Dodson 1908). In addition, authors believe interactions and quadratics are more likely than their reported occurrence in published survey research suggests (e.g., Aiken and West 1991; Blalock 1965; Cohen 1968; Cohen and Cohen 1975, 1983; Darlington 1990; Friedrich 1982; Kenny 1985; Howard 1989; Jaccard, Turrisi and Wan 1990; Lubinski and Humphreys 1990; Neter, Wasserman and Kunter 1989; Pedhazur 1982).

Cubics in survey data have received comparatively little theoretical attention. They also have yet to appear in published structural equation models, perhaps because, among other things, there is no guidance for estimating cubics. This paper sheds additional light on several powered linear latent variables (e.g., XX, XZ, ZZ, XXX, and ZZZ in Equation 1), especially cubics. Specifically, it will discuss the estimation and interpretation of latent variable interactions, quadratics and cubics, and it will propose a specification for latent variable cubics. We begin with a discussion of non-linear variables (e.g., XX, XZ, ZZZ, etc. in Equation 1) which leads to a proposed specification of cubics involving latent variables and a pedagogical example that illustrates their estimation and interpretation.

NONLINEAR ASSOCIATIONS IN SURVEY MODELS

To motivate the admittedly novel topic of cubics in theoretical model tests using survey data it is instructive to begin with a discussion of interactions and quadratics that lays the groundwork for cubics. There have been several proposals for specifying latent variable quadratics including (1) Kenny and Judd 1984; (2) Bollen 1995; (3) Jöreskog and Yang 1996; (4) Ping 1995; (5) Ping 1996a; (6) Ping 1996b; (7) Jaccard and Wan 1995; (8) Jöreskog 2000; (9) Wall and Amemiya 2001; (10) Mathieu, Tannenbaum and Salas 1992; (11) Algina and Moulder 2001; (12) Marsh, Wen and Hau 2004; (13) Klein and Moosbrugger 2000/Schermelleh-Engle, Kein and Moosbrugger 1998/Klein and Muthén 2002; and (14) Moulder and Algina 2002.

Several of these proposals have not been evaluated for possible bias and lack of efficiency (i.e., techniques 8 and 10).

Most of these techniques are based on the Kenny and Judd product indicators (x_1z_1, x_1z_2, ... x_1z_m, x_2z_1, x_2z_2, ... x_2z_m, ... x_nz_m, where n and m are the number of indicators of X and Z respectively, or x_1x_1, x_1x_2, ... x_1x_n, x_2x_2, x_2x_3, ... x_2x_n, ... x_nx_n). However, for most theoretical model testing situations where reliability and validity are important, and where "interesting" models (i.e., models with more than 3 exogenous constructs, not including interactions or quadratics), over-determined latent variables (i.e., latent variables with 4 to 6 or more indicators), and real world survey data are the rule, specifying all the Kenny and Judd product indicators produces model-to-data fit problems. The resulting specification of XZ, for example, in a single construct measurement model will not fit the data, and the full measurement and the structural models containing XZ will usually exhibit unacceptable model-to-data fit) (techniques 1 and 5).

Several proposals use subsets of the Kenny and Judd (1984) product indicators or indicator aggregation to avoid these inconsistency problems (techniques 3, 4, 5, 7, 9, 11, 12 and 14). Unfortunately, weeding the Kenny and Judd product indicators raises questions about the face or content validity of the resulting interaction or quadratic (e.g., if all the indicators of X are not present in XX, for example, is XX still the "product of X and X"?) (techniques 3, 7, 9, 11, 12 and 14). In addition, using the formula for the reliability of XX is problematic for a weeded XX. The formula for the reliability of XX is a function of (unweeded) X, and thus it assumes XX is operationally (unweeded) X times (unweeded) X). Weeded Kenny and Judd product indicators also produce interpretation problems using factored coefficients (discussed below) because XX is no longer (unweeded) X times (unweeded) X operationally. Some of these proposals do not involve Maximum Likelihood estimation, or commercially available estimation software (techniques 2, 6 and 13).

The following will rely on the Ping 1995 proposal for several reasons including that it has the fewest drawbacks. He proposed using a single indicator that is the product of sums of the first order variables' indicators. For example for X with indicators x_1 and x_2, and Z with the indicators z_1 and z_2 the single indicator of XZ would be $xz = (x_1+x_2)(z_1+z_2)$. He further suggested that under the Kenny and Judd (1984) normality assumptions,[1] the loading, λ_{xz}, and error variance, $\theta\varepsilon_{xz}$, for this single indicator are

$$\lambda_{xz} = \Lambda_X\Lambda_Z \, , \tag{2}$$

and

$$\theta\varepsilon_{xz} = \Lambda_X^2Var(X)\theta_Z + \Lambda_Z^2Var(Z)\theta_X + \theta_X\theta_Z \, , \tag{2a}$$

where $\Lambda_X = \lambda_{x1} + \lambda_{x2}$, Var indicates error disattenuated variance, $\theta_X = Var(\varepsilon_{x1}) + Var(\varepsilon_{x2})$, $\Lambda_z = \lambda_{z1} + \lambda_{z2}$, $\theta_z = Var(\varepsilon_{z1}) + Var(\varepsilon_{z2})$, $\lambda_{xz} = \Lambda_X\Lambda_Z$, $\theta\varepsilon_{xz} = \Lambda_X^2Var(X)\theta_Z + \Lambda_Z^2Var(Z)\theta_X + \theta_X\theta_Z$, and λ and θ are loadings and measurement error variances.

Using simulated data sets with data conditions that were representative of that encountered in surveys (i.e., four indicators per X and Z latent variable, X and Z loadings ranging from 1 to .6, sample sizes of 100 and 300 cases, and linear latent variable reliabilities of .6 and .9), Ping's (1995) results suggested that the proposed single indicator for an interaction or quadratic (discussed later) produced unbiased and consistent coefficient estimates.

[1] The Kenny and Judd normality assumptions were that each of the latent variables X and Z is independent of their measurement errors (ε_{x1}, ε_{x2}, ε_{z1}, and ε_{z2}), the measurement error are mutually independent, the indicators x_1, x_2, z_1, and z_2 , and the measurement errors (ε_{x1}, ε_{x2}, ε_{z1}, and ε_{z2}) are multivariate normal.

While this specification can be used in several ways, it is most often used in two estimation steps. In this two-step version of the Ping (1995) technique the measurement parameters in Equations 2 and 2a (i.e., λ_{x1}, λ_{x2}, λ_{z1}, λ_{z2}, $\theta_{\varepsilon x1}$, $\theta_{\varepsilon x2}$, $\theta_{\varepsilon z1}$, $\theta_{\varepsilon z2x1}$, Var(X), and Var(Z)) are estimated in a first-orders-only measurement model (e.g., a model that excludes XZ). Then, the single indicators of the interaction and quadratic latent variables are created as products of sums of the indicators of linear latent variables for each case in the data set. Next, using the first-orders-only measurement model parameter estimates the loadings and measurement error variances for the interaction indicator (λ_{xz} and θ_{xz}) are computed using the equations above. Finally, specifying the calculated loadings and error variances λ_{xz} and θ_{xz} for the product indicator as constants, the structural model is estimated.

If the structural model estimates of the measurement parameters for X and Z (i.e., λ_{x1}, λ_{x2}, λ_{z1}, λ_{z2}, Var(ε_{x1}), Var(ε_{x2}), Var(ε_{z1}), Var(ε_{z2}), Var(X), and Var(Z)) are not similar to those from the measurement model (i.e., equal in the first two decimal places) the loadings and error variances of the product indicators can be recomputed using the structural model estimates of their measurement parameters. Our experience is that zero to two of these iterations are sufficient to produce exact estimates (i.e., equal to direct estimates-- see Ping 1995).

QUADRATICS AND CUBICS

In some ways the graphs of (observed) quadratics and cubics are similar. A quadratic such as XX, for example, has a graph of XX versus X that is shaped like a portion of the capital letter U. This letter U opens upwards when β_3 is positive, and downward when β_3 is negative. In general the graph of a cubic such as XXX is shaped like a portion of a stylized capital letter N (i.e., with rounded vertices) when β_6 is positive. This N is backward (again with rounded vertices) when β_6 is negative. When the range of X is suitably restricted, the graph of a cubic can be a "more-V-like" quadratic (i.e., when one of the outside legs of the N is not present in the graph). Thus, in theory tests involving survey data where the range of X, for example, is frequently restricted (e.g., from 1 to 5), XX and XXX could have similarly shaped graphs (e.g., one side of a U or an N), and in effect the estimation issue would be which one (i.e., XX or XXX), if any, provides a better representation (in theory testing terms--e.g., is significant) of hypothesized "satiation" or "diminishing returns," for example.

Ping (1995) in effect proposed that a quadratic could be specified the same way as an interaction: as the product of the sum of indicators. Thus for $xx = (x_1+x_2)(x_1+x_2)$ as the single indicator of XX, for example, the loading, λ_{xx}, and a measurement error variance, $\theta\varepsilon_{xx}$ of xx is

$$\lambda_{xx} = \Lambda_X\Lambda_X \,, \qquad (3$$

and

$$\theta\varepsilon_{xx} = 4\Lambda_X^2\mathrm{Var}(X)\theta_X + 2\theta_X^2 \,, \qquad (3a$$

where $\Lambda_X = \Sigma\lambda_{xi}$, λ_{xi} is the loading of x_i on X, Var(X) is the error disattenuated (measurement model) variance of X, $\Theta_X = \Sigma\mathrm{Var}(\varepsilon_{xi})$, and $\mathrm{Var}(\varepsilon_{xi})$ is the measurement error variance of x_i. The indicators x_i are usually zero- or mean-centered by subtracting the mean of x_i from x_i in each case, and xx becomes

$$x_cx_c = [\Sigma(x_i^u - \mathbf{M}(x_i^u)]\ [\Sigma(x_i^u - \mathbf{M}(x_i^u)] \,,$$

where x_i^u is an uncentered indicator, **M** denotes a mean, and Σ is a sum taken before any multiplication. Specifying XX with this indicator reduces the high correlation or nonessential ill-conditioning (Marquardt, 1980; see Aiken and West, 1991) of (mean centered) X and XX that produces inefficient or unstable structural coefficient estimates that vary widely across studies.

Using simulated data sets with data conditions that were again representative of those encountered in surveys, Ping's (1995) results suggested that the proposed single indicator for a quadratic produced unbiased and consistent coefficient estimates for a latent variable quadratics.

As with an interaction, this approach could be used in two ways, but the two step procedure described earlier is the least tedious.

A cubic might be specified in a manner similar to a quadratic, as the product of the sum of indicators such as $xxx = (\Sigma x_i)^3$. However, mean centering x_i does not reduce the high correlation between X and XXX (Dunlap and Kemery, 1988 Marquardt, 1980; see Aiken and West, 1991), and inefficient structural coefficient estimates usually obtain. An alternative specification that avoids this difficulty is the indicator

$$x_c(xx)_c = [\Sigma x_i^u - \mathbf{M}(\Sigma x_i^u)][(\Sigma x_i^u)(\Sigma x_i^u) - \mathbf{M}((\Sigma x_i^u)(\Sigma x_i^u))] \,, \qquad (\cancel{3,\ 3c,}\ 3b$$

where Σx_i^u is the sum of uncentered x_i's. The indicator $x_c(xx)_c$ is thus the product of mean-centered x_i's and mean-centered x_ix_j's, and thus $x_c(xx)_c$ could be used as a single indicator of $X_c(X_uX_u)_c$ (i.e., with $x_c(xx)_c$ =

$\lambda_{Xc(XuXu)c}X_c(X_uX_u)_c + \varepsilon_{Xc(XuXu)c})$.

The loading and error variance of $x_c(xx)_c$ is derived next. Under the above Kenny and Judd (1984) assumptions the variance of the product of $x = (\Sigma x_i) = [\Sigma(\lambda_{xi}X + \varepsilon_{xi})]$ and $w = (\Sigma w_i) = [\Sigma(\lambda_{wj}W + \varepsilon_{wj})$ (where λ_{wj} and ε_{wj} are the loading and measurement error, respectively, of w_j on its latent variable W, and j = 1 to m, where m is the number if indicators of W), is

$$\mathrm{Var}(xw) = (\Lambda_X\Lambda_W)^2\mathrm{Var}(XW) + \Lambda_X^2\mathrm{Var}(X)\Theta_W + \Lambda_W^2\mathrm{Var}(W)\Theta_X + \Theta_X\Theta_W \ , \qquad (4a$$

(e.g., Ping 1995) where $\Lambda_W = \Sigma\lambda_{wj}$, λ_{wj} is the loading of w_j on W, Var(XW) is the error disattenuated variance of XW, $\Theta_W = \Sigma\theta\varepsilon_{wj}$, Var(W) is the error disattenuated variance of W, and $\theta\varepsilon_{wj}$ is the measurement error variance of w_j. This would provide a specification of XW using the indicator xw with the calculated loading $\Lambda_{XW} = \Lambda_X\Lambda_W$ and the calculated measurement error variance $\Theta_{XW} = \Lambda_X^2\mathrm{Var}(X)\Theta_W + \Lambda_W^2\mathrm{Var}(W)\Theta_X + \Theta_X\Theta_W$, if estimates of these parameters are available (e.g., Ping, 1995).

To provide estimates of the parameters involving a specific $W = (X_uX_u)_c$,

$$\mathrm{Var}(W) = \mathrm{Var}[(X_uX_u)_c] = 4\mathbf{E}^2(X_u)\mathrm{Var}(X_u) + 2\mathrm{Var}^2(X_u)$$

(Bohrnstedt and Goldberger, 1969), where $\mathbf{E}^2(X_u)$ denotes the square of the mean of X_u, and $\mathrm{Var}^2(X_u)$ is the square of the variance of X_u (X_u is uncentered and thus has a non-zero mean). Since first moments are unaffected by measurement error, $\mathbf{E}(X_u) = \mathbf{E}(\Sigma x_i^u)$, and $\mathrm{Var}(X_u) = \mathrm{Var}(X)$. Thus, for $W = (X_uX_u)_c$, Var(W) in Equation 4a is a function of the mean of Σx_i^u and the variance of X:

$$\mathrm{Var}(W) = 4\mathbf{E}^2(\Sigma x_i^u)\mathrm{Var}(X) + 2\mathrm{Var}^2(X) \ , \qquad (4b$$

where $\mathbf{E}(\Sigma x_i^u)$ is available from SAS, SPSS, etc. and Var(X) is available from the measurement model for X.

For $(xx)_c$,

$$\begin{aligned}(xx)_c &= (\Sigma x_i^u)(\Sigma x_i^u) - \mathbf{M}((\Sigma x_i^u)(\Sigma x_i^u)) \\ &= [(\Sigma\lambda_{xiu})X_u + \Sigma\varepsilon_{xiu}][(\Sigma\lambda_{xiu})X_u + \Sigma\varepsilon_{xiu}] - \mathbf{M}((\Sigma x_i^u)(\Sigma x_i^u)) \\ &= (\lambda_{Xu}X_u + \varepsilon_{Xu})(\lambda_{Xu}X_u + \varepsilon_{Xu}) - \mathbf{M}((\Sigma x_i^u)(\Sigma x_i^u)) \\ &= \lambda_{Xu}^2X_uX_u + 2\lambda_{Xu}X_u\varepsilon_{Xu} + \varepsilon_{Xu}\varepsilon_{Xu} - \mathbf{M}((\Sigma x_i^u)(\Sigma x_i^u)),\end{aligned}$$

where $\lambda_{Xu} = \Sigma\lambda_{xiu}$ and $\varepsilon_{Xu} = \Sigma\varepsilon_{xiu}$, and with the usual assumptions that X_u is independent of measurement errors, measurement errors have zero expectations, and measurement errors are independent, and thus

$$\mathrm{Var}((xx)_c) = \mathrm{Var}(\lambda_{Xu}{}^2X_uX_u + 2\lambda_{Xu}X_u\varepsilon_{Xu} + \varepsilon_{Xu}\varepsilon_{Xu} - \mathbf{M}((\Sigma x_i{}^u)(\Sigma x_i{}^u)))$$

$$= \mathrm{Var}(\lambda_{Xu}{}^2X_uX_u + 2\lambda_{Xu}X_u\varepsilon_{Xu} + \varepsilon_{Xu}\varepsilon_{Xu})$$

$$= \mathrm{Var}(\lambda_{Xu}{}^2X_uX_u) + \mathrm{Var}(2\lambda_{Xu}X_u\varepsilon_{Xu}) + \mathrm{Var}(\varepsilon_{Xu}\varepsilon_{Xu}) + 2\mathrm{Cov}(\lambda_{Xu}{}^2X_uX_u, 2\lambda_{Xu}X_u\varepsilon_{Xu}) + 2\mathrm{Cov}(\lambda_{Xu}{}^2X_uX_u, \varepsilon_{Xu}\varepsilon_{Xu}) + 2\mathrm{Cov}(2\lambda_{Xu}X_u\varepsilon_{Xu}, \varepsilon_{Xu}\varepsilon_{Xu})$$

$$= (\lambda_{Xu}{}^2)^2\mathrm{Var}(X_uX_u) + 4\lambda_{Xu}{}^2\mathrm{Var}(X_u\varepsilon_{Xu}) + \mathrm{Var}(\varepsilon_{Xu}\varepsilon_{Xu})$$

$$= (\lambda_{Xu}{}^2)^2\mathrm{Var}(X_uX_u) + 4\lambda_{Xu}{}^2[\mathbf{E}^2(X_u)\mathrm{Var}(\varepsilon_{Xu}) + \mathrm{Var}(X_u)\mathrm{Var}(\varepsilon_{Xu})] + 2\mathrm{Var}^2(\varepsilon_{Xu}) ,$$

because the covariances involving measurement errors are zero, and the expansions of $\mathrm{Var}(X_u\varepsilon_{Xu})$ and $\mathrm{Var}(\varepsilon_{Xu}\varepsilon_{Xu})$ contain zero $\mathbf{E}(\varepsilon_{Xu})$ terms.

As a result, in Equation 4a for $W = X_uX_u$

$$\Lambda_W = \lambda_{Xu}{}^2 = (\Sigma\lambda_{xi})^2 \qquad (4c$$

and

$$\Theta_W = 4\lambda_{Xu}{}^2[\mathbf{E}^2(X_u)\mathrm{Var}(\varepsilon_{Xu}) + \mathrm{Var}(X_u)\mathrm{Var}(\varepsilon_{Xu})] + 2\mathrm{Var}^2(\varepsilon_{Xu})$$

$$= 4(\Sigma\lambda_{xiu})^2[\mathbf{E}^2(X_u)\mathrm{Var}(\Sigma\varepsilon_{xiu}) + \mathrm{Var}(X_u)\mathrm{Var}(\Sigma\varepsilon_{xiu})] + 2\mathrm{Var}^2(\Sigma\varepsilon_{xiu}) .$$

Since $(\Sigma\lambda_{xiu}) = (\Sigma\lambda_{xi})$, $\mathbf{E}(X_u) = \mathbf{E}(\Sigma x_i{}^u)$, $\mathrm{Var}(\Sigma\varepsilon_{xiu}) = \mathrm{Var}(\Sigma\varepsilon_{xi}) = \Sigma\mathrm{Var}(\varepsilon_{xi})$, and $\mathrm{Var}(X_u) = \mathrm{Var}(X)$,

$$\Theta_W = 4(\Sigma\lambda_{xi})^2[\mathbf{E}^2(\Sigma x_i{}^u)\Sigma\mathrm{Var}(\varepsilon_{xi}) + \mathrm{Var}(X)\Sigma\mathrm{Var}(\varepsilon_{xi})] + 2(\Sigma\mathrm{Var}(\varepsilon_{xi}))^2 \qquad (4d$$

and the parameters in Equation 4a are all parameters involving X. Specifically, the loading of XW, Λ_{XW}, is $\Lambda_X\Lambda_W$ and the measurement error variance of XW, Θ_{XW}, is $\Lambda_X{}^2\mathrm{Var}(X)\Theta_W + \Lambda_W{}^2\mathrm{Var}(W)\Theta_X + \Theta_X\Theta_W$, where $\Lambda_X = \Sigma\lambda_{xi}$, i = 1 to n and n is the number of indicators of X, λ_{xi} are the loadings of λ_{xi} on X, $\Lambda_W = (\Sigma\lambda_{xi})^2$**,** Var(X) is (error disattenuated) variance available in the measurement model for X, $\Theta_W = 4(\Sigma\lambda_{xi})^2[\mathbf{E}^2(\Sigma x_i{}^u)\Theta_X + \mathrm{Var}(X)\Theta_X] + 2\Theta_X{}^2$, $\mathbf{E}^2(\Sigma x_i{}^u)$ is the square of the mean of the sum of the uncentered x_i (available from SAS, SPSS, etc.), $\Theta_X = \Sigma\mathrm{Var}(\varepsilon_{xi})$, $\mathrm{Var}(\varepsilon_{xi})$ are the measurement error variances of x_i, $\Theta_X{}^2$ is the square of Θ_X, and $\mathrm{Var}(W) = 4\mathbf{E}^2(\Sigma x_i{}^u)\mathrm{Var}(X) + 2\mathrm{Var}^2(X)$.[2]

---

[2] Using Equation 4a, this simplifies to the loading, Λ_{XXX}, of the cubic XXX

$$\Lambda_{XXX} = (\Lambda_X)^3, \qquad (4e$$

and the error variance Θ_{XW} of the cubic XXX is

$$\Theta_{XW} = \Lambda_X{}^2\mathrm{Var}(X)\Theta_W + \Lambda_W{}^2\mathrm{Var}(W)\Theta_X + \Theta_X\Theta_W .$$

Significant higher orders such as XX, XZ and XXX create factor coefficients of X, for example. We discuss these next.

FACTORED COEFFICIENTS AND THEIR STANDARD ERRORS

Equation 1 can be factored to produce a coefficient of Z due to the interaction XZ, i.e.,

$$Y = \beta_1 X + (\beta_2 + \beta_4 X)Z + \beta_3 XX + \beta_5 ZZ + \beta_6 XXX + \beta_7 ZZZ + \zeta_Y . \qquad (5$$

Similarly Equation 1 can be refactored to produce a coefficient of X due to the interaction XZ (i.e., $[\beta_1 + \beta_4 Z]$ in $[\beta_1 + \beta_4 Z]X$). These factored coefficients are important to understanding interactions in survey data, and later they will help shed more light on cubics. When the XZ interaction in Equation 1 is significant (i.e., β_4 is

---

Substituting for Θ_W and Var(W) using Equations 4b and 4d respectively and simplifying

$$\Theta_{XW} = \Lambda_X^2 Var(X)[4(\Sigma\lambda_{xi})^2[\mathbf{E}^2(\Sigma x_i^u)Var(\Sigma\varepsilon_{xi}) + Var(X)Var(\Sigma\varepsilon_{xi})] + 2Var^2(\Sigma\varepsilon_{xi})]$$
$$+ \Lambda_W^2 [4\mathbf{E}^2(\Sigma x_i^u)Var(X) + 2Var^2(X)]\Theta_X$$
$$+ \Theta_X[4(\Sigma\lambda_{xi})^2[\mathbf{E}^2(\Sigma x_i^u)Var(\Sigma\varepsilon_{xi}) + Var(X)Var(\Sigma\varepsilon_{xi})] \qquad + 2Var^2(\Sigma\varepsilon_{xi})]$$
$$= \Lambda_X^2 Var(X)[4(\Sigma\lambda_{xi})^2\mathbf{E}^2(\Sigma x_i^u)Var(\Sigma\varepsilon_{xi}) + 4(\Sigma\lambda_{xi})^2 Var(X)Var(\Sigma\varepsilon_{xi}) + 2Var^2(\Sigma\varepsilon_{xi})]$$
$$+ \Lambda_W^2\, 4\mathbf{E}^2(\Sigma x_i^u)Var(X)\,\Theta_X + \Lambda_W^2 2Var^2(X)\Theta_X$$
$$+ \Theta_X 4(\Sigma\lambda_{xi})^2\mathbf{E}^2(\Sigma x_i^u)Var(\Sigma\varepsilon_{xi}) + 4(\Sigma\lambda_{xi})^2 Var(X)Var(\Sigma\varepsilon_{xi})\,\Theta_X + 2Var^2(\Sigma\varepsilon_{xi})\,\Theta_X$$
$$= \Lambda_X^2 Var(X)4(\Sigma\lambda_{xi})^2\mathbf{E}^2(\Sigma x_i^u)Var(\Sigma\varepsilon_{xi}) + \Lambda_X^2 Var(X)4(\Sigma\lambda_{xi})^2 Var(X)Var(\Sigma\varepsilon_{xi})$$
$$+ \Lambda_X^2 Var(X)2Var^2(\Sigma\varepsilon_{xi})$$
$$+ \Lambda_W^2\, 4\mathbf{E}^2(\Sigma x_i^u)Var(X)\Theta_X + \Lambda_W^2 2Var^2(X)\Theta_X$$
$$+ \Theta_X 4(\Sigma\lambda_{xi})^2\mathbf{E}^2(\Sigma x_i^u)Var(\Sigma\varepsilon_{xi}) + 4(\Sigma\lambda_{xi})^2 Var(X)Var(\Sigma\varepsilon_{xi})\Theta_X + 2Var^2(\Sigma\varepsilon_{xi})\Theta_X$$
$$= 4\Lambda_X^4 Var(X)\Theta_X\mathbf{E}^2(\Sigma x_i^u) + 4\Lambda_X^{\ 4}Var(X)^2\Theta_X + 2\Lambda_X^2 Var(X)\Theta_X^2$$
$$+ 4\Lambda_X^4 Var(X)\Theta_X\mathbf{E}^2(\Sigma x_i^u) + 2\Lambda_X^4 Var^2(X)\Theta_X$$
$$+ 4\Lambda_X^2\Theta_X^2\mathbf{E}^2(\Sigma x_i^u) \qquad + 4\Lambda_X^2 Var(X)\Theta_X^2 + 2\Theta_X^{\ 3}$$
$$= 8\Lambda_X^4 Var(X)\Theta_X\mathbf{E}^2(\Sigma x_i^u) + 6\Lambda_X^4 Var(X)^2\Theta_X + 6\Lambda_X^2 Var(X)\Theta_X^2$$
$$+ 4\Lambda_X^2\Theta_X^2\mathbf{E}^2(\Sigma x_i^u) + 2\Theta_X^3 , \qquad (4f$$

which are available in a "first-orders only" measurement model containing X, and for unidimensional first-significant) the factored coefficient of Z, for example, in Equation 5 is not constant over the range of X in the study. Depending on the signs and magnitudes of β_2 and β_4, the (factored) coefficient of Z, $(\beta_2 + \beta_4 X)$, can be positive for X at one end of the range of X in the study, zero near the middle of the range of X, and negative at the other end of the range of X in the study.

The standard error of the factored coefficient of Z also varies over the range of X in the study. Determined by the square root of $Var(\beta_2 + \beta_4 X)$, where Var indicates variance, the standard error of the factored coefficient of Z is

$$[Var(\beta_2) + X^2 Var(\beta_4) + 2XCov(\beta_2,\beta_4)]^{1/2}, \qquad (6$$

where Cov indicates covariance (which is available as an output option in structural equation packages such as LISREL, EQS, etc.) and the exponents 2 and 1/2 indicate the square and the square root respectively (e.g., Jaccard, Turrisi and Wan 1990). In different words, the standard error of the factored coefficient of Z, $(\beta_2 + \beta_4 X)$, is a function of the standard errors/variances of β_2, β_4 and the value of X at which the coefficient is evaluated. Thus the factored coefficient of Z can not only vary with the values of X in the study, but it can also be significant for some X in the study but nonsignificant for other values of X in the study.

Other factored coefficients are obviously possible. For example Equation 1 could be refactored to produce a factored coefficient of Z due to the quadratic ZZ (i.e., $[\beta_2 + \beta_5 Z]$ in $[\beta_2 + \beta_5 Z]Z$), and a factored coefficient of X due to the quadratic XX (i.e., $[\beta_1 + \beta_3 X]$ in $[\beta_1 + \beta_3 X]X$). The factored coefficient of Z due to ZZ and ZZZ is $[(\beta_2 + (\beta_5 + \beta_7 Z)Z]$ in $[(\beta_2 + (\beta_5 + \beta_7 Z)Z]Z$, and the factored coefficient of X due to XX and XXX is $[(\beta_1 + (\beta_3 + \beta_6 X)X)]$ in $[(\beta_1 + (\beta_3 + \beta_6 X)X)]X$. Additional factorizations are also possible. For example the factored coefficient of Z in Equation 1 is shown in

$$[(\beta_2 + \beta_4 X + (\beta_5 + \beta_7 Z)Z)]Z, \qquad (7$$

and the factored coefficient of X is shown in $[(\beta_1 + \beta_4 Z + (\beta_3 + \beta_6 X)X)]X$ In addition, each of these factored coefficients has a nonconstant standard error that is a function of the (constant) standard errors of the coefficients that comprise it (i.e., the β's) and values of the variables (e.g., X, Z, etc.) that also comprise it. For example the standard error of the Equation 7 factored coefficient of Z, $(\beta_2 + \beta_4 X + (\beta_5 + \beta_7 Z)Z)$, is

---

order LV's, can be reestimated using the structural model (i.e., using "2-steps, see Ping 1995).

$$[Var(\beta_2 + \beta_4X + (\beta_5+ \beta_7Z)Z)]^{1/2} = [Var(\beta_2 + \beta_4X + \beta_5Z+ \beta_7ZZ)]^{1/2}$$

$$= [Var(\beta_2) + Var(\beta_4X + \beta_5Z+ \beta_7ZZ) + 2Cov(\beta_2,\beta_4X + \beta_5Z+ \beta_7ZZ)]^{1/2}$$

$$= [Var(\beta_2) + X^2Var(\beta_4) + Var(\beta_5Z+ \beta_7ZZ) + 2Cov(\beta_4X,\beta_5Z+ \beta_7ZZ)$$

$$+ 2XCov(\beta_2,\beta_4) + 2ZCov(\beta_2,\beta_5) + 2ZZCov(\beta_2,\beta_7)]^{1/2}$$

$$= [Var(\beta_2) + X^2Var(\beta_4) + Z^2Var(\beta_5) + (ZZ)^2Var(\beta_7) + 2ZZZCov(\beta_5,\beta_7)$$

$$+ 2XZCov(\beta_4,\beta_5) + 2XZZCov(\beta_4,\beta_7) + 2XCov(\beta_2,\beta_4)$$

$$+ 2ZCov(\beta_2,\beta_5) + 2ZZCov(\beta_2,\beta_7)]^{1/2} , \qquad (8$$

where the exponent 1/2 indicates the square root. Thus the standard error of this factored coefficient of Z, $(\beta_2 + \beta_4X + (\beta_5+ \beta_7Z)Z)$, varies with X and Z, and the factored coefficient of Z could be negative and significant for some combination of X's and Z's, it could be nonsignificant for other combinations of X's and Z's, and the coefficient of Z could be positive and significant for still other combinations of X's and Z's in the study.

INTERPRETATIONS

While Equation 1 may appear to be more complicated than a more traditional survey data model that ignores non-linear or powered terms, its interpretation is simplified by the use of factored coefficients. We will illustrate this by interpreting an interaction, a quadratic and two factored coefficients involving a cubic.

Interactions Suppose in a study that only X, Z and the interaction XZ were significant in Equation 1 (i.e., β_3, β_5, β_6 and β_7 were nonsignificant). In addition, assume β_2 in the resulting factored coefficient of Z, $(\beta_2 + \beta_4X)$, was -.191 and β_4 was .142. Table B (from the pedagogical example, discussed later) shows the factored coefficient of Z at several different levels of X, where X is the latent variable A and Z is the latent variable I, and $(\beta_2 + \beta_4X)Z$ = (-.191+.142A)I. When A was low in this study, for example equal to 1, the coefficient of the variable I was -.191+.142*(1 - 2.54) ∇ -.411 (A had a mean of 2.54, and it was mean centered) (see Column 2 of Table B) (the Table B calculation of Column 3 involved β_2 and β_4 with more than 3 significant decimal digits). When A was higher, for example A = 2.54 the coefficient of the variable I was -.191+.142*(2.54 - 2.54) = 0, and when A was 5 the coefficient of the variable I was -.191+.142*(5 - 2.54) = .1583 (again the table value used β_2 and β_4 with more than 3 significant decimal digits). A summary statement for the variable coefficient results for the I-Y association shown in the left hand portion of Table B would be that when A was low in the study the variable I was negatively associated with Y. However, for A above study average the variable I was positively associated with Y.

In addition, because the standard error of the factored coefficient of I also varied with A, the variable I was significantly associated with Y for A's at the low end of the range of A's in the study. However, the variable I was nonsignificantly associated with Y for A's above the study average. Overall, when A was low in the study, the variable I was significantly and negatively associated with Y. However, for A above study average the variable I was not associated with Y.

Because β_4 was significant, Equation 1 could be refactored to produce a factored coefficient of X, (β_1 + β_4Z). The interpretation of the factored coefficient (β_1 + β_4I)A is also shown in Table B. To summarize this portion of Table B, when the variable I was low in the study, the A-Y association was not significant. However, for I at or above the study average the A-Y association was positive and significant.

Quadratics Turning a quadratic, in a different study suppose in Equation 1 that only Z and ZZ were significant (i.e., β_1, β_3, β_4, β_6 and β_7 are nonsignificant). In addition assume β_2 in the resulting factored coefficient of Z, (β_2 + β_5Z), was -.191 and β_5 was -.092. Table C (also from the pedagogical example) shows the factored coefficient of Z at several different levels of Z, where Z is the latent variable I, and (β_2 + β_5Z)Z = (-.191 -.092I)I. When the level of I was high in the study (e.g., I = 5) the coefficient of I was (-.191 - .092(5 - 3.8)) ∇ -.303 (the variable I had a mean of 3.8, and it was mean centered) (see Column 3 of Table C-- the calculation of Column 3 involved β_2 and β_4 with more than 3 significant decimal digits). Thus when the level of I was 5 in the study small changes in that level of I were negatively associated with Y. However when I was lower in the study, small changes in I were less strongly associated with Y, and when the level of I was below 2 in the study small changes in the variable I were positively associated with Y.

Because the standard error of the factored coefficient of I also varied with I, the variable I was significant for I's at the high end of the range of I's, and I was nonsignificant for I's below the study average (see Column 5 of Table C).

CUBICS

Turning to cubics, in a different study suppose in Equation 1 that only X and XXX were significant (i.e., β_2, β_3,

β_4, β_5 and β_7 are nonsignificant). In addition assume β_1 in the resulting factored coefficient of X, $(\beta_1 + \beta_6XX)X$, was .191 and β_6 was .015. Table D (from the pedagogical example) shows the factored coefficient of X at several different levels of XX, where X is the latent variable A, and $(\beta_2 + \beta_6XX)X = (.191+.015AA)A$. When the level of A was low in the study (e.g., I = 1) the coefficient of A was $(.191 + .015(5 - 2.54)^2) \nabla .229$ (A had a mean of 2.54, and it was mean centered) (see Column 4 of Table C-- the calculation of Column 4 involved β_2 and β_6 with more than 3 significant decimal digits). Thus when the level of A was 1 in the study small changes in that level of A were positively associated with Y. However when the level of A increased in the study, the association between A and Y declined, then for A above the study average it increased again.

Because the standard error of the factored coefficient of A also varied slightly with A, the variable A was more significant at the extremes of the range of A in the study.

COMBINATIONS

Finally, suppose X, XZ and the cubic XXX were significant in a study. Further suppose β_1, β_4 and β_6 were observed to be the same as they were in the earlier studies, so that the factored coefficient of X was (.191+.142Z+.015XX)X. Table E (from the example) provides a slightly different tabulation for the factored coefficient of X at several different levels of Z and XX, where X is the latent variable A, and Z is the latent variable I. To interpret the factored coefficient of X = I, $(\beta_1 + \beta_4 I + \beta_6 AA)$, when I and A are low in the study (e.g., = 1) the factored coefficient of A is $(.191+.142{*}I+.015AA) = (.191+.142{*}(1-3.8)+.015{*}(1-2.54)^2 \nabla -.173$ (I had a mean of 3.8 and A had a mean of 2.54, and both were mean centered) (see Column 1 row 1 of Table E-- the calculation involved β's that had more than 3 significant decimal digits). Thus when I and A were low in the study, (small) changes in A were negatively (but nonsignificantly) associated with Y (see Column 1 row 3). As the Column 1 level of I increased in the study, A's association with Y weakened, until between I equal 2 and 3 it turned positive, and it became significant for values of I just below the study average of I. For higher values of A in Columns 2 through 6 this pattern of nonsignificance for lower values of I, and significance for higher values of I was repeated.

The standard error of the factored coefficient is determined by its variance which can be derived by inspection from Equation 8:

$$[\mathrm{Var}(\beta_1 + \beta_4 Z + \beta_6 XX)]^{1/2} = [\mathrm{Var}(\beta_1 + \beta_4 Z + \beta_6 XX)]^{1/2}$$

$$= [\mathrm{Var}(\beta_1) + \mathrm{Var}(\beta_4 Z + \beta_6 XX) + 2\mathrm{Cov}(\beta_1,\beta_4 Z + \beta_6 XX)]^{1/2}$$

$$= [\mathrm{Var}(\beta_1) + Z^2\mathrm{Var}(\beta_4) + XX^2\mathrm{Var}(\beta_6) + 2Z{*}XX{*}\mathrm{Cov}(\beta_4,\beta_6)$$

$$+ 2Z\mathrm{Cov}(\beta_1,\beta_4) + 2XX\mathrm{Cov}(\beta_1,\beta_6)]^{1/2}$$

where the exponent 1/2 indicates the square root. Thus the standard error of this factored coefficient for X varies with the levels of Z and X, and X was negative and nonsignificant for some combination of Z's and X's, and X was be positive and significant for other combinations of X's and Z's in the study.

AN EXAMPLE

For pedagogical purposes a real-world data set will be reanalyzed.[3] The structural model

$$Y = b_1S + b_2A + b_3I + b_4C + b_5AI + b_6II + b_7SS + b_8AA + \zeta \qquad (9$$

was tested in response to hypotheses that postulated that S, A, I and C were associated with Y; that I moderated the A-Y association, and that there were "diminishing returns-like" behavior in S, A and I (e.g., as S increased, it's association with, or its "effect" on, Y diminished) (see Howard (1989); Jaccard, Turrisi and Wan (1990); and Kenny and Judd (1984) for accessible discussions and examples of quadratics). S, A, I, C and Y were operationalized as latent variables with multiple indicators measured with Likert scales.

Quadratics in S, A and I were specified in Equation 9 because the hypotheses did not stipulate the form of the "diminishing returns-like" relationships between Y, and S, A and I, and quadratics are more familiar (and perhaps more parsimonious) that cubics. In addition, cubics are difficult to jointly estimate with their related quadratics in real-world data because of nonessential ill-conditioning in X, XX and XXX (high intercorrelations) (Dunlap and Kemery, 1988 Marquardt, 1980; see Aiken and West, 1991) that can usually be attenuated only in pairs of X, XX and XXX (e.g., X and XX, X and XXX, XX and XXX).

Before specifying the nonlinear variables (i.e., AI, II, SS, and AA), the consistency, reliability and validity of the first-order latent variables (i.e., S, A, I, C and Y) was verified using LISREL 8 and Maximum Likelihood (ML) estimation. Consistency was attained for the first-order latent variables by estimating a single

[3] The variable names have been disguised to skirt non pedagogical issues such as the theory behind the model, etc.

construct measurement model for S, for example, and omitting the item with the largest sum of first derivatives without regard to sign.[4] The single construct measurement model with the remaining indicators of S was then reestimated, and the indicator with the resulting largest sum of first derivatives without regard to sign was omitted. This process of omitting, reestimating, and then omitting the indicator with the resulting largest sum of first derivatives without regard to sign in each reestimation was repeated until the p-value for χ^2 in the single construct measurement model for the remaining indicators of S became (just) non zero. This process was repeated for the other first-order latent variables using the EXCEL template "For "weeding" a measure so it "fits the data"..." shown on the previous web page (Ping 2006a), and the resulting measures were judged to have acceptable internal consistency. These first-orders also were judged to be externally consistent (e.g., external consistency was judged to be adequate for these purposes using the model-to-data fit of the measurement model containing S, A, I, C and Y: χ^2/df/p-value/GFI/AGFI/CFI/RMSEA = 644/199/0/.894/.861/.950/.070). (Experience suggests that in real-world data the first derivative approach typically produces internally consistent latent variables that are also externally consistent.) They were also judged to be reliable (ρ_S = .95, ρ_A = .91, ρ_I = .90, ρ_C = .89 and ρ_Y = .87, where ρ denotes reliability) and valid.[5]

To specify the non-linear variable (i.e., AI, II, SS, and AA), single indicators and Equations2, 2a, 3 and 3a were used. First, the indicators of all the first-order latent variables were mean centered to reduce the nonessential ill-conditioning in A, for example, and its related second-order variables (i.e., AI, and AA). Then the mean centered indicators of A were summed and multiplied by the mean centered and summed indicators

---

[4] Omitting an item must be done with concern that the omitted item does not degrade content or face validity.

[5] Authors disagree on what comprises an adequate set of validity criteria (e.g., Bollen 1989, Campbell 1960, DeVellis 1991, Heeler and Ray 1972, Nunnally 1978, Peter 1981). Nevertheless, a minimal demonstration of validity might include content or face validity (how well a latent variable's indicators tap its conceptual definition), construct validity (its correlations with other latent variables are theoretically sound), convergent validity (e.g., its average extracted variance (Fornell and Larker 1981) is greater than .5), and discriminant validity (e.g., its correlations with other measures are less than .7) (e.g., Bollen 1989, DeVellis 1991, Fornell and Larker 1981, Nunnally 1978). The validity of the study measures was qualitatively assessed considering their reliability and their performance over this minimal set of validity criteria.

of I in each case (i.e., to form $(a_1+a_2+\ldots)(i_1+i_2+\ldots)$). Similarly, the single indicators of II, SS and AA were formed in a similar manner in each case.

Then, the consistency, reliability and validity of these second-orders were gauged. The second-orders were judged to be trivially internally consistent, and to judge their external consistency, the Equation 9 measurement model was estimated using LISREL 8 and Maximum Likelihood estimation by fixing the loadings and measurement error variances for the single indicators of AI, II, SS and AA at their Equations 2, 2a, 3 and 3a computed values. Because the resulting measurement model was judged to fit the data (χ^2/df/p-value/GFI/AGFI/CFI/RMSEA = 795/267/0/.886/.852/.939/.061), the second-orders were judged to be externally consistent.

The reliability of an interaction XZ, ρ_{XZ}, involving mean centered X and Z is

$$\rho_{XZ} = \frac{\mathrm{Corr}^2_{X,Z} + \rho_X\rho_Z}{\mathrm{Corr}^2_{X,Z} + 1} \tag{9a}$$

(Bohrnstedt and Marwell 1978, see Busemeyer and Jones 1983), where $\mathrm{Corr}^2_{X,Z}$ is the square of the correlation between X and Z, and ρ_X and ρ_Z are the reliabilities of X and Z.

Under the usual assumptions for a mean centered latent variable X,[6] the reliability of the quadratic XX, ρ_{XX}, is

$$\begin{aligned}\rho_{XX} &= \frac{\mathrm{Var}(X_TX_T)}{\mathrm{Var}(XX)} \qquad (10\\ &= \frac{2\mathrm{Var}^2(X_T)}{2\mathrm{Var}^2(X)}\\ &= (\rho_X)^2\end{aligned}$$

(Kendall and Stewart, 1958), where $(\rho_X)^2$ is the square of the reliability of X, since XX is composed of true-score variance, $\mathrm{Var}(X_TX_T)$, and error variance, and $\mathrm{Var}(X_T) = \rho_X\mathrm{Var}(X)$.

[6] For normally distributed $X = X_T + e_X$, where X_T denotes the true-score and e_X is measurement error, the expectation of the measurement error is assumed to be zero, and X and X_T are independent of the measurement error.

Using these results, the second-order latent variables were judged to be reliable (ρ_{AI} = .83, ρ_{II} = .81, ρ_{SS} = .87, and ρ_{AA} = .84) and valid (e.g., their AVE's were above .5).[7]

Next the measurement model for a reduced Equation 9 that excluded AI, II, SS and AA that was used to gauge the external consistency of the first-order latent variables was examined. Using its loadings and measurement error estimates for S, A, I, Equations 2, 2a, 3 and 3a were used to calculate the loadings and measurement error variances for AI, II, SS and AA.

Then the Equation 9 measurement model was estimated using LISREL and Maximum Likelihood estimation by fixing the loadings and measurement error variances for AI, II, SS and AA at their Equations 2, 2a, 3 and 3a computed values. This measurement model was judged to fit the data, suggesting AI, II, SS and AA were externally consistent. (They were trivially internally consistent, and experience suggests that for consistent first order latent variables, their interactions and quadratics specified using the Ping 1995 approach will be externally consistent in real world data.) The resulting loadings and measurement error variances for S, A, I, C and Y in the structural model were judged sufficiently similar to those from the measurement model for S, A, I, C and Y, and that a second measurement model estimation was not necessary.

Then the Equation 9 structural model was estimated using LISREL and Maximum Likelihood estimation. This structural model was judged to fit the data (χ^2/df/p-value/GFI/AGFI/CFI/RMSEA = 794/267/0/.884/.850/.940/.064). The resulting loadings and measurement error variances for S, A, I, C and Y in the structural model also were judged sufficiently similar to those from the measurement model for S, A, I, C and Y, and that a second structural model estimation was not necessary.

The results suggested that AI and II were significant in Equation 9, but they also suggested that SS and AA were not significant.

CUBICS IN S AND A

Because the exact mathematical form of the hypothesized "diminishing returns-like" association between Y,

---

[7] AI is content or face valid if A and I are content valid and the specification of AI includes all the indicators of A and I. The formula for the Average Variance Extracted (AVE) of a quadratic is $\Sigma(\lambda_{xi}\lambda_{xj})^2\text{Var}(XX)/[\Sigma(\lambda_{xi}\lambda_{xj})^2\text{Var}(XX) + \Sigma\text{Var}(\varepsilon_{xx})]$, where $\Sigma(\lambda_{xi}\lambda_{xj})^2$ is the sum of squares of $\lambda_{xi}\lambda_{xj}$, i = 1 to n, j = 1 to n, i can equal j, n is the number of indicators of X, and Var(XX) is the error dissattenuated variance of XX (available in the structural model) (Fornell and Larker 1981) (see Equations 2 and 2a). However, the construct

and S and A were not hypothesized, the nonsignificant SS and AA were replaced in the Equation 9 model with $S_c(SS)_c$ and $A_c(AA)_c$ to produce the structural model,

$$Y = b_1S + b_2A + b_3I + b_4C + b_5AI + b_6II + b_9SSS + b_{10}AAA + \zeta \qquad (\cancel{9a,}\ 11$$

The consistency, reliability and validity of the cubics were gauged next.

The latent variable $S_c(SS)_c$ was judged to be trivially internally consistent, but to judge its external consistency, several steps were required. To specify $S_c(SS)_c$ in the Equation 11 measurement model, the product indicator $s_c(ss)_c = [\Sigma s_j^u - \mathbf{M}(\Sigma s_j^u)][(\Sigma s_j^u)(\Sigma s_j^u) - \mathbf{M}((\Sigma s_j^u)(\Sigma s_j^u))]$ was computed in each case, where s_j^u are the uncentered indicators of S, and **M** denotes a mean. Specifically, using the results from the measurement model that was used to gauge the external consistency of S, A, I, C and Y, and Equations 4e and 4f, the loading and measurement error variance of $S_c(SS)_c$ was computed. Similarly, the loading and measurement error variance of $A_c(AA)_c$ was computed. Then, the Equation 11 measurement model was estimated using LISREL 8 and Maximum Likelihood estimation by fixing the loadings and measurement error variances for the single indicators of AI and II at their previous values, and the loadings and measurement error variances for $S_c(SS)_c$, and $A_c(AA)_c$ at their computed values using the cubic EXCEL template shown on the previous web page. The resulting loadings and measurement error variances for S, A, I, C and Y in the structural model were judged sufficiently similar to those from the measurement model for S, A, I, C and Y, and that a second measurement model estimation was not necessary. Because the resulting measurement model fit the data $S_c(SS)_c$, and $A_c(AA)_c$ were judged to be externally consistent. (Experience suggests that in real world data this cubic specification is externally consistent.)

The reliability of $S_c(SS)_c$ is

$$\rho_{Sc(SS)c} = \frac{\mathrm{Corr}^2(S_c,(SS)_c) + \rho_{Sc}\rho_{(SS)c}}{\mathrm{Corr}^2(S_c,(SS)_c) + 1}, \qquad (12$$

(see Equation 9a) where ρ denotes reliability and $\mathrm{Corr}^2(S_c,(SS)_c)$ is the square of the correlation between S_c and $(SS)_c$ (Bohrnstedt and Marwell 1978). However, estimates of $\mathrm{Corr}(S_c,(SS)_c)$ were difficult to obtain because S_c and $(SS)_c$ were nonessentially ill-conditioned (highly correlated) and their measurement model

---

(correlational) validity of a second-order is usually impossible to judge.

produced correlations greater than one. Nevertheless, assuming $Corr^2(S_c,(SS)_c)$ is 1, and substituting the reliability of S_c (= ρ_S) and the reliability of $(SS)_c = S_uS_u$ - $\mathbf{M}(S_uS_u)$ into Equation 12,

$$\rho_{(SS)c} = \frac{\mu^2\rho_{Su} + \mu^2\rho_{Su} + 2\mu^2 Corr(S_u,S_u) + Corr\ ^2(S_u,S_u) + \rho_{Su}\rho_{Su}}{\mu^2 + \mu^2 + 2\mu^2 Corr(S_u,S_u) + Corr\ ^2(S_u,S_u) + 1},$$

$$= \frac{2\mu^2\rho_{Su} + 2\mu^2 + 1 + \rho_{Su}\rho_{Su}}{4\mu^2 + 2} \qquad (12a$$

$$= \frac{2\mu^2\rho_S + 2\mu^2 + 1 + \rho_S^2}{4\mu^2 + 2} \qquad (12b$$

(Bohrnstedt and Marwell 1978) (μ = $\mathbf{M}(\Sigma i_j)/SD(S)$, where **M** denotes mean and SD denotes standard deviation), produced a rough estimate of the reliability of $S_c,(SS)_c$ ($\rho_{Sc(SS)c} \approx .81$). Another reliability estimate for $S_c,(SS)_c$ was available from its squared multiple correlation in the Equation 11 measurement model (and the Equation 11 structural models, estimated later) ($\rho_{Sc(SS)c} \approx .80$).[8] Using these roughly similar results, $S_c,(SS)_c$ was judged to be reliable, and it was judged to be valid.[9] Using similar estimates, $A_c,(AA)_c$ was judged to be reliable and valid.

The results of the Maximum Likelihood estimation using LISREL are shown in Table A. In summary, the first-order variables were associated with Y as hypothesized, except for S. In addition, I moderated the A-Y association, and the Y associations with S, A and I were "diminishing returns-like."

Because of the significant non-linear terms the coefficients of S, A and I should be interpreted using factored coefficients. The interpretation for the factored coefficient for A ($(b_2+b_5I+b_{10}AA)A$) is shown in Table

---

[8] The squared multiple correlation of $S_c,(SS)_c$ is equivalent to Werts, Linn and Jöreskog's (1974) proposed formula for computing latent variable reliability, $(\Sigma\lambda_{xi})^2Var(X)/[(\Sigma\lambda_{xi})^2Var(X) + Var(\Sigma\varepsilon_{xi})]$. Experience suggests that for second-order latent variables it is practically equivalent to Bohrnstedt and Marwell (1978) calculated results for unstandardized X in real-world data.

[9] $S_c,(SS)_c$ is content or face valid if S is content valid and the specification of $S_c,(SS)_c$ includes all the indicators of S. The formula for the Average Variance Extracted (AVE) of a cubic in the proposed specification is $\Sigma(\lambda_{xi}{}^2\lambda_{xj})^2Var(XW)/[\Sigma(\lambda_{xi}{}^2\lambda_{xj})^2Var(XW) + \Theta_{XW}]$ (Fornell and Larker 1981) (see Equations 4a through 4d), where $\Sigma(\lambda_{xi}{}^2\lambda_{xj})^2$ is the sum of the squares of $\lambda_{xi}{}^2\lambda_{xj}$, i = 1 to n, j = 1 to n, j ≥ i, Var(XW) is the error dissattenuated variance of XW (available in the structural model) and n is the number of indicators of X. However, the construct (correlational) validity of a third-order is usually impossible to evaluate.

E and has already been discussed. The interpretation of the factored coefficient for I is similar (not shown). It turns out that S does not have a factored coefficient in the population because the coefficient of S is NS (i.e., the factored coefficient is $(0+b_9SS)S = b_9SSS$). Thus SSS cab be interpreted using more familiar techniques such as a graph of Y (= SSS) versus S (see Table F).

Since the Y associations with S A and I were hypothesized to exhibit "diminishing returns-like" behavior, graphs of Y versus S, Y versus A, and Y versus I may be useful in order to verify this hypothesized behavior (see Table F and G).

COMMENTS

While it is reasonable to expect the Equation 4 and 4a specification of a cubic to be acceptably unbiased and consistent, such a demonstration is beyond the scope of this paper. (However, a simulated data set generated with the Equation 9 structural model parameters and the psychometrics of S, A, I, C, AI, II, SS and AA produced results that were practically equivalent to those shown in Table A. This hints that the specification *may* be trustworthy.)

As mentioned earlier, quadratics and cubics are difficult to estimate together. Centering the first-order variables (e.g., X and Z) reduces the collinearity between X, for example, and XX. However, it does not reduce the collinearity between XX and XXX. In the above example, the error-free correlation between SS and SSS, for example, was almost 1, and estimation was impossible without using LISREL's Ridge Option, and attendant inflated structural coefficient estimates and standard errors. Such estimates are inconsistent (i.e., they are likely to be very different in the next study), and thus of little use in theory testing. While there are several ways to center XX to reduce the collinearity between XX and XXX, they invariably increase the collinearity between X and XX, which again produces inconsistent estimates. Thus, we may never know if quadratics and cubics can actually occur together in (real world) survey data.

Some authors believe interactions and quadratics are more likely than their reported occurrence in published survey research suggests. Our experience and the above example hint that quadratics and cubics may be even more common that interactions in survey data (e.g., Howard 1989) (also see McClelland and Judd 1993 for evidence that suggests that interactions in survey data should not be common).

CONCLUSION

The purpose of the paper was to suggest a specification for a latent variable cubic. A pedagogical example was also provided that illustrated its estimation and interpretation. The reliability and standard error of a cubic was derived, and as a separate matter, the pedagogical example contained a suggested sequential procedure for testing hypothesized satiation or diminishing returns in a first-order latent variable (e.g., the latent variables S and A).

While the suggested specification of a latent variable cubic is a small step forward, more work remains to be done to demonstrate that this specification is acceptably unbiased and consistent. Specifically, because the suggested specification used the Kenny and Judd (1984) normality assumptions, and the pedagogical example used maximum likelihood estimation, the suggested specification should be formally investigated with the nonnormal data and small sample sizes typical of substantive theoretical model testing, and with low reliabilities and small cubic coefficient sizes (e.g., Ping 1995). These investigations were beyond the exploratory purposes of this paper.

REFERENCES

Aiken, Leona S. and Stephen G. West (1991), *Multiple Regression: Testing and Interpreting Interactions*, Newbury Park, CA: SAGE Publications.

Algina, James and Bradley C. Moulder (2001), "A Note on Estimating the Jöreskog-Yang Model for Latent Variable Interaction Using LISREL 8.3," Structural Equation Modeling, 8 (1) 40-52.

Anderson, James C. and David W. Gerbing (1984), "The Effect of Sampling Error on Convergence, Improper Solutions, and Goodness of Fit Indices for Maximum Likelihood Confirmatory Factor Analysis," *Psychometrika*, 49, 155-173.

Bandura, Albert (1966), "Vicarious Processes: a Case of No Trial Learning, in *Advances in Experimental Social Psychology*, L. Berkowitz, ed., NY: Academic Press.

Blalock, Hubert M., Jr. (1965), "Theory Building and the Concept of Interaction," *American Sociological Review*, 30, 374-381.

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables,* New York: Wiley.

Bollen, Kenneth A. (1995), "Structural Equation Models that are Nonlinear in Latent Variables: A Least Squares Estimator," *Sociological Methodology*, 25, 223-251.

Bohrnstedt, G.W. and G. Marwell (1978), "The Reliability of Products of Two Random Variables," in Sociological Methodology, K. F. Schuessler ed., San Francisco: Jossy Bass, 254-273.

Browne, Michael W. and Robert Cudeck (1993), "Alternative Ways of Assessing Model Fit," in *Testing Structural Equation Models*, K. A. Bollen et al. eds, Newbury Park CA: SAGE Publications.Kendall,

M.G. and A. Stuart (1958), *The Advanced Theory of Statistics*, Vol. 1, London: Charles Griffith.
Busemeyer, Jerome R. and Lawrence E. Jones (1983), "Analysis of Multiplicative Combination Rules When the Causal Variables are Measured With Error," *Psychological Bulletin*, 93 (May), 549-62.
Campbell, Donald T. (1960), "Recommendations for APA Test Standards Regarding Construct, Trait and Discriminant Validity," *American Psychologist*, 15, 546-553.
Cohen, Jacob (1968), "Multiple Regression as a General Data-Analytic System," *Psychological Bulletin*, 70, 426-43.
_______ and Patricia Cohen (1975), *Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences*, Hillsdale, NJ: Lawrence Erlbaum.
_______ and Patricia Cohen (1983), *Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences*, Hillsdale, NJ: Lawrence Erlbaum.
Darlington, R. B. (1990), *Regression and Linear Models*, New York: McGraw-Hill.
DeVellis, Robert F. (1991), *Scale Development: Theory and Applications*, Newbury Park, CA: SAGE Publications.
Dunlap, William P. and Edward R. Kemery (1988), "Effects of Predictor Intercorrelations and Reliabilities on Moderated Multiple Regression," *Organizational Behavior and Human Decision Processes*, 41, 248-258.
Fornell, Claes and David F. Larker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.
Friedrich, R. J. (1982), "In Defense of Multiplicative Terms in Multiple Regression Equations," *American Journal of Political Science*, 26, 797-833.
Heeler, Roger M. and Michael L. Ray (1972), "Measure Validation in Marketing," *Journal of Marketing Research*, 9 (November), 361-70.
Homans, G. C. (1974), *Social Behavior: Its Elementary Forms*, 2nd Edn., New York: Harcourt, Brace Jovanovich.
Howard, John A. (1989), *Consumer Behavior in Marketing Strategy*, Englewood Cliffs, NJ: Prentice Hall.
Jaccard, James, Robert Turrisi and Choi K. Wan (1990), *Interaction Effects in Multiple Regression*, Newbury Park, CA: SAGE Publications.
Jaccard, James and C. K. Wan (1995), "Measurement Error in the Analysis of Interaction Effects Between Continuous Predictors Using Multiple Regression: Multiple Indicator and Structural Equation Approaches," *Psychological Bulletin*, 117 (2), 348-357.
Jöreskog, Karl G. (1993), "Testing Structural Equation Models," in *Testing Structural Equation Models*, Kenneth A. Bollen and J. Scott Long eds., Newbury Park, CA: SAGE.
Jöreskog, Karl G. and Fan Yang (1996), "Nonlinear Structural Equation Models: The Kenny and Judd Model with Interaction Effects," *Advances in Structural Equation Modeling Techniques*, G.A. Marcoulides, R.E. Schumacker, eds., Hillsdale, NJ: LEA.
Jöreskog, Karl G. (2000), "Latent Variable Scores and Their Uses," (on-line paper), http://www.ssicentral.com/lisrel/ techdocs/lvscores.pdf.
Kendall, M.G. and A. Stuart (1958), *The Advanced Theory of Statistics*, Vol. 1, London: Charles Griffith.
Kenny, D. and C. M. Judd (1984), "Estimating the Nonlinear and Interactive Effects of Latent Variables," *Psychological Bulletin*, 96, 201-10.
Kenny, (1985), "Quantitative Methods for Social Psychology," in *Handbook of Social Psychology*, Vol. 1, G. Lindsey et al. eds., New Your: Random House, 487-508.
Klein, A.G. and H. Moosbrugger (2000), "Maximum Likelihood Estimation of Latent Interaction Effects with the LMS Method," *Psychometrika*, 65, 457-474.
Klein, A.G. and B.O. Muthén (2002), "Quasi Maximum Likelihood Estimation of Structural Equation Models with Multiple Interactions and Quadratic Effects," Unpublished ms., Graduate School of Education, UCLA.
Laroche, M. and J. A. Howard (1980), "Nonlinear Relations in a Complex Model of Buyer Behavior," *Journal of Consumer Research*, 6 (4) (march), 377-388.

Lubinski, D. and Humphreys, L.G. (1990), "Assessing Spurious Moderator Effects: Illustrated Substantively with the Hypothesized ("Synergistic") Relation Between Spatial and Mathematical Ability," *Psychological Bulletin*, 107, 385-393.

Marquardt, D.W. (1980), "You Should Standardize the Predictor Variables in your Regression Models," *Journal of the American Statistical Association*, 75, 87-91.

Marsh, Herbert W., Zhonglin Wen and Kit-Tai Hau (2004), "Structural Equation Models of Latent Interactions: Evaluation of Alternative Estimation Strategies and Indicator Construction," *Psychological Methods*, 9 (3), 275-300.

Mathieu, J.E., S.I. Tannenbaum and E. Salas (1992), "Influences of Individual and Situational Characteristics on Measuring of Training Effectiveness," *Academy of Management Journal*, 35, 828-847.

McClelland, G. H. and C. M. Judd (1993), "Statistical Difficulties of Detecting Interactions and Moderator Effects," *Psychological Bulletin*, 114 (2), 376-390.

Molder, Bradley C. and James Algina (2002), "Comparison of Methods for Estimating and Testing Latent Variable Interactions," *Structural Equation Modeling*, 9 (1), 1-19.

Neter, John, William Wasserman and Michael H. Kunter (1989), *Applied Linear Regression Models*, 2nd ed'n., Homewood, IL: Irwin.

Nunnally, Jum C. (1978), *Psychometric Theory*, 2nd Ed., New York: McGraw-Hill.

Pedhazur, E.J. (1987), *Multiple Regression in Behavioral Research*, New York: Holt, Rinehart and Winston.

Peter, J. Paul (1981), “Construct Validity: A Review of Basic Issues and Marketing Practices,” *Journal of Marketing Research*, 18 (May), 133-45.

Ping, R. (1995), "A Parsimonious Estimating Technique for Interaction and Quadratic Latent Variables," *The Journal of Marketing Research*, 32 (August), 336-347.

Ping, R. (1996a), "Latent Variable Interaction and Quadratic Effect Estimation: A Two-Step Technique Using Structural Equation Analysis," *Psychological Bulletin*, 119 (January), 166-175.

Ping, R. (1996b), "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients," *Multivariate Behavioral Research*, 31 (1), 95-120.

Ping, R.A. (2006), "Frequently Asked Questions (FAQ's) About Interactions and Quadratics." [on-line paper]. http://home.att.net/~rpingjr/Faq.doc .

Ping, R.A. (2006a). "An excel template for obtaining an internally consistent set of items." [on-line paper]. http://home.att.net/~rpingjr/weeding1.xls.

Ping, R.A. (2006a), "An excel template for specifying cubics." [on-line paper]. http://home.att.net/~rpingjr/Cubic.xls.

Schermelleh-Engle, K., A. Kein and H. Moosbrugger (1998), "Estimating Nonlinear Effects using a Latent Moderated Structural Equations Approach," in *Interaction and Nonlinear Effects in Structural Equation Modeling*, R. E, Schumacker and G. A. Marcoulides, eds., Mahwah, NJ: Erlbaum.

Wall, M.M. and Y. Amemiya (2001), "Generalized Appended Product Indicator Procedure for Nonlinear Structural Equation Analysis," *Journal of Educational and Behavioral Statistics*, 26, 1-29.

Werts, C.E., R.L. Linn and K.G. Jöreskog (1974), "Intraclass Reliability Estimates: Testing Structural Assumptions," *Educational and Psychological Measurement*, 34, 25-33.

Wheaton, Blair (1985), "Personal Resources and Mental Health: Can There Be Too Much of a Good Thing?," in *Research in Community and Mental Health*, J.R. Greenley, ed., Greenwich, CT: JAI.

Yerkes, R.M. and J.D. Dodson (1908), "The Relation of Strength of Stimulus to Rapidity of Habit Formation, *Journal of Comparative Neurology of Psychology*, 18, 459-482.

Table A- Equation 9a Estimation Results[a]

$$Y = -.061S + .191A - .191I + .066C + .142AI - .092II - 0.020SSS + 0.015AAA + \zeta\ (= .346)$$

	S	A	I	C	AI	II	SSS	AAA	ζ	
	(0.063)	(0.048)	(0.060)	(0.021)	(0.058)	(0.045)	(0.005)	(0.004)	(0.037)	(SE)
	-0.97	3.95	-3.17	3.12	2.43	-2.03	-3.91	3.37	9.32	t-value

$\chi^2 = 806$ GFI[b] = .88 CFI[c] = .940

df = 267 AGFI[b] = .85 RMSEA[d] = .067 R^2 for Y = .216

---

[a] Using LISREL and Maximum Likelihood.

[b] Shown for completeness only-- GFI and AGFI may be inadequate for fit assessment in larger models (see Anderson and Gerbing 1984).

[c] .90 or higher indicates acceptable fit (see McClelland and Judd 1993).

[d] .05 suggests close fit, .051-.08 suggests acceptable fit (Brown and Cudeck 1993, Jöreskog 1993).

Table B- Variable Y Associations with A and I Due to the AI Interaction in Table A

I-Y Association Moderated by A[a]					A-Y Association Moderated by I[f]					
A Level[b]	Centered A[c]	I Coefficient[d]	SE of I Coefficient[e]	t-value of I Coefficient	I Level[g]	Centered I[h]	A Coefficient[i]	SE of A Coefficient[j]	t-value of A Coefficient	
1	-1.54	-0.411	0.102	-4.05	1	-2.8	-0.208	0.169	-1.23	
2	-0.54	-0.269	0.064	-4.18	2	-1.8	-0.065	0.114	-0.57	
2.54[2]	0	-0.192	0.060	-3.17	3	-0.8	0.077	0.066	1.18	
3	0.46	-0.126	0.069	-1.81	3.8[k]	0	0.192	0.048	3.96	
4	1.46	0.017	0.111	0.15	4	0.2	0.220	0.050	4.36	
5	2.46	0.160	0.164	0.97	5	1.2	0.363	0.088	4.14	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(Column Number)

[a] The table displays the unstandardized associations of A and I with Y produced by the significant AI interaction in Table A (see Footnotes d and i). In Columns 1-5 when the existing level of A was low in Column 1, (small) changes in I were negatively associated with Y (see Columns 3 and 5). At higher levels of A however, I was less strongly associated with Y, until above the study average for A, the association was nonsignificant (see Column 5). For higher levels of A, I was negatively but nonsignificantly associated with Y.

[b] A is determined by the observed variable (indicator) with the loading of 1 on A (i.e., the indicator that provides the metric for A). The value of this indicator of A ranged from 1 (= low A) to 5 in the study.

[c] Column 1 minus 2.54.

[d] The factored (unstandardized) coefficient of I was (-.191+.142A)I with A mean centered. E.g., when A = 1 the coefficient of I was -.191+.142*(1 - 2.54) ∇ -.411.

[e] The Standard Error of the A coefficient was:

$$\%\overline{\mathrm{Var}(\beta_I+\beta_{AI}A)} = \%\overline{\mathrm{Var}(\beta_I) + A^2\mathrm{Var}(\beta_{AI}) + 2A\mathrm{Cov}(\beta_I,\beta_{AI})} = \%\overline{\mathrm{SE}(\beta_I)^2 + A^2\mathrm{SE}(\beta_{AI})^2 + 2A\mathrm{Cov}(\beta_I,\beta_{AI})}\ ,$$

where Var and Cov denote variance and covariance, SE is the standard error, and β denotes unstandardized structural coefficients from Table A.

[f] This portion of the table displays the unstandardized associations of A and Y moderated by I. When I was low in Column 6, the A association with Y was not significant (see Column 10). However, as I increased, A's association with Y strengthened, until at the study average it was positively associated with Y (see columns 8 and 10).

[g] I is determined by the observed variable (indicator) with the loading of 1 on I (i.e., the indicator that provides the metric for I). The value of this indicator of I ranged from 1 (= low Z) to 5 in the study.

[h] Column 6 minus 3.8.

[i] The factored (unstandardized) coefficient of A was (.191+.142I)A with I mean centered. E.g., when I = 1 the coefficient of A was .191+.142*(1-3.8) ∇ -.208.

[j] The Standard Error of the A coefficient is

$$\%\overline{\mathrm{Var}(\beta_A+\beta_{AI}I)} = \%\overline{\mathrm{Var}(\beta_A) + I^2\mathrm{Var}(\beta_{AI}) + 2I\mathrm{Cov}(\beta_A,\beta_{AI})} = \%\overline{\mathrm{SE}(\beta_A)^2 + I^2\mathrm{SE}(\beta_{AI})^2 + 2I\mathrm{Cov}(\beta_A,\beta_{AI})}\ ,$$

where Var and Cov denote variance and covariance, SE is the standard error, and β denotes unstandardized structural coefficients from Table A.

[k] Mean value in the study.

Table C- Variable Y Associations with I Due to the II Quadratic in Table A

I-Y Association Moderated by I[a]				
			SE of	t-value
	Cen-	I	I	of I
I	tered	Coef-	Coef-	Coef-
Level[b]	I[c]	ficient[d]	ficient[e]	ficient
1	-2.8	0.067	0.110	0.61
2	-1.8	-0.025	0.073	-0.34
3	-0.8	-0.118	0.053	-2.23
3.8[f]	0	-0.192	0.060	-3.17
4	0.2	-0.210	0.065	-3.21
5	1.2	-0.303	0.099	-3.04
(1)	(2)	(3)	(4)	(5)

(Column Number)

[a] The table displays the unstandardized associations of I with Y produced by the significant quadratic II shown in Table A (see Footnote d). When the existing level of I was low in Column 1, small changes in I were not associated with Y (see Column 2). As the Column 1 level of I increased in the study however, I's association with Y strengthened (i.e., became larger in Column 2), and small changes in I were significantly associated with Y when the level of I was at or above 3 (see Column 4).

[b] I is determined by the observed variable (indicator) with the loading of 1 on I (i.e., the indicator that provides the metric for I). The value of this indicator of I ranged from 1 (= low I) to 5 in the study.

[c] Column 1 minus 3.8.

[d] The factored (unstandardized) coefficient of I was (-.191-.092I)I with I mean centered. E.g., when I = 1 the coefficient of I was -.191 -.092*(1-3.8) = .067.

[e] The Standard Error of the I coefficient was:

$$\sqrt{Var(\beta_I+\beta_I I)} = \sqrt{Var(\beta_I)+I^2*Var(\beta_{II})+2*I*Cov(\beta_I,\beta_{II})} = \sqrt{SE(\beta_I)^2 + I^2 SE(\beta_{II})^2 + 2ICov(\beta_I,\beta_{II})} ,$$

where Var and Cov denote variance and covariance, SE is the standard error, and β denotes unstandardized structural coefficients from Table A.

[f] The mean of I in the study.

Table D- Variable Y Associations with A Due to the AAA Cubic in Table A

A-Y Association Moderated by the Level of A[a]

A Level[b]	Centered A Value[c]	Centered AA Value[d]	A Coefficient[e]	SE of A Coefficient[f]	t-value of A Coefficient
1	-1.54	2.37	0.229	0.047	4.90
2	-0.54	0.29	0.196	0.048	4.08
2.54[g]	0	0	0.192	0.048	3.96
3	0.46	0.21	0.195	0.048	4.05
4	1.46	2.13	0.225	0.047	4.81
5	2.46	6.05	0.286	0.049	5.84
(1)	(2)	(3)	(4)	(5)	(6)

(Column Number)

[a] The table displays the unstandardized association of A with Y produced by the significant cubic AAA in Table A (see Footnote f). When the existing level of A was low in Column 1, small changes in the level of A were positively associated with in Y (see Column 4). As the Column 1 level of A increased in the study, A's association with Y weakened slightly (i.e., became smaller in Column 4), and it's significance declined. However, for A above the study average its association with Y increased again and significance increased (see Column 6).

[b] A is determined by the observed variable (indicator) with the loading of 1 on A (i.e., the indicator that provides the metric for A). This indicator of A ranged from 1 (= low A) to 5 in the study.

[c] Column 1 minus 2.54.

[d] Equals the square of Column 2.

[e] The factored (unstandardized) coefficient of A was (.191+.015AA) with A mean centered. E.g., when A = 1 the coefficient of A is .191 + .015*2.37 ∇ .229.

[f] The Standard Error of the A coefficient is:

$$\%\overline{\mathrm{Var}(\beta_A+\beta_{AAA}AA)} = \%\overline{\mathrm{Var}(\beta_A)+AA^2\mathrm{Var}(\beta_{AAA})+2AA\mathrm{Cov}(\beta_A,\beta_{AAA})}$$

$$= \%\overline{\mathrm{SE}(\beta_A)^2 + AA^2\mathrm{SE}(\beta_{AA})^2 + 2AA\mathrm{Cov}(\beta_A,\beta_{AA})}\ ,$$

where Var and Cov denote variance (= SE^2) and covariance, and β denotes the unstandardized structural coefficients shown in Footnote e.

[g] The mean of X in the study.

Table E- Variable Y Associations with A Due to the AI Interaction and the AAA Cubic in Table A

A-Y Association Moderated by the Level of I and AA[a]								
I Level[b]	Centrd I[c]	A Coefficient[d]						
1	-2.8	-0.173	-0.204	-0.208	-0.204	-0.173	-0.111	
		0.166	0.168	0.169	0.168	0.166	0.162	SE[e]
		-1.04	-1.21	-1.23	-1.21	-1.04	-0.68	t-value
2	-1.8	-0.030	-0.061	-0.065	-0.061	-0.030	0.032	
		0.111	0.113	0.114	0.113	0.111	0.107	SE[e]
		-0.27	-0.54	-0.57	-0.54	-0.27	0.30	t-value
3	-0.8	0.112	0.081	0.077	0.081	0.112	0.175	
		0.062	0.065	0.066	0.065	0.062	0.061	SE[e]
		1.80	1.25	1.18	1.25	1.80	2.88	t-value
3.8	0	0.227	0.195	0.192	0.195	0.227	0.289	
		0.047	0.048	0.048	0.048	0.047	0.049	SE[e]
		4.86	4.06	3.96	4.06	4.86	5.87	t-value
4	0.2	0.255	0.224	0.220	0.224	0.255	0.318	
		0.049	0.050	0.050	0.050	0.049	0.053	SE[e]
		5.17	4.46	4.36	4.46	5.17	6.01	t-value
5	1.2	0.398	0.367	0.363	0.367	0.398	0.460	
		0.089	0.088	0.088	0.088	0.089	0.094	SE[e]
		4.48	4.18	4.14	4.18	4.48	4.91	t-value
		2.25	0.25	0	0.25	2.25	6.25	Centrd AA[f]
		-1.5	-0.5	0	0.5	1.5	2.5	Centrd A[g]
		1	2	2.54	3	4	5	A Level
		(1)	(2)	(3)	(4)	(5)	(6)	(Column Number)

[a] The table displays the unstandardized associations of A with Y produced by the significant interaction AI and cubic AAA in Table A (see Footnote d). When the existing levels of I and A were low in Column 1, row 1, small changes in the level of A were negatively but nonsignificantly associated with in Y. As the Column 1 level of I increased in the study, A's association with Y weakened then turned positive and became significant at and above the study average of I. This pattern was more or less consistent for existing levels of A in Columns 2 through 6. (Thus the cubic contributed little)

[b] I is determined by the observed variable (indicator) with the loading of 1 on I (i.e., the indicator that provides the metric for I). This indicator of I ranged from 1 (= low I) to 5 in the study.

[c] Rows minus 3.8.

[d] The factored (unstandardized) coefficient of A was (.191+.142I+.015AA) (see Table A) with A mean centered. E.g., when I = 1 and A = 1 the coefficient of A was .191 + .142*(1-3.8) + .015*$(1-2.54)^2$ ∇ -.173.

[e] The Standard Error of the A coefficient is:

$$\%\overline{Var(\beta_I+\beta_{AI}I+\beta_{AAA}AA)} = \%\overline{Var(\beta_I)+I^2Var(\beta_{AI})+AA^2Var(\beta_{AAA})+2IAACov(\beta_{AI},\beta_{AAA})+2ICov(\beta_I,\beta_{AI})+2AACov(\beta_I,\beta_{AAA})}$$

$$= \%\overline{SE(\beta_I)^2+I^2SE(\beta_{AI})^2+AA^2SE(\beta_{AA})^2+2IAACov(\beta_{AI},\beta_{AAA})+2ICov(\beta_I,\beta_{AI})+2AACov(\beta_I,\beta_{AAA})}) ,$$

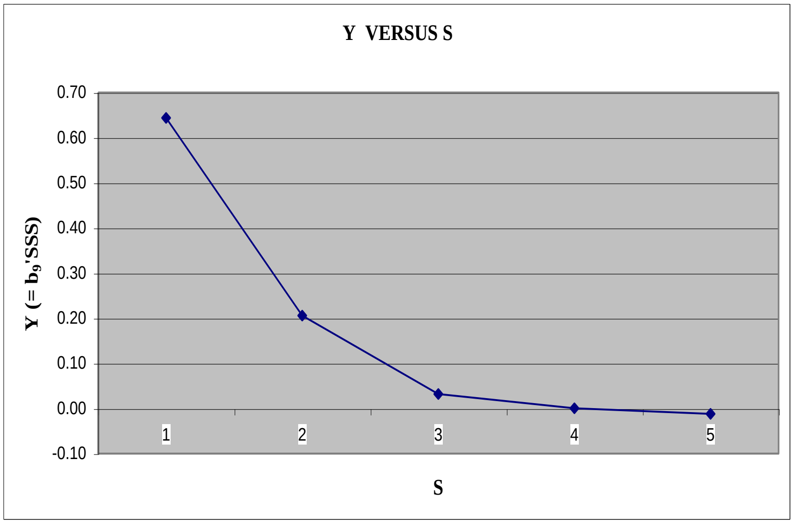
where Var and Cov denote variance (= SE^2) and covariance respectively, and β denotes the unstandardized structural coefficients shown in Footnote d.

[f] Equals the square of the values in the row below.

[g] Column values minus 2.54.

Table F- The Observed Relationship Between Y and S ($Y = b_9SSS$) in Equation 9a[a] (Note: to view the complete graph zoom to 150%)

Y at Selected Levels of S		
S Level	Centrd S	Y = b_9SSS[b]
1	-3.16	0.64
2	-2.16	0.21
3	-1.16	0.03
4	-0.16	0
5	0.84	-0.01



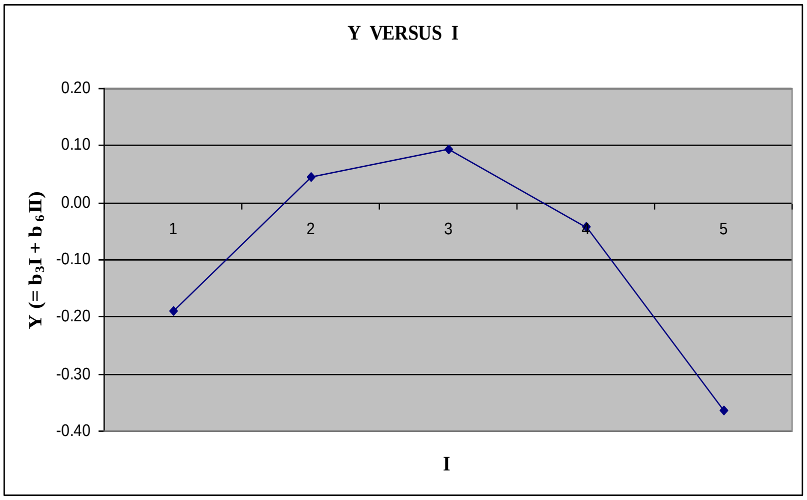
[a] The table and graph display (predicted) $Y = b_9SSS$ at selected values of S in the study (with the other variables held constant). As suggested by the graph, as S increased from 1 to 5 in the study (with the other variables held constant), Y was likely to decrease at a declining rate, with a "scree point" at S = 3. From that point on Y was likely to change little as S increased. Thus S's association with Y exhibited "diminishing returns" to S, or satiation in Y.

[b] Using centered S and centered Y. This illustrates a difficulty with the required use of centered data. Y actually ranged from 1 to 5 in the study, and did not assume values below 1. Nevertheless, the shape of the graph is correct, and thus the interpretation in Footnote a is also correct.

Table G- The Observed Relationships Between Y and I and A in Equation 9a[a] (Note: to view the complete graphs zoom to 150%)

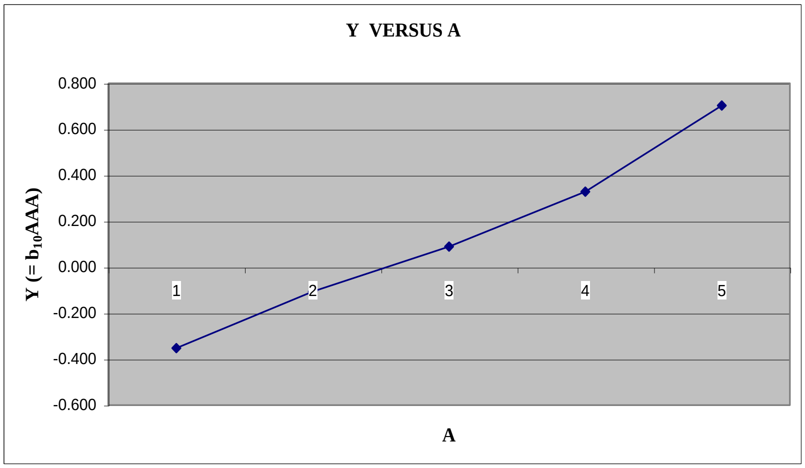
(a) Y at Selected Levels of I[b]

I Level	Centrd I	$Y = b_3I + b_6II$
1	-2.8	-0.19
2	-1.8	0.05
3	-0.8	0.09
4	0.2	-0.04
5	1.2	-0.36



(b) Y at Selected Levels of A[c]

A Level	Centrd A	$Y = b_{10}AAA$
1	-1.54	-0.35
2	-0.54	-0.11
3	0.46	0.09
4	1.46	0.33
5	2.46	0.7



[a] The tables and graphs display (predicted) Y at selected values of I and A in the study (with the other variables held constant).

[b] Using centered I and centered Y. As suggested by the graph, as the variable I increased from 1 to 5 in the study (with the other variables held constant), Y was likely to increase at a declining rate, with maximum at I = 3. From that point on Y was likely to decrease as the variable I increased. Thus I was associated with Y as hypothesized, and it exhibited "diminishing returns" to S, or satiation in Y at lower I. However at higher levels of I, increasing the variable I increased or amplified Y.

[c] Using centered A and centered Y. As suggested by the graph, as the variable A increased from 1 to 5 in the study (with the other variables held constant), Y was likely to increase at a slightly declining rate. However, at approximately A = 3, further increases in A were likely to increase Y at a slightly increasing rate. Thus A was associated with Y as hypothesized, and it exhibited "diminishing returns" to A, or satiation in Y at lower I. However at higher levels of A, the association between Y and A changed slightly, requiring further explanation.

ESTIMATING LATENT VARIABLE INTERACTIONS AND QUADRATICS: EXAMPLES, SUGGESTIONS AND NEEDED RESEARCH

Robert A. Ping, Jr.
Department of Marketing
College of Business Administration
Wright State University
Dayton, OH 45435
513-873-3047

December, 1995

ESTIMATING LATENT VARIABLE INTERACTIONS AND QUADRATICS: EXAMPLES, SUGGESTIONS AND NEEDED RESEARCH

(Note: most of the revisions are in parentheses)

ABSTRACT

Because there is little guidance for substantive researchers in the estimation of latent variable interactions and quadratics in theory tests, the paper provides several examples of their estimation using EQS and LISREL 8. After summarizing the available structural equation techniques for estimating these latent variables, procedures for their use are suggested including procedures for obtaining convergence and proper estimates. Examples of several techniques for the direct and indirect estimation of these latent variables using EQS and LISREL are provided. The paper concludes with suggestions for future research in this area.

INTRODUCTION

In studies reported in the social science literature that involve categorical independent variables (ANOVA studies), interactions (e.g., XZ in $Y = b_0 + b_1X + b_2Z + b_3XZ + b_4XX$) and quadratics (XX in the equation just mentioned) are routinely investigated to help interpret significant first-order (main) effects. Interaction and quadratic variables are also investigated in studies involving continuous variables and regression, although not routinely, and not to aid interpretation as they are in ANOVA studies. Typically, continuous interactions and quadratics are investigated in response to theory that proposes their existence.

However, researchers in the social sciences have called for the investigation of interaction and quadratic variables to improve interpretation in models that involve continuous variables (Aiken & West, 1991; Blalock, 1965; Cohen, 1968; Cohen & Cohen, 1975, 1983; Howard, 1989; Jaccard, Turrisi & Wan, 1990; Kenny, 1985). Their argument is that failing to consider the presence of interactions and quadratics in the population model increases the risk of false negative research findings, and positive research findings that are conditional in the population equation. To explain, in a model such as $Y = b_0 + b_1X + b_2Z + b_3XZ$, the coefficient of Z is given by $Y = b_0 + b_1X + (b_2 + b_3X)Z$. The significance of this coefficient could be very different from the significance of the Z coefficient in the same model specified without the XZ interaction (i.e., $Y = b_0' + b_1'X + b_2'Z$). Specifically, b_2' could be nonsignificant while $b_2 + b_3X$ could be significant over part(s) of the range of X. Failing to specify this significant interaction would lead to a false disconfirmation of the Z-Y association. Alternatively b_2' could be significant while $b_2 + b_3X$ could be nonsignificant over part of the range of X. Failing to specify this significant interaction could produce a misleading picture of the contingent Z-Y association. The implications for failing to specify a population quadratic are similar.

Fortunately there has been considerable progress in the estimation of interactions and quadratics using structural equation analysis recently. Several specification techniques have been proposed (see Hayduk, 1987; Kenny & Judd, 1984; Ping, 1995, 1996; Wong & Long, 1987), and LISREL 8 has become commercially available. Along with subgroup analysis (Jöreskog, 1971a), these alternatives offer substantive researchers considerable power and flexibility in estimating latent variable interaction and quadratic effects in structural equation models. After briefly summarizing these alternatives, an illustration of the use of several of them using field survey data, EQS, and LISREL 8 is provided. The paper concludes with suggested procedures for the use of several techniques, and a discussion of topics for further research in this area.

LATENT VARIABLE INTERACTION AND QUADRATIC SPECIFICATION AND ESTIMATION

Estimation techniques for interaction and quadratic latent variables in structural equation models can be classified into direct and indirect approaches. Direct estimation approaches produce structural coefficient estimates without introducing additional convenience variables to the model. Examples include multiple indicator specification (Kenny & Judd, 1984) (see Jaccard & Wan, 1990; Jöreskog &Yang, forthcoming) and single-indicator specification (see Ping, 1995). These specifications can be estimated directly with COSAN, LISREL 8, and similar software that accommodates nonlinear constraint equations.[1]

With indirect estimation approaches structural coefficient estimates may or may not be available, and convenience variables or several estimation steps are required. Examples include subgroup analysis (Jöreskog,1971a), convenience-variable techniques (Hayduk, 1987; Wong & Long, 1987), indirect multiple indicator specification (Ping, 1996), and indirect single-indicator specification (Ping, 1995).

[1] COSAN is available from SAS, Inc. Other software that will directly estimate latent variable interactions and quadratics includes LINCS (distributed by APTEC Systems), RAMONA (distributed by Michael W. Browne, Ohio State University), MECOSA (distributed by SLI-AG, Frauenfeld Switzerland).

These techniques can be used with EQS and the other structural equation software packages mentioned above.

Direct Estimation

Kenny and Judd (1984) proposed that latent variable interactions and quadratics could be specified adequately using all unique products of the indicators of their constituent linear latent variable. For the linear latent variables X and Z with indicators x_1, x_2, z_1, z_2, and z_3, respectively, the latent variable interaction XZ can be specified with the product-indicators x_1z_1, x_1z_2, x_1z_3, x_2z_1, x_2z_2, and x_2z_3. X*X can be specified with the product-indicators x_1x_1, x_1x_2, and x_2x_2. Under the Kenny and Judd normality assumptions,[2] the variance of the product-indicator x_1z_1, for example, is given by

$$\begin{aligned} \mathrm{Var}(x_1z_1) &= \mathrm{Var}[(8_{x1}X + ,_{x1})(8_{z1}Z + ,_{z1})] \\ &= 8_{x1}{}^2 8_{z1}{}^2\mathrm{Var}(XZ) + 8_{x1}{}^2\mathrm{Var}(X)\mathrm{Var}(,_{z1}) \\ &\qquad + 8_{z1}{}^2\mathrm{Var}(Z)\mathrm{Var}(,_{x1}) + \mathrm{Var}(,_{x1})\mathrm{Var}(,_{z1}) \;, \quad (1) \end{aligned}$$

where 8_{x1} and 8_{z1} are the loadings of the indicators x_1 and z_1 on the latent variables X and Z; $,_{x1}$ and $,_{z1}$ are the error terms for x_1 and z_1; Var(a) is the variance of a; and Cov(X,Z) is the covariance of X and Z. The variance of XZ is given by

$$\mathrm{Var}(XZ) = \mathrm{Var}(X)\mathrm{Var}(Z) + \mathrm{Cov}(X,Z)^2 \quad (1a)$$

(Kendall & Stuart, 1958).

In the quadratic case the variance of the product-indicator x_1x_1 is given by

$$\begin{aligned} \mathrm{Var}(x_1x_1) &= \mathrm{Var}[(8_{x1}X + ,_{x1})^2] \\ &= \mathrm{Var}[8_{x1}{}^2X^2 + 28_{x1}{}^2,_{x1} + ,_{x1}{}^2] \\ &= 8_{x1}{}^2 8_{x1}{}^2\mathrm{Var}(X^2) + 48_{x1}{}^2\mathrm{Var}(X)\mathrm{Var}(,_{x1}) + 2\mathrm{Var}(,_{x1})^2 \;. \quad (2) \end{aligned}$$

The variance of XX is given by

$$\mathrm{Var}(XX) = 2\mathrm{Var}(X)^2 \quad (2a)$$

using the equation (1a) result.

Kenny and Judd (1984) provided examples of a COSAN implementation of their technique, and examples of a LISREL 8 implementation of their technique are provided in Jaccard and Wan (1995) and

---

[2] The Kenny and Judd normality assumptions were that each of the latent variables X and Z is independent of the errors (ε_{x1}, ε_{x2}, ε_{z1}, and ε_{z2}), the error terms are mutually independent, the indicators x_1, x_2, z_1, and z_2 are multivariate normal, and the errors (ε_{x1}, ε_{x2}, ε_{z1}, and ε_{z2}) are multivariate normal.

Jöreskog and Yang (forthcoming). The implementation procedure involves centering the indicators for X, Z and Y;[3] [4] adding the product-indicators for XZ and/or XX to each case; specifying XZ and/or XX in the structural model with the appropriate product-indicators, constraining the loadings and error variances of these product-indicators to equal their equation (1) and (2) forms, constraining the variance of XZ and XX to their equation (1a) and (2a) forms, and subsequently estimating this model with COSAN, LISREL 8, or other software that accommodates nonlinear constraints. The loading of the product-indicator x_1z_1 for instance, would be constrained to equal $8_{x1}8_{z1}$ in equation (1), and the error variance of x_1z_1 would be constrained to equal $8_{x1}{}^2\text{Var(X)Var}(,_{z1}) + 8_{z1}{}^2\text{Var(Z)Var}(,_{x1}) + \text{Var}(,_{x1})\text{Var}(,_{z1})$ also in equation (1). Similarly the loading of x_1x_1 for example, would be constrained to equal $8_{x1}8_{x1}$ in equation (2), and the error variance of x_1x_1 would be constrained to equal $48_{x1}{}^2\text{Var(X)Var}(,_{x1}) + 2\text{Var}(,_{x1})^2$ also in equation (2).

This technique is powerful because it models several interaction and/or quadratic latent variables and provides coefficient estimates for these variables, but it has been infrequently used in the substantive literatures. This may have been because it is not widely known, and can be difficult to use (Aiken & West, 1991; see Jaccard & Wan, 1995; Jöreskog & Yang, forthcoming). Coding the constraint equations for the product-indicators in COSAN can be a daunting task. LISREL 8 provides a nonlinear constraint capability that is different from that available in COSAN, and the constraint coding effort is reduced. However, this task can still become tedious for larger models (Jöreskog & Yang, forthcoming; Ping, 1995). In addition for larger models, the size of the covariance matrix created by the addition of product-indicators, and the number of additional variables implied by the constraint equations, can create model convergence and other model estimation problems.

---

[3] Centering an indicator involves subtracting the mean of the indicator from each case value. As a result the indicator has a mean of zero (see Aiken & West, 1991; Bollen, 1989:13; Jaccard, Turrisi & Wan, 1990:28; Kenny & Judd, 1984). While this is a standard assumption in structural equation modeling for variables with an arbitrary zero point such as Likert-scaled and other rating-scaled variables (see Bollen, 1989), mean or zero centering has been the subject of much confusion over the interpretation of centered variables. Aiken and West (1991) present a compelling and exhaustive argument for the efficacy of centering.

[4] Centering the indicators of the dependent variable Y is optional and has no effect on the structural coefficient estimates. However it is recommended and used throughout this presentation to produce coefficient estimates that are equivalent to those produced when an intercept is specified (see Footnote 12).

As an alternative to the Kenny and Judd (1984) product-indicator specification, single-indicator specification has been proposed (Ping, 1995). The product-indicators of XZ and XX are replaced by single-indicators $x{:}z = [(x_1 + x_2)/2][z_1 + z_2 + z_3)/3]$ for XZ and $x{:}x = [(x_1+x_2)/2][(x_1+x_2)/2]$ for XX. The loadings and error variances for these single-indicators are similar in appearance to equations (1) and (2):

$$
\begin{aligned}
\mathrm{Var}(x{:}z) &= \mathrm{Var}\{[(x_1+x_2)/2][(z_1+z_2+z_3)/3] \\
&= \mathrm{Var}\{(x_1+x_2)/2\}\mathrm{Var}\{(z_1+z_2+z_3)/3\} + \mathrm{Cov}\{(x_1+x_2)/2,(z_1+z_2+z_3)/3\}^2 \\
&= [϶_X{}^2\mathrm{Var}(X)+2_X][϶_Z{}^2\mathrm{Var}(Z)+2_Z] + [϶_X϶_Z\mathrm{Cov}(X,Z)]^2 \\
&= ϶_X{}^2϶_Z{}^2[\mathrm{Var}(X)\mathrm{Var}(Z)+\mathrm{Cov}(X,Z)^2] + ϶_X{}^2\mathrm{Var}(X)2_Z \\
&\qquad + ϶_Z{}^2\mathrm{Var}(Z)2_X + 2_X2_Z \\
&= ϶_X{}^2϶_Z{}^2\mathrm{Var}(XZ) + ϶_X{}^2\mathrm{Var}(X)2_Z + ϶_Z{}^2\mathrm{Var}(Z)2_X + 2_X2_Z \quad (3)
\end{aligned}
$$

where $϶_X = (8_{x1} + 8_{x2})/2$, $2_X = (\mathrm{Var}(,_{x1}) + \mathrm{Var}(,_{x2}))/2^2$, $϶_Z = (8_{z1} + 8_{z2} + 8_{z3})/3$, $2_Z = (\mathrm{Var}(,_{z1}) + \mathrm{Var}(,_{z2}) + \mathrm{Var}(,_{z3}))/3^2$.

Similarly

$$
\begin{aligned}
\mathrm{Var}(x{:}x) &= \mathrm{Var}\{[(x_1+x_2)/2][(x_1+x_2)/2]\} \\
&= 2\mathrm{Var}\{(x_1 + x_2)/2\}^2 \\
&= 2\mathrm{Var}[(8_{x1}X+,_{x1})+(8_{x2}X+,_{x2})]^2/2^4 \\
&= 2[\mathrm{Var}((8_{x1}+8_{x2})X)+\mathrm{Var}(,_{x1})+\mathrm{Var}(,_{x2})]^2/2^4 \\
&= 2[\mathrm{Var}(϶_XX) + 2_X]^2/2^4 \\
&= ϶_X{}^4\mathrm{Var}(X^2) + 4϶_X{}^2\mathrm{Var}(X)2_X + 22_X{}^2\,. \quad (4)
\end{aligned}
$$

These results extend to latent variables with an arbitrary number of indicators.[5]

Estimation using single-indicator specification also involves centering the indicators for X, Z and Y; adding the single-indicators x:z and/or x:x to each case; specifying XZ and/or XX in the structural model using the single-indicators x:z and/or x:x, constraining the loadings and error variances of these single-indicators to equal their equation (3) and (4) forms, and constraining the variance of XZ and/or XX to their equation (1a) and (2a) forms; and subsequently estimating this structural model using COSAN, LISREL 8, or similar software that provides a nonlinear constraint capability. The loading of

[5] By induction, for an arbitrary latent variable X, $\Gamma_X = (\lambda_{x1} + \lambda_{x2} +...+ \lambda_{zm})/m$ and $2_X = (\mathrm{Var}(\varepsilon_{x1}) + \mathrm{Var}(\varepsilon_{x2}) +...+ \mathrm{Var}(\varepsilon_{xm}))/m^2$, where m is the number of indicators of X, and equations (3), and (4) apply to X and Z with an arbitrary number of indicators.

the single-indicator x:z for instance, would be constrained to equal э$_X$э$_Z$ in equation (3), and the error variance of x:z would be constrained to equal э$_X^2$Var(X)2$_Z$ + э$_Z^2$Var(Z)2$_X$ + 2$_X$2$_Z$ also in equation (3). Similarly the loading of x:x for example, would be constrained to equal э$_X$э$_X$ in equation (4), and the error variance of x:x would be constrained to equal 4э$_X^2$Var(X)2$_X$ + 22$_X^2$ also in equation (4). An example using this technique is provided later in the paper.

Single-indicator specification appears to be equivalent to the Kenny and Judd (1984) specification (Ping, 1995); it has the power of the Kenny and Judd (1984) technique because it can model several interaction and/or quadratic variables and provide coefficient estimates for these variables. In addition it requires less specification effort than the Kenny and Judd (1984) approach, and produces an input covariance matrix with fewer elements.[6] However it is new and has yet to appear in the substantive literatures.

Indirect Estimation

Subgroup Analysis Subgroup analysis (Jöreskog, 1971a) generally involves splitting the sample and assessing differences in model fit when the model is restricted to the resulting groups of cases. The procedure involves dividing a sample into subgroups of cases based on different levels (e.g., low and high) of a suspected interaction X. Constraining the model coefficients to be equal between subgroups for the model estimated in each subgroup, the coefficients of the linear-terms-only model (e.g., the model without any interaction or quadratic latent variables present) are then estimated for this model using each of the resulting subgroups and structural equation analysis. The result is a Π^2 statistic for the model's fit across the two subgroups with this coefficient equality constraint across the subgroups. Relaxing this equality constraint, the model is re-estimated, and the resulting Π^2 statistic is compared with that from the first estimation. A significant difference between the Π^2 statistics for these two nested models suggests that there is at least one coefficient difference between the two groups. The coefficients from the second estimation are then tested for significant differences between the groups using a coefficient difference test (see for instance Jaccard, Turissi & Wan, 1990:49). A significant coefficient difference

[6] Each additional product-indicator adds an input variable, and a row and column, to the sample covariance matrix. Adding product-indicators can become statistically detrimental in that each additional product-indicator places additional demands on the sample covariance matrix: the number of resulting variables can become too large to yield a reasonably stable matrix (Jaccard & Wan, 1995).

between the Z coefficients suggests an interaction between that variable and X, the variable used to create the subgroups. A significant coefficient difference between the X coefficients suggests the presence of a quadratic.

This technique is popular in the substantive literatures, and is a preferred technique in some situations. Jaccard, Turissi and Wan (1990) state that subgroup analysis may be appropriate when the model could be structurally different for different subgroups of subjects. They also point out that an interaction need not be of the form "X times Z" (interaction forms include X/Z, and the possibilities are infinite (Jaccard, Turissi & Wan, 1990) (see Hanushek & Jackson, 1977)), and that three group analysis may be more appropriate in these cases. Sharma, Durand and Gur-Arie (1981) recommend subgroup analysis to detect what they term a homologizer: a variable W, for example, that affects the strength of an association between two variables, X and Y, yet is not related to either X or Y. However, subgroup analysis is criticized in the regression literature for its reduction of statistical power and increased likelihood of Type II error (Cohen & Cohen, 1983; Jaccard, Turissi & Wan, 1990) (see also Bagozzi, 1992). In addition, coefficient estimates for significant interactions or quadratics are not available in subgroup analysis, and interpretation of a significant interaction or quadratic is problematic. Sample size requirements also limit the utility of subgroup analysis. Samples of 100 cases per subgroup are considered by many to be the minimum sample size, and 200 cases per group are usually recommended (Boomsma, 1983) (see Gerbing & Anderson (1985) for an alternative view).

Convenience Variables Hayduk (1987) and Long and Wong (1987) suggested an approach involving the addition of convenience variables to the structural model in order to accomplish the estimation of an interaction or quadratic effect. Hayduk (1987) for instance, suggested specifying the loading of x_1z_1 on XZ by creating an intervening latent variable between XZ and x_1z_1 that has an error of zero. The indirect effect of XZ on x_1z_1 via this variable could then be used to specify the loading of x_1z_1 on XZ. The error term for x_1z_1 is handled similarly.

These convenience variable approaches are powerful because they can specify multiple interaction and quadratic variables, and provide coefficient estimates for interactions and quadratics. However the effort required to specify a model using these techniques can be considerable (see Hayduk, 1987:Chapter 7). Perhaps as a result they are infrequently used in the substantive literatures.

Two-Step Approaches Two-step estimation techniques involve calculating the loadings and error variance for the (product- or single-) indicator(s) of XZ using measurement model parameter estimates, then fixing these loadings and error variances at their calculated values in the structural model.

Estimates of the parameters comprising the loadings and error variances of the indicator(s) of latent variable interactions and quadratics are available in a linear-terms-only measurement model corresponding to the structural model of interest (i.e., a measurement model that includes the linear latent variables but excludes the interactions and quadratics). With sufficient unidimensionality, that is the indicators of each construct have only one underlying construct each (Aaker & Bagozzi, 1979; Anderson & Gerbing, 1988; Burt, 1973; Hattie, 1985; Jöreskog, 1970, 1971b; McDonald, 1981), these measurement model parameter estimates change trivially, if at all, between the measurement model and alternative structural models (Anderson & Gerbing, 1988). As a result, instead of specifying the loadings and error variances for the indicators of XZ or XX as variables in the structural model, they can be calculated using parameter estimates from the linear-terms-only measurement model (i.e., involving only X, Z and Y), and specified as constants in the structural model if X and Z are each sufficiently unidimensional. This is possible because the unidimensionality of X and Z permits the omission of the XZ and XX latent variables from the linear-terms-only measurement model: because X and Z are each unidimensional, their indicator loadings and error variances are unaffected by the presence or absence of other latent variables in a measurement or structural model, in particular XZ and XX.

To use this technique X, Z and Y are unidimensionalized,[7] X, Z and Y are centered, and the (product- or single-) indicator(s) for XZ and/or XX are added to each case as before. Then the equation (1), (1a), (2), (2a), (3) and/or (4) parameters (i.e., 8_{x1}, 8_{x2}, 8_{z1}, 8_{z2}, 8_{z3}, Var($,_{x1}$), Var($,_{x2}$), Var($,_{z1}$), Var($,_{z2}$), Var($,_{z3}$), Var(X), Var(Z), and Cov(X,Z)) are estimated in a linear-terms-only measurement model (e.g., one that excludes XX and XZ). Then the loadings and error variances for the indicators of XZ and/or XX, plus the variances of XZ and/or XX are calculated by substituting these measurement model parameter estimates into the appropriate versions of equations (1) through (4).[8] Finally, using a structural

---

[7] Procedures for obtaining unidimensionality are suggested in Anderson and Gerbing (1982:454), Gerbing and Anderson (1988), and Jöreskog (1993:313).

[8] Ping (1996) showed that in general the loading of a product-indicator xz is given by $\lambda_x\lambda_z$ and the error variance is given by $K\lambda_x^2$Var(X)Var(□ε_z) + $K\lambda_z^2$Var(Z)Var(□ε_x) + KVar(□x)Var(□z) (K=2 if x=z, K=1 otherwise). For the single-indicator technique Γ and □ in this equation are replaced by □ and 2 defined at equation (3).

model in which these calculated loadings and error variances for the indicators of XZ and/or XX, plus the variance of XZ and/or XX, are specified as constants (i.e., fixed rather than free), the structural model is estimated using EQS or LISREL.[9] An example of this technique is provided in the next section.

If the structural model estimates of the linear latent variable measurement parameters (e.g., 8_{x1}, 8_{x2}, 8_{z1}, 8_{z2}, 8_{z3}, Var($,_{x1}$), Var($,_{x2}$), Var($,_{z1}$), Var($,_{z2}$), Var($,_{z3}$), Var(X), Var(Z), and Cov(X,Z)) are not similar to the linear-terms-only measurement model estimates (i.e., equal in the first two decimal places) and the calculated values fixed in the specification of XZ and/or XX in the structural model therefore, do not adequately reflect the structural model measurement parameter values, the equations (1) and (1a), (2) and (2a), (3) or (4) values can be re-computed using the structural model estimates of these measurement parameters. The structural model can then be re-estimated with these updated fixed values. Zero to two of these re-estimations are usually sufficient to produce consecutive structural models with the desired trivial difference in measurement parameters and coefficient estimates that are equivalent to direct LISREL 8 or COSAN estimates.

Two-step techniques appear to be equivalent to direct estimation (Ping, 1995; 1996). They have the power of the Kenny and Judd (1984) technique because they can model several interaction and/or quadratic variables and provide coefficient estimates for these variables. In addition they require less specification effort than the direct approaches, and with a single-indicator produce a sample covariance matrix with fewer elements than the Kenny and Judd (1984) approach. However, two-step techniques are also new.

EXAMPLES

In the Marketing literature Ping (1993) reported that relationship neglect (reduced physical contact) (NEG) in a buyer-seller relationship (i.e., an economic plus social exchange relationship) had antecedents that included overall relationship satisfaction (SAT), alternative attractiveness (ALT), and relationship investment (INV) (see also Rusbult, Zembrodt & Gunn, 1982). However an hypothesized switching cost (SCT) association with NEG was not significant.

---

[9] LISREL 8 is available for microcomputers only, and according to individuals at SSI and SPSS when this paper was written (December, 1995), there are no plans to release a mainframe version of LISREL 8 in the near future. As a result mainframe LISREL 7 is likely to remain in use. The two-step technique can also be used with LISREL 8 in problem situations (see page 14).

Recalling the previous remarks concerning the role of population interactions or quadratics in disconfirmed theory test results, the nonsignificant SCT-NEG effect may have been due to an unmodeled interaction or quadratic present in the population model. Because a quadratic SCT- NEG effect and several SCT interaction effects were plausible, the SCT*SCT quadratic, and the SCT interactions with the other antecedents were estimated using direct estimation, a single-indicator per interaction or quadratic latent variable, and the following structural model,

$$NEG = \exists_1 SAT + \exists_2 ALT + \exists_3 INV + \exists_4 SCT + \exists_5 SAT{*}SCT + \exists_6 ALT{*}SCT + \exists_7 INV{*}SCT + \exists_8 SCT{*}SCT + ..$$

(5

The LISREL 8 program code for this specification is shown in Figure 1, and the results are shown in Table1.

The interactions of SCT with SAT and INV were significant when all the interactions and the quadratic latent variable involving SCT were jointly specified (see Table1). When these interactions and the quadratic were estimated singly, however, only the INV*SCT interaction was significant. The results for equation (5) re-specified with only the significant INV*SCT interaction, i.e.,

$$NEG = \exists_1 SAT + \exists_2 ALT + \exists_3 INV + \exists_4 SCT + \exists_7 INV{*}SCT + . , \quad (6$$

are shown in Table 2, and the EQS program code for this specification is shown in Figure 2. The significance of the INV*SCT interaction effect on NEG is shown in Table 3.

Discussion

The LISREL 8 coefficient estimates for equation (5) shown in Table1 were produced following the direct estimation procedure described above. Each indicator for SAT, for instance, was centered by subtracting the indicator's average from its value in each case. The values for the single-indicators of the four interactions and the quadratic were added to each case. The single-indicator value for SAT*INV for instance, was computed in each case using sat:neg = $[(sa_2 + sa_4 + sa_5 + sa_6 + sa_7)/5][(in_1 + in_3 + in_4 + in_5)/4]$. Next the structural model was specified using PAR variables (Jöreskog & Sörbom, 1993:14), constraint equations (Jöreskog & Sörbom, 1993:11) for the loadings, errors, and variances for the interactions and the quadratic, and latent variable metrics assigned using unit variances for the exogenous variables (see Figure 1 and Jöreskog & Sörbom, 1993:7); and the structural model was estimated using maximum likelihood. The use of PAR variables in this manner is sensitive to the sequence and location of the PAR and constraint (CO) statements in the program. In general PAR's should not be used recursively (Jöreskog & Sörbom, 1993). In this application they should appear at the end of the program, just before the output (OU) card. In addition the PAR variables and the variables constrained in the CO statements (e.g., lx(18,5) in Figure 1) should be defined in their natural numerical order (e.g., PAR(1), PAR(2), etc., not PAR(2), PAR(1); lx(18,5), lx(19,6), etc. not vice versa), and a PAR variable should be used in a CO statement as soon as possible after it is defined (see Figure 1).

The EQS estimates for equation (6) shown in Table 2 were produced using the procedure for two-step estimation described above. SAT, ALT, INV, SCT and NEG were unidimensionalized. In this application unidimensionality was conceptualized as: the indicators of each construct have only one underlying construct each. While there have been many proposals for developing unidimensional constructs (see Footnote 7), for this example unidimensionality was established as follows. Unidimensionality was operationalized as i) for each construct the Π^2 p-value resulting from the single construct measurement model (see Jöreskog, 1993) for each construct is .01 or greater; and ii) for the linear-terms-only measurement model (i.e., containing only the linear constructs) not rejecting H_0: The root mean square error of approximation < .05 (i.e., its p-value > .05), and a comparative fit index of .95 or larger. The measurement model for SAT, for instance, was estimated using all of its items, then re-estimated until a target Π^2 p-value .01 or greater was attained by serially deleting items that did not appear to degrade content validity (not reported). This is obviously an area where judgement was required, and considerable care was taken to avoid sacrificing content validity for low Π^2. Because several itemizations for each construct were acceptable under criteria (i), judges were used to evaluate the content validity of each of these itemizations and select the final itemization for each construct. A linear-terms-only measurement model involving the final itemization for each construct suggested the constructs were unidimensional using the step (ii) criteria (see Table 4).

Next the indicators for SAT, ALT, INV, SCT and NEG were centered, and the single-indicator inv:sct = $[(in_1 + in_3 + in_4 + in_5)/4][(sc_2 + sc_3 + sc_4 + sc_5)/4]$ was added to each case. Using the linear-terms-only measurement model parameter estimates (see Table 4) the loading and error variance of inv:sct plus the variance of INV:SCT were calculated using equations (3) and (1a). The structural model then was specified with the loading of inv:sct and its error variance fixed at the values calculated with equation (3), and the variance of INV*SCT fixed at the equation (1a) value. Specifically, the loading of inv:sct was fixed at $\ni_{INV}\ni_{SCT}$ = .9280, the error variance of inv:sct was fixed at $\ni_X^2 Var(X) 2_Z$ +

$э_Z^2$Var(Z)2_X + $2_X 2_Z$ = . 0573, and the variance of INV*SCT was fixed at Var(INV)Var(SCT) + Cov(INV,SCT)^2 = .7532 (see Figure 2). Finally the structural model was estimated using maximum likelihood estimation and EQS.

Indicator Nonnormality While the measurement parameter, structural disturbance, and coefficient estimates (e.g., 8's, 2's, N's, P's, ('s, and ∃'s) produced by maximum likelihood are robust to departures from normality (Anderson & Amemiya, 1985, 1986; Bollen, 1989; Boomsma, 1983; Browne, 1987; Harlow, 1985; Jöreskog & Sörbom, 1989; Sharma, Durvasula & Dillon, 1989; Tanaka, 1984), the model fit and significance statistics (i.e., Π^2 and standard errors) may not be (Bentler, 1989; Bollen, 1989; Jöreskog & Sörbom, 1989) (see Jaccard & Wan (1995) for evidence to the contrary). Because product- and single-indicators are per se nonnormal, and model fit and significance statistics from the maximum likelihood-robust estimator (Satorra & Bentler, 1988) appear to be less sensitive to departures from normality (see Hu, Bentler & Kano, 1992), maximum likelihood-robust estimates were added to the Table 2 results.[10] However, comparing the standard error and Π^2 estimates for maximum likelihood and maximum likelihood-robust, they were generally similar. The limited evidence to date (Jaccard & Wan, 1995; Ping, 1995) suggests that for most practical purposes (Jöreskog & Yang, forthcoming) model fit and significance statistics from maximum likelihood estimation may be sufficiently robust to the addition of a few product-indicators that involve linear indicators that are normal. However, the robustness of these statistics from this estimator to the addition of many product-indicators (i.e., over four) or product-indicators comprised of nonnormal linear indicators typical of survey data is unknown.

Re-estimation Comparing the Table 2 and 4 measurement parameter estimates, they are similar. Had they been dissimilar (i.e., different in the second decimal place) the re-estimation process could have been used. Using the re-estimation technique, the requirement for strict unidimensionality in the linear latent variables can be relaxed somewhat, although the practical limits of how different the measurement parameters can be between the linear-terms-only measurement model and the structural model in order to produce stable measurement and structural coefficient estimates is unknown.

*Interpreting INV*SCT* Table 3 provides information regarding the contingent nature of the SCT

---

[10] Asymptotic Distribution Free estimates (WLS and DWLS methods in LISREL and the AGLS method in EQS) appear to be inappropriate for small samples (i.e., fewer than 400 cases) (Aiken & West, 1991; Hu, Bentler & Kano, 1992; Jaccard & Wan, 1995) (see also Jöreskog & Yang, forthcoming).

relationship with NEG (and the INV relationship with NEG). The significance of the SCT coefficient (i.e., .013-.111INV) varies with INV: the size of the coefficient and its standard error depend on the level of INV. In addition the standard error of the coefficient of SCT involves the variance and covariance of the INV and SCT*INV coefficients. Table 3 also demonstrates the effect of an interaction. When INV is at its study average, the SCT coefficient was small and nonsignificant. For smaller values of INV it was positive and approached significance. As this example suggests had the interaction been significant at both ends of the range of INV, offering plausible explanations for disordinal interactions can be challenging (see Aiken & West, 1991).[11]

Intercepts In realistic social science research situations with centered indicators, the omission of an intercept term in a structural equation with an interaction or quadratic biases the resulting coefficients slightly (see Jöreskog & Yang, forthcoming).[12] In these situations this bias is typically in the third or fourth decimal place. For instance the Table 2 results with an intercept were

NEG =	.043 -	.363SAT +	.163ALT -	.173INV -	.013SCT -	.116INV*SCT.[13]
t =	1.06	-4.68	2.69	-2.41	.25	-2.26

As a result unless mean structures are of interest it is usually unnecessary to estimate intercepts with latent variable interactions and quadratics, and the Table 3 results used the Table 2 coefficient and standard error estimates, which assumed a zero intercept.

---

[11] An XZ interaction can be ordinal or disordinal (Lubin, 1961) . For an ordinal interaction the Z-Y association becomes weaker over the range of the interacting variable X. A disordinal interaction, however, is characterized by the Z-Y association changing signs over the range of the interacting variable X.

[12] As Jöreskog and Yang (forthcoming) point out, a structural equation containing an interaction should in general be specified with an intercept term to avoid interpretational difficulties. Neglecting to do so biases structural coefficients and their standard errors, and produces an interaction coefficient that represents a centered interaction, which may be difficult to interpret. However, the specification of an intercept along with interactions and quadratics can create estimation difficulties, and limit the accessability of estimators that are less sensitive to departures from normality. Because no-intercept bias is substantially reduced in realistic social science research situations by centering the indicators of the independent and dependent latent variables, this presentation centers all indicators and reports estimation results that typically omit the specification of intercepts.

[13] These estimates were produced by adding METHOD=MOMENT to the /SPECIFICATIONS section of Figure 2, and adding a *V999 variable to the equation for F5.

However, for structural coefficients in a neighborhood of twice their maximum likelihood standard error (e.g., between |t| = 1.85 and 2.15), intercept-influenced coefficients probably should be estimated to avoid Type I and II errors. Because modeling intercepts can produce model identification, convergence and improper solution problems (see Bentler, 1989), and adding interactions and quadratics frequently aggravate these problems, an intercept model containing interactions and quadratics should be estimated in two steps. First the interaction and quadratic model should be estimated as described above (i.e., without an intercept). Then the model should be reestimated with the intercept(s) specified, using starting values from the no-intercept model and OLS regression estimates of the intercept(s). If estimation problems are encountered in the second step, one or more structural coefficients for linear latent variables not involved in the interactions and quadratics could be fixed at their no-intercept values to obtain starting values for the other structural coefficients and the intercept. For larger models, however, it may be impossible to estimate an intercept model with interactions, quadratics, and centered linear latent variables. For unknown reasons the equation (5) model with an intercept would not converge using any of the techniques discussed above. OLS regression results for equation (5) with an intercept and those for regression-through-the-origin (which is equivalent to the equation (5) model) were

NEG = .048 - .290SAT + .137ALT - .170INV - .004SCT + .158SAT*SCT - .004ALT*SCT - .133INV*SCT -.022SCT*SCT

and

NEG = - .299SAT + .137ALT - .168INV - .001SCT + .160SAT*SCT - .002ALT*SCT - .132INV*SCT +.005SCT*SCT .

Comparing these coefficient estimates suggests that any intercept bias in the Figure 1 results may be small.

Alternatives In the first example two-step estimation with single-indicators could have been used instead of direct estimation. The estimation procedure would have been the same as in the second example except that the computed starting values for the loadings and error variances of the interactions and the quadratic (plus their variances if the diagonal N's had not equaled 1) would have been fixed. Both direct and two-step estimation using single-indicators are useful for probing for the existence of interactions or quadratics because they are relatively easy to specify, and they both produce sample covariance matrices with fewer elements than the Kenny and Judd (1984) approach. The single-indicator direct approach works well in LISREL 8 for smaller models, but because the PAR's are used recursively, larger models can misbehave. For instance the standard errors and significance for several interaction loadings and errors did not print in the first example. Eliminating the single-indicator PAR's by replacing them with their expanded equivalents does not usually execute in larger models such as the first example.

Similarly the second example could have used product-indicators with either direct or indirect estimation. Specification effort would have been higher than for a single-indicator if the full set of product-indicators were used for each nonlinear latent variable (to avoid concerns about their content validity-- discussed later). For the calculations involved in fixed and/or starting values for product- or single-indicators an Excel or Lotus spreadsheet is useful. An Excel spreadsheet was used for starting values in both examples, and for the fixed single-indicator loading and error variance, plus the fixed variance of INV*SCT, in the second example (see Table 5). The linear-terms-only measurement model loadings, error variances, and variances and covariances for the linear latent variables involved in the interactions and quadratic (e.g., SAT, ALT, INV, and SCT) (see Table 4) were keyed into the spreadsheet, and the product-indicators' loadings and error variances, plus the interaction and quadratic variances, were calculated using equations (1) through (4).

CONVERGENCE AND PROPER ESTIMATES

These techniques can produce their share of convergence and improper solution headaches. As with any structural equation model solution, the output should be examined for negative squared multiple correlations, linearly dependent parameter estimates, parameter estimates constrained at zero, etc. When convergence or improper solution difficulties are encountered using the techniques discussed above, the first step should be to verify that the indicators of the linear latent variables were centered.

User-specified starting values for the latent variable variances and covariances, the structural coefficients (i.e., the ('s and ∃'s), and the variances of the structural disturbances (i.e., Var(.)'s) are frequently required. Error-attenuated variance and covariance estimates from SAS or SPSS, and OLS regression coefficient estimates are frequently sufficient to solve convergence and improper solution problems. The estimated disturbance term variance for Y, for instance, can be calculated using Var(.) = Var(Y)(1-R^2), where Var(Y) is the error-attenuated variance of Y (e.g., from SAS or SPSS) and R^2 is the OLS regression estimate of the explained variance for Y regressed on the summated linear (e.g., $(x_1+x_2)/2$), interaction, and quadratic variables (e.g., $[(x_1+x_2)/2][(z_1+z_2+z_3)/3]$) involved in the structural equation model.

Occasionally disattenuated variance and covariance estimates are required. The adjusted variance and covariance estimates for the independent variables can be calculated using attenuated variances and covariances (i.e., SAS or SPSS estimates), linear-terms-only measurement model estimates, and the calculations shown in the Appendix. As an alternative a measurement model involving all the latent variables in the structural model of interest (e.g., X, Z, Y, XZ, and XX) should also provide useful disattenuated variance and covariance estimates.

If problems persist, constraining the variance of the structural disturbance terms (i.e., . in equations 5 or 6) to more than 10% of the variance of their respective endogenous variables may be effective (few models in the social sciences explain 90% of the variance of an endogenous variable). This would be accomplished in the first example by constraining PSI(1), and in the second example by constraining d1.

In addition, scaling the exogenous variables by setting their variance to 1 (see Jöreskog & Sörbom, 1993:7) is sometimes useful. This approach was used in the first example: the loadings of SAT, ALT, INV, and SCT were all freed, and their variances plus the variances of the interactions and the quadratic were fixed at 1 (see Figure 1).

The equation (1a) and (2a) constraints on the variances of the interactions and quadratics could also be relaxed in problem situations. This would be accomplished in the first example by deleting the constraint equations for the variances of the interactions and quadratic. In the second example the fixed variances for the interaction would be freed by adding a *. The resulting interaction or quadratic coefficient estimates are typically attenuated and closer to their OLS regression estimates.

In addition fixing some of the covariances among the exogenous variables at zero is occasionally necessary. In particular zeroing the covariances between the linear latent variables (e.g., X and Z) and the interactions/quadratics (e.g., XZ), and/or zeroing the covariances among the interactions/quadratics (e.g., between XZ and XX) may be required. This was done in the first example to compensate for the multi-collinearity among the interactions and the quadratic.

If problems continue to persist, the model may not be identified. Otherwise identified models can become nonidentified with the specification of correlated errors or nonrecursive relationships (see Berry, 1984). More often with interactions and quadratics, otherwise identified models can be empirically underidentified, or weakly identified (see Hayduk, 1987). Berry (1984), Bollen (1989:251), and Hayduk (1987:139) provide accessible discussions of the sources of lack of identification and identification checking.

NEEDED RESEARCH

Although the above techniques provide considerable improvement over regression coefficient estimates for theory tests involving unobserved interaction and quadratic variables, additional work is needed on the specification, estimation, and interpretation of these variables using structural equation analysis. The following is an incomplete enumeration of areas where additional research on these matters might be useful, in no particular order of importance.

Specification

The number of product-indicators in a model with interactions and quadratics involving over-identified constituent latent variables can become large. Equation (5) for instance would have required 62 product-indicators. As mentioned earlier, specifying many product-indicators can add to execution times, convergence, and improper solution problems. However, a reduced number of product-indicators may adequately specify a latent variable interaction or quadratic. Jaccard and Wan (1995), for instance, used a subset of four product-indicators.

It could be argued, however, that concern for the content validity of a latent variable interaction or quadratic requires the use of all its product-indicators. All of the indicators of the interaction and quadratic latent variables were used in the examples presented above because it was not clear which product-indicators could have been safely dropped without impairing the content validity of interactions and the quadratic. As a result, it would be useful to know the conditions under which product-indicators could safely be dropped without impairing the content validity of the resulting interaction or quadratic latent variable, or to have guidelines for this endeavor.

In an investigation of the detection spurious interactions in regression, Lubinski and Humphreys (1990) noted that interactions and quadratics are correlated. As the correlation between X and Z for instance approaches 1, the correlation between XX (or ZZ) and XZ also approaches 1. As a result, they argued a population quadratic can be mistaken for an interaction in regression. Consequently they proposed that quadratic combinations of the linear latent variables comprising a latent variable interaction should be entered in the regression model along with the interaction of interest.

It is plausible that these results may extend to structural equation analysis. In the first example the quadratic was included because it was a second-order latent variable involving switching cost.

Lubinski and Humphreys' (1990) results suggest that it should have been included also because a significant INV*SCT interaction could be induced by significant SCT*SCT or INV*INV quadratics in the population equation. Consequently it would be helpful to know if latent variable interactions and quadratics can be mistaken for each other in structural equation analysis, and if so, what remediation steps would be appropriate.

Kenny and Judd (1984) proposed that, under their normality assumptions, the variances of latent variable interactions and quadratics should be constrained to their respective equation (1a) and (2a) forms, which are based on Kendall and Stuart's (1958) results. Kendall and Stuart also showed under these conditions that interactions and quadratics are associated with each other and linear latent variables in a predictable manner (e.g., Cov(XZ,XX) = 2Var(x)Cov(X,Z), Cov(XZ,X) = 0, etc.). However these constraints were not specified in Kenny and Judd (1984), and they have not been specified subsequently (see Jaccard & Wan, 1995; Jöreskog & Yang, forthcoming; Ping, 1995; 1996). For instance in the first example, the covariances of the interactions with each other and the quadratic were not constrained, nor were their covariances with the linear latent variables constrained. While Jaccard and Wan's (1995) results suggest that the omission of Cov(XZ,*) constraints, where * is a linear latent variable, may not materially affect interaction coefficients, their study may not have been designed to investigate this matter. As a result, it would be interesting to know what effect, if any, the omission of Cov(XZ,*) constraints, and Cov(XZ,**) constraints (when an interaction and a quadratic, and/or multiple interactions and/or quadratics are jointly specified) has on the resulting interaction and quadratic coefficients.

It is not obvious how a nonrecursive latent variable interaction or quadratic should be estimated. A nonrecursive interaction or quadratic specification may be appropriate when a hypothesized feedback relationship is not linear in one or both directions. This situation is plausible in the examples presented above. It could be argued that investment in a relationship should reduce relationship neglect, and neglect should also reduce investment. Assuming the neglect-investment relationship would still be moderated by switching cost (see Table 3) in a nonrecursive specification of this relationship, and recalling the requirements for identification and instrumental variables that are not correlated with their indirect endogenous variables in these specifications (see Berry, 1984), it would be useful to know how to specify this relationship in a structural equation model using a minimum of additional nonrecursive paths and instrumental variables.

It would also be helpful to have some guidance regarding the estimation of mixed formative and reflexive models involving interactions and quadratics. While adequate techniques such as PLS (Lohmöller, 1981) exist for estimating these models (see Fornell & Bookstein, 1982), substantive researchers frequently use OLS regression when their structural model contains a mixture of formative and reflexive variables (see for instance Heide & John, 1990) (however, see Bristor, 1993). As a result at least some of the structural coefficient estimates are biased and inefficient (Bohrnstedt & Carter, 1971). Hence it would be useful to know how to specify formative interactions and quadratics, and mixed formative-reflexive interactions.

Estimation

A hierarchical procedure for sequentially adding interactions and quadratics to a model such as that used in ANOVA or hierarchical regression analysis would also be useful to avoid interaction or quadratic latent variables that are significant but explain little additional variance and are therefore of little substantive value. The second example is a case in point. An equation (6) model that excluded the INV*SCT interaction, explained 35.6 percent of the variance of NEG (not reported). This is slightly more than one percentage point less than the multiple squared correlation for NEG shown in Table 2 with INV*SCT specified, which would be nonsignificant in a hierarchical regression analysis.

Research design could affect the detection of a latent variable interaction or quadratic. In an exploration of the difficulties of detecting interactions using survey data and regression, McClelland and Judd (1993) showed that because field studies are similar to an experiment with unequal cell sizes, field studies are generally less efficient than experiments in detecting interactions (see also Stone-Romero, Alliger and Aguinis, 1994). They concluded an optimal experimental design for detecting interactions exhibits an independent variable distribution that is polar (i.e., has many cases containing extreme independent variable values) and balanced (i.e., has equal cell sizes). The most efficient of these distributions McClelland and Judd (1993) characterized as a "four-cornered" data distribution (which has, for two independent variables with more than two levels, a three-dimensional frequency distribution that looks like the four legs on an upside-down kitchen table), and an "X-model" (which has, for two independent variables with more than two levels, a three-dimensional frequency distribution that resembles a bas-relief X anchored on the four polar cells).

Because field studies in the social sciences typically produce censored mound-shaped distributions for independent variables instead of uniform (balanced), four-cornered, or X distributions, they are not usually as efficient as experiments in detecting interactions. Comparing an experiment with two independent variables and a four-cornered data distribution to the equivalent mound-shaped field study distribution, McClelland and Judd (1993) argued that the field study would produce a data distribution that is 90% to 94% less efficient in detecting interactions as the four-cornered distribution.

Since it is plausible that their results extend to structural equation analysis, it would be helpful to have guidelines for non-experimental research designs that would produce a high likelihood of detecting an hypothesized population interaction or quadratic. For instance McClelland and Judd (1993) suggested over-sampling the extremes or poles of the scales in such studies. Based on their results, a stratified sample that produces a uniform distribution for two independent variables increases the efficiency of detecting an interaction between these two variables using regression by a factor of between 2.5 and 4. A field study that samples only the poles, such as Dwyer and Oh's (1987) study of output sector munificence effects in marketing channels, improves the efficiency of interaction detection using regression by 1250% to 1666% (although they did not test for interactions).

It would also be useful to have suggestions regarding analytical technique refinements that would increase the likelihood of detecting an hypothesized population interaction or quadratic using a field-survey design. Possibilities might include a case-weighting approach that emphasizes the polar cases in a set of responses, so that a more-nearly uniform or polar distribution would be produced.

There may be other data-related factors that affect the detection of interactions and quadratics in structural equation analysis. For instance, correlated (systematic) error between the independent and dependent variables attenuates the observed coefficient sizes of interactions in regression (Evans, 1985). In the examples presented above it was assumed that the effects of these errors were adequately modeled with uncorrelated indicator error terms. While the techniques discussed earlier can be extended to models involving correlated errors for the linear latent variables (see Ping, 1995; 1996), it would be helpful to have guidance for any implications this has for correlations among the error terms for product-indicators.

Interpretation

As mentioned earlier, product- and single-indicators are not normal even of their constituent variables are normal. However the evidence to date (Jaccard & Wan, 1995; Kenny & Judd, 1984; Ping, 1995; 1996), including the results presented in the second example, suggest the addition of a few of these indicators does not materially bias standard errors or the chi squared statistic of the resulting structural model using maximum likelihood. As a result, it would be useful to know the limits of these results. This may be important to substantive researchers because, when compared to maximum likelihood, estimators that are less dependent on distributional assumptions are practically unknown to them. While studies involving Monte Carlo analyses with realistic research situations would be valuable (see for instance Jaccard & Wan, 1995), it would be interesting to see results derived from bootstrap techniques (however, see Bollen & Stine, 1993).

The techniques discussed in this paper and elsewhere (e.g., Aiken & West, 1991; Jaccard, Turissi & Wan, 1990) have been applied exclusively to exogenous variable interactions and quadratics. While these techniques may extend with little or no modification to endogenous interactions and quadratics, it is not clear what a dependent interaction represents. Such a situation could arise in the examples presented above. It is plausible that satisfaction is an antecedent of both investment and switching cost: as overall relationship satisfaction increases, investments in the relationship should increase, and the perception of the difficulty (cost) of switching relationships should also increase. However, if INV and SCT still combine (interact) in their relationship with NEG with a SAT antecedent, it is not clear how to conceptualize or interpret the SAT-to-NEG via INV, SCT and INV*SCT relationship.

While model fit assessment is a controversial area (see Bollen & Long, 1993), guidelines for model fit when interactions and quadratics are specified would be useful. It has been my experience with field survey data and simulated data closely mimicking field survey data that as the number of specified product-indicators for a significant interaction or quadratic increases, some model fit indices suggest model-data fit is improved while others suggest the opposite. The addition of a single-indicator for a significant interaction or quadratic also appears to produce conflicting model-to-data fit results, even when compared to a misspecified model that excludes a population interaction or quadratic.

It may also be helpful to revisit the reporting and interpretation of interactions and quadratics from a theory testing perspective. While Aiken and West (1991) provide an accessible treatment of this topic for interactions, there is no equivalent treatment of quadratics. In addition, the fact that the SCT effect on NEG for instance depends on the range and mean of INV, and each of these statistics has a confidence interval, suggests that the Table 3 presentation may be simplistic for conclusions regarding a theory test. It could be argued that in other contexts the significant disordinal interaction may be ordinal, and the SCT effect could be positive or negative over the range of INV. To explain, in a different study the range of the interacting variable INV could be different, the sample could produce a different mean for INV, and the observed interaction could be ordinal and positive, ordinal and negative, or disordinal and both. Hence the most that might be concluded from the second example is there is an interaction between INV and SCT in their association with NEG, and the SCT-NEG effect could be positive, negative or both over the range of INV.

Standardized structural coefficients are used in some disciplines (e.g., Marketing) to compare the relative impact of latent variables with significant coefficients. However, standardized regression coefficients for interactions and quadratics are not invariant to centering (see Aiken & West 1991), and comparisons among standardized coefficients in models involving interactions and quadratics that also involve centered linear variables may be misleading. Friedrich (1982) proposed using Z-scores to produce standardized coefficients in regression involving interactions (see Aiken & West, 1991), and it would be useful to have equivalent results for interaction and quadratic latent variables.

Finally, it may be helpful to revisit the interpretation of an interaction when the coefficient for the linear variable is nonsignificant. The current practice is to interpret the first derivative of the combination of the nonsignificant linear variable coefficient and the significant interaction variable as shown in Table 3 (Aiken & West, 1991; Jaccard, Turissi & Wan, 1990). It could be argued, however, that since the linear variable coefficient is nonsignificant, only the contribution of the interaction variable should be interpreted. For instance in Table 3, the first derivative (.013-.111INV) was interpreted. However, since the constant term in this first derivative is nonsignificant, should only the expression -.111INV*SCT be interpreted? In the Table 3 case the SCT coefficient would have been significant at lower and higher values of INV (not reported) and the conclusions would not be the same as that suggested by the Table 3 results.

There may be other matters as well. For instance, noninterval data analyzed as interval data produces biased estimates (Jöreskog & Sörbom, 1989) and the situation may or may not be aggravated by the specification of interactions and quadratics comprised of ordinal latent variables. In addition, the limits of departures from unidimensionality for the re-estimation technique to work in two-step estimation are not known. In summary it is likely there are useful additions to what is known in this emerging area.

REFERENCES

Aiken, L.S. & S.G. West (1991), Multiple Regression: Testing and Interpreting Interactions, Newbury Park, CA: SAGE Publications.

Aaker, D. A., & R.P. Bagozzi (1979), “Unobservable variables in structural equation models with an application in industrial selling, Journal of Marketing Research, 16, 147-158.

Anderson, J.C. & D.W. Gerbing (1982), "Some Methods for Respecifying Measurement Models to Obtain Unidimensional Construct Measurement," Journal of Marketing Research, 19 (November), 453-60.

Anderson, J.C. & D.W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," Psychological Bulletin, 103, 411-23.

Anderson, T.W. & Y. Amemiya (1985), "The Asymptotic Normal Distribution of Estimators in Factor Analysis under General Conditions," Technical Report 12, Econometric Workshop, Stanford U.

_____ & Y. Amemiya (1986), "Asymptotic Distribution in Factor Analysis and Linear Structural Relations," Technical Report 18, Econometric Workshop, Stanford University.

Bagozzi, R.P. (1992), “State versus Action Orientation and the Theory of Reasoned Action: An Application to Coupon Usage,” Journal of Consumer Research, 18, 505-518.

Bentler, P.M. (1983), "Some Contributions to Efficient Statistics for Structural Models: Specification and Estimation of Moment Structures," Psychometrika, 48, 493-517.

Bentler, P.M. (1989), EQS Structural Equations Program Manual, Los Angeles: BMDP Statistical Software.

______ (1990), "Comparative Fit Indexes in Structural Models," Psychological Bulletin, 107 (March), 238-46.

Berry, W.D. (1984), Nonrecursive Causal Models, Beverly Hills: SAGE.

Blalock, H.M., Jr. (1965), "Theory Building and the Concept of Interaction," American Sociological Review, 30, 374-81.

Bohrnstedt, G.W. & T.M. Carter (1971), "Robustness in Regression Analysis," in Sociological Methodology, H.L. Costner ed., San Francisco: Jossey-Bass 118-46.

Bollen, K.A. (1989), Structural Equations with Latent Variables, New York: Wiley.

_____ & J.S. Long (1993), Testing Structural Equation Models, Newbury Park, CA: SAGE.

_____ & R.A. Stine (1993), “Bootstrapping Goodness of Fit Measures in Structural Equation Models,” in Testing Structural Equation Models, K.A. Bollen, J. Scott Long eds., Newbury Park, CA: Sage Publications.

Boomsma, A. (1983), On the Robustness of LISREL (Maximum Likelihood Estimation) Against Small Sample Size and Nonnormality, Unpublished Dissertation, University of Groningen.

Bristor, J.M. (1993), “Influence Strategies in Organizational Buying: The Importance of Connections to the Right People in the Right Places,” Journal of Business-to-Business Marketing, 1 (1), 63-98.

Burt, R.S. (1973). Confirmatory factor-analysis structures and the theory construction process. Sociological Methods and Research 2, 131-187.

Cohen, J. (1968), "Multiple Regression as a General Data-Analytic System," Psychological Bulletin, 70, 426-43.

______ & P. Cohen (1975), Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences, Hillsdale, NJ: Lawrence Erlbaum.

______ (1983), Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences, Hillsdale, NJ: Lawrence Erlbaum.

Dwyer, F.R. & S. Oh 1987, "Output Sector Munificence Effects on the Internal Political Economy of Marketing Channels,” Journal of Marketing Research, 24 (November), 347-358.

Evans, M.T. (1985), “A Monte Carlo Study of the Effects of Correlated Methods Variance in Moderated Multiple

Regression Analysis," Organizational Behavior and Human Decision Processes, 36, 305-323.

Fornell, C. & F.L. Bookstein (1982), "Two Structural Equation Models: LISREL and PLS Applied to Exit-Voice Theory," Journal of Marketing Research, 19 (November), 440-452.

Friedrich, R.J. (182), "In Defense of Multiplicative Terms in Multiple Regression Equations, American Journal of Political Science, 26, 797-833.

Gerbing, D.W. & J.C. Anderson (1985), "The Effects of Sampling Error and Model Characteristics on Parameter Estimation for Maximum Likelihood Confirmatory Factor Analysis," Multivariate Behavioral Research, 20, 255-271.

______ (1988), "An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment," Journal of Marketing Research, 25 (May), 186-92.

Hanushek, E.A. & J.E. Jackson (1977), Statistical Methods for Social Scientists, New York: Academic Press.

Harlow, L.L. (1985), Behavior of Some Elliptical Theory Estimators with Nonnormal Data in a Covariance Structures Framework: A Monte Carlo Study, Unpublished Dissertation, University of California, Los Angeles.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. Applied Psychological Measurement 9, 139-164.

Hayduk, L.A. (1987), Structural Equation Modeling with LISREL: Essential and Advances, Baltimore, MD: Johns Hopkins Press.

Heide, J. B. & G. John (1990), "Alliances in Industrial Purchasing: The Determinants of Joint Action in Buyer-Seller Relationships," Journal of Marketing Research, 27 (February), 24-36.

Howard, J.A. (1989), Consumer Behavior in Marketing Strategy, Englewood Cliffs, NJ: Prentice Hall.

Hu, L., P.M. Bentler & Y. Kano (1992), "Can Test Statistics in Covariance Structure Analysis Be Trusted?" Psychological Bulletin, 112, 351-362.

Jaccard, J. & C.K. Wan (1995), "Measurement Error in the Analysis of Interaction Effects Between Continuous Predictors Using Multiple Regression: Multiple Indicator and Structural Equation Approaches," Psychological Bulletin, 117 (2), 348-357.

Jaccard, J., R. Turrisi & C.K. Wan (1990), Interaction Effects in Multiple Regression, Newbury Park, CA: SAGE Publications.

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. Biometrika 57, 239-251.

_____ (1971a), "Simultaneous Factor Analysis in Several Populations," Psychometrika, 57, 409-26.

_____ (1971b), "Statistical Analysis of Sets of Congeneric Tests," Psychometrika, 36, 109-133.

_____ (1993), "Testing Structural Equation Models," in Testing Structural Equation Models, K.A. Bollen and J.S. Long eds., Newbury Park, CA: SAGE.

_____ & D. Sörbom (1989), LISREL 7 A Guide to the Program and Applications, 2nd ed., Chicago: SPSS, Inc.

_____ (1993), LISREL 8 User's Reference Guide, Chicago: Scientific Software International.

_____ & F. Yang (forthcoming), "Nonlinear Structural Equation Models: The Kenny and Judd Model with Interaction Effects," Advances in Structural Equation Modeling Techniques, G.A. Marcoulides, R.E. Schumacker, eds., Hillsdale, NJ: LEA.

Kendall, M.G. & A. Stuart (1958), The Advanced Theory of Statistics, Volume 1, London: Griffin.

Kenny, D.A. (1985), "Quantitative Methods for Social Psychology," in Handbook of Social Psychology, 3rd Ed., Vol. 1, G. Lindzey and E. Aronson eds., New York: Random House.

______ & C.M. Judd (1984), "Estimating the Nonlinear and Interactive Effects of Latent Variables," Psychological Bulletin, 96, 201-10.

Lohmöller, J. (1981), LVPLS 1.6 Latent Variables Path Analysis with Partial Least Squares Estimation, University of the Federal Armed Forces, Munich, Federal Republic of Germany.

Lubin, A. (1961), "The Interpretation of Significant Interaction," Educational and Psychological Measurement, 21, 807-817.

Lubinski, D. & L.G. Humphreys (1990), "Assessing Spurious 'Moderator Effects': Illustrated Substantively with the Hypothesized ('Synergistic') Relation Between Spatial and Mathematical Ability," Psychological Bulletin, 107, 385-93.

McClelland, G.H. & C.M. Judd (1993), "Statistical Difficulties of Detecting Interactions and Moderator Effects," Psychological Bulletin, 114 (2), 376-390.

McDonald, R.P. (1981). The dimensionality of tests and items. British Journal of Mathematical and Statistical Psychology 34, 100-117.

Ping, R.A. Jr. (1993), "The Effects of Satisfaction and Structural Constraints on Retailer Exiting, Voice, Loyalty, Opportunism, and Neglect," Journal of Retailing, 69 (Fall), 320-352.

______ (1995), "A Parsimonious Estimating Technique for Interaction and Quadratic Latent Variables," The Journal of Marketing Research, 32 (August), 336-347.

______ (1996), "Latent Variable Interaction and Quadratic Effect Estimation: A Two-Step Technique Using Structural Equation Analysis," Psychological Bulletin, 119 (January), 166-175.

Rusbult, C. E., I.M. Zembrodt & Gunn, L. K. (1982). Exit, voice, loyalty, and neglect: responses to dissatisfaction in romantic involvement. Journal of Personal and Social Psychology 43, 1230-1242.

Satorra, A. & P.M. Bentler (1988), "Scaling Corrections for Chi-Squared Statistics in Covariance Structure Analysis," Proceedings of the American Statistical Association, 308-13.

Sharma, S., R.M. Durand & O. Gur-Arie (1981), "Identification and Analysis of Moderator Variables," Journal of Marketing Research, 18 (August), 291-300.

______ S. Durvasula & W.R. Dillon (1989), "Some Results on the Behavior of Alternative Covariance Structure Estimation Procedures in the Presence of Nonnormal Data," Journal of Marketing Research, 26 (May), 214-21.

Stone-Romero, E.F., G.M. Alliger & H. Aguinis (1994), "Type II Error Problems in the Use of Moderated Multiple Regression for the Detection of Moderating Effects of Dichotomous Variables," Journal of Management, 20 (1), 167-178.

Tanaka, J.S. (1984), Some Results on the Estimation of Covariance Structure Models, Unpublished Dissertation, University of California, Los Angeles.

Wong, S.K. & J.S. Long (1987), "Reparameterizing Nonlinear Constraints in Models with Latent Variables," Technical Report, Washington State University.

Table1- STRUCTURAL MODEL RESULTS FOR THE SINGLE-INDICATOR SPECIFICATION AND LISREL8

Structural Equation Analysis Estimates:

Parameter	Estimate[a]	Parameter	Variance Estimate[a]	Parameter	Estimate[a]	t-value
8_{s1}	0.559	$,_{s1}$	0.166	N_{SAT}	1.000	
8_{s2}	0.625	$,_{s2}$	0.130	N_{ALT}	1.000	
8_{s3}	0.699	$,_{s3}$	0.110	N_{INV}	1.000	
8_{s4}	0.620	$,_{s4}$	0.119	N_{SCT}	1.000	
8_{s5}	0.663	$,_{s5}$	0.101	$N_{SAT:SCT}$	1.000	
8_{a1}	0.843	$,_{a1}$	0.272	$N_{ALT:SCT}$	1.000	
8_{a2}	0.818	$,_{a2}$	0.249	$N_{INV:SCT}$	1.000	
8_{a3}	0.910	$,_{a3}$	0.077	$N_{SCT:SCT}$	1.000	
8_{a4}	0.713	$,_{a4}$	0.244	$N_{SAT,ALT}$	-0.550*	
8_{i1}	0.683	$,_{i1}$	0.454	$N_{SAT,INV}$	0.347*	
8_{i2}	0.752	$,_{i2}$	0.120	$N_{SAT,SCT}$	0.257*	
8_{i3}	0.766	$,_{i3}$	0.089	$N_{ALT,INV}$	-0.278*	
8_{i4}	0.750	$,_{i4}$	0.124	$N_{ALT,SCT}$	-0.398*	
8_{sc1}	0.884	$,_{sc1}$	0.292	$N_{INV,SCT}$	0.532*	
8_{sc2}	0.962	$,_{sc2}$	0.209	Θ_{NEG}	0.319*	
8_{sc3}	0.947	$,_{sc3}$	0.175	$(_{NEG,SAT}$	-0.241	-4.394
8_{sc4}	0.979	$,_{sc4}$	0.210	$(_{NEG,ALT}$	0.132	2.398
$8_{sat:sct}$	0.597	$,_{sat:sct}$	0.046	$(_{NEG,INV}$	-0.138	-2.601
$8_{alt:sct}$	0.774	$,_{alt:sct}$	0.087	$(_{NEG,SCT}$	-0.018	-0.341
$8_{inv:sct}$	0.696	$,_{inv:sct}$	0.076	$(_{NEG,SAT:SCT}$	0.120	2.762
$8_{sct:sct}$	0.889	$,_{sct:sct}$	0.203	$(_{NEG,ALT:SCT}$	0.001	0.024
8_{n1}	0.691	$,_{n1}$	0.188	$(_{NEG,INV:SCT}$	-0.118	-2.684
8_{n2}	0.717	$,_{n2}$	0.107	$(_{NEG,SCT:SCT}$	-0.022	-0.482
8_{n3}	0.865	$,_{n3}$	0.063			
8_{n4}	1.000	$,_{n4}$	0.141			

Fit Indices:

Chi-Square Statistic Value	681
Chi-Square Degrees of Freedom	269
p-Value of Chi-Square Value	.000
GFI	.803
AGFI	.762
Comparative Fit Index	.910
Standardized RMS Residual	.088
RMSEA	.083
p-value for RMSEA < 0.05	.413E-06

OLS Regression Estimates:

Dependent Variable	Independent Variable[b]	b Coef- ficient	p-value	F-value and (p)	R^2
NEG	SAT	-.290	.0001	5.45 (.000)	.367
	ALT	.137	.006		
	INV	-.170	.002		
	SWC	.004	.915		
	SAT*SCT	.158	.014		
	ALT*SCT	-.004	.941		
	INV*SCT	-.133	.021		
	SCT*SCT	-.022	.646		
	Constant	.048	.311		

[a] Maximum likelihood.
[b] The independent and dependent variables were averaged and centered.

* t-value > 2.

Table2- STRUCTURAL MODEL RESULTS FOR THE SINGLE-INDICATOR SPECIFICATION AND EQS

Structural Equation Analysis Estimates:

Parameter	Estimate[a]	Parameter	Variance Estimate[a]	Parameter	Estimate[a]	t-value ML	t-value ML-Robust
8_{s1}	0.792	$,_{s1}$	0.167	N_{SAT}	0.518		
8_{s2}	0.886	$,_{s2}$	0.130	N_{ALT}	0.849		
8_{s3}	1.000	$,_{s3}$	0.109	N_{INV}	0.608		
8_{s4}	0.879	$,_{s4}$	0.119	N_{SCT}	0.969		
8_{s5}	0.940	$,_{s5}$	0.102	$N_{INV:SCT}$	0.753		
8_{a1}	0.926	$,_{a1}$	0.271	$N_{SAT,ALT}$	-0.371*		
8_{a2}	0.904	$,_{a2}$	0.249	$N_{SAT,INV}$	0.197*		
8_{a3}	1.000	$,_{a3}$	0.077	$N_{SAT,SCT}$	0.188*		
8_{a4}	0.783	$,_{a4}$	0.244	$N_{SAT,INV:SCT}$	0.055		
8_{i1}	0.893	$,_{i1}$	0.452	$N_{ALT,INV}$	-0.203*		
8_{i2}	0.986	$,_{i2}$	0.120	$N_{ALT,SCT}$	-0.366*		
8_{i3}	1.000	$,_{i3}$	0.089	$N_{ALT,INV:SCT}$	-0.056		
8_{i4}	0.978	$,_{i4}$	0.124	$N_{INV,SCT}$	0.411*		
8_{sc1}	0.902	$,_{sc1}$	0.293	$N_{INV,INV:SCT}$		-0.171*	
8_{sc2}	0.981	$,_{sc2}$	0.211	$N_{SCT,INV:SCT}$	-0.020		
8_{sc3}	0.967	$,_{sc3}$	0.174	Θ_{NEG}	0.327*		
8_{sc4}	1.000	$,_{sc4}$	0.208	$(_{NEG,SAT}$	-0.363	-4.671	-3.622
$8_{inv:sct}$	0.928	$,_{inv:sct}$	0.057	$(_{NEG,ALT}$	0.163	2.697	-2.640
8_{n1}	0.695	$,_{n1}$	0.189	$(_{NEG,INV}$	-0.172	-2.367	-2.092
8_{n1}	0.695	$,_{n1}$	0.189	$(_{NEG,SCT}$	0.013	0.245	0.224
8_{n3}	0.872	$,_{n3}$	0.063	$(_{NEG,INV:SCT}$	-0.111	-2.094	-2.403
8_{n4}	1.000	$,_{n4}$	0.141				

Fit Indices:	ML	ML-Robust
Chi-Square Statistic Value	321	276
Chi-Square Degrees of Freedom	196	
p-Value of Chi-Square Value	.000	
GFI	.882	
AGFI	.848	
Comparative Fit Index	.971	.972
Standardized RMS Residual	.045	
RMSEA	.053	
p-value for RMSEA < 0.05	.268	

Squared Multiple Correlation for NEG = .372

OLS Regression Estimates:

Dependent Variable	Independent Variable[b]	b Coef- ficient	p-value	F-value and (p)	R^2
NEG	SAT	-.318	.0002	2.68 (.000)	.344
	ALT	.155	.001		
	INV	-.157	.005		
	SWC	.021	.623		
	INV*SCT	-.100	.031		
	Constant	.042	.284		

[a] Maximum likelihood.
[b] The independent and dependent variables were averaged and centered.
* t-value > 2.

Table 3- INV-SCT INTERACTION SIGNIFICANCE

SCT-NEG Assoc.				INV-NEG Assoc.			
INV Value[a]	SCT Coef-ficient [b]	SE of SCT Coef-ficient[c]	t-value	SCT Value[d]	INV Coef-ficient[e]	SE of INV Coef-ficient[f]	t-value
				1	0.07	0.11	0.66
2	0.21	0.11	1.83	2	-0.03	0.07	-0.42
3	0.10	0.07	1.38	3	-0.14	0.06	-2.08
3.80[g]	0.01	0.05	0.24	3.25[g]	-0.17	0.07	-2.36
4	-0.00	0.05	-0.15	4	-0.25	0.09	-2.70
5	-0.11	0.07	-1.51	5	-0.36	0.13	-2.68

[a] The values ranged from 2 (=low) to 5 in the study.
[b] The coefficient of SCT is given by (.013-.111INV)SCT with INV centered.
[c] The Standard Error of the SCT coefficient is given by

$$\sqrt{Var(b_{SCT}+b_{INV*SCT}INV)}$$

$$= \sqrt{(Var(b_{SCT})+INV^2Var(b_{INV*SCT})+2INVCov(b_{SCT},b_{INV*SCT})}$$

[d] The values ranged from 1 (=low) to 5 in the study.
[e] The coefficient of INV is given by (-.172-.111SCT)INV with SCT centered.
[f] The Standard Error of the INV coefficient is given by

$$\sqrt{Var(b_{INV}+b_{INV*SCT}SCT)}$$

$$= \sqrt{Var(b_{INV})+SCT^2Var(b_{INV*SCT})+2SCTCov(b_{INV},b_{INV*SCT})}$$

[g] Mean value.

Table 4- LINEAR-TERMS-ONLY MEASUREMENT MODEL RESULTS

Parameter	Estimate[a]	Parameter	Variance Estimate[a]	Parameter	Estimate[a]
8_{s1}	0.792	$,_{s1}$	0.167	N_{SAT}	0.517
8_{s2}	0.886	$,_{s2}$	0.130	N_{ALT}	0.849
8_{s3}	1.000	$,_{s3}$	0.109	N_{INV}	0.602
8_{s4}	0.879	$,_{s4}$	0.119	N_{SCT}	0.968
8_{s5}	0.940	$,_{s5}$	0.102	N_{NEG}	0.520
8_{a1}	0.926	$,_{a1}$	0.271	$N_{SAT,ALT}$	-0.371*
8_{a2}	0.904	$,_{a2}$	0.249	$N_{SAT,INV}$	0.199*
8_{a3}	1.000	$,_{a3}$	0.077	$N_{SAT,SCT}$	0.188*
8_{a4}	0.783	$,_{a4}$	0.244	$N_{SAT,NEG}$	-0.286*
8_{i1}	0.891	$,_{i1}$	0.454	$N_{ALT,INV}$	-0.205*
8_{i2}	0.986	$,_{i2}$	0.120	$N_{ALT,SCT}$	-0.366*
8_{i3}	1.000	$,_{i3}$	0.089	$N_{ALT,NEG}$	0.309*
8_{i4}	0.978	$,_{i4}$	0.124	$N_{INV,SCT}$	0.411*
8_{sc1}	0.902	$,_{sc1}$	0.293	$N_{INV,NEG}$	-0.118*
8_{sc2}	0.982	$,_{sc2}$	0.210	$N_{SCT,NEG}$	-0.184*
8_{sc3}	0.967	$,_{sc3}$	0.175		
8_{sc4}	1.000	$,_{sc4}$	0.209		
8_{n1}	0.694	$,_{n1}$	0.190		
8_{n1}	0.730	$,_{n1}$	0.107		
8_{n3}	0.872	$,_{n3}$	0.063		
8_{n4}	1.000	$,_{n4}$	0.141		

Fit Indices:	
Chi-Square Statistic Value	305
Chi-Square Degrees of Freedom	179
p-Value of Chi-Square Value	.000
GFI	.883
AGFI	.850
Comparative Fit Index	.970
Standardized RMS Residual	.045
RMSEA	.056
p-value for RMSEA < 0.05	.153

[a] Maximum likelihood.
* t-value > 2.

Table 5- SPREADSHEET FOR THE SINGLE-INDICATOR LOADINGS AND ERRORS

	A	B	C	D	E	F	G	
1	EXCEL spreadsheet to calculate interaction and quadratic fixed and starting values							
2	using linear terms only measurement model values (see Table 4)							
3								
4	Unstandardized Linear-Terms-Only Measurement Model Values:							
5	Lambda:						Sum	
6	SAT	0.7923	0.88286	1	0.87807	0.93786	0.898218	(=SUM(B6:F6)/5)
7	ALT	0.92619	0.90489	1	0.78421		0.9038225	(=SUM(B7:E7)/4)
8	INV	0.8907	0.9848	1	0.97772		0.963305	(=SUM(B8:E8)/4)
9	SCT	0.90298	0.98275	0.96771	1		0.96336	(=SUM(B9:E9)/4)
10	Theta:						Sum	
11	SAT	0.16612	0.13209	0.10738	0.11977	0.10302	0.025135	(=SUM(B11:F11)/5^2)
12	ALT	0.27292	0.24845	0.0784	0.24354		0.052707	(=SUM(B12:E12)/4^2)
13	INV	0.45457	0.12132	0.08869	0.12417		0.049297	(=SUM(B13:E13)/4^2)
14	SCT	0.29309	0.21006	0.17531	0.2094		0.055491	(=SUM(B14:E14)/4^2)
15	Phi:							
16		SAT	ALT	INV	SCT			
17	SAT	0.51985						
18	ALT	-0.37181	0.8487					
19	INV	0.19978	-0.20527	0.603				
20	SCT	0.18902	-0.36652	0.41129	0.96864			

Calculated Interaction and Quadratic Values:

	Lambda	
sat:sct	0.865307292	(=+G6*G9)
alt:sct	0.870706444	(=+G7*G9)
inv:sct	0.928009505	(=+G8*G9)
sct:sct	0.92806249	(=+G9^2)
	Theta	
sat:sct	0.233968111	(=+G6^2*B17*G14+G9^2*E20*G11+G11*G14)
alt:sct	0.16890411	(=+G7^2*B18*G14+G9^2*E20*G12+G12*G14)
inv:sct	0.262181552	(=+G8^2*B19*G14+G9^2*E20*G13+G13*G14)
sct:sct	0.896686172	(=+4*G9^2*E20*G14+2*G14^2)
	Phi	
sat:sct	0.539276064	(=+B17*E20+B20^2)
alt:sct	0.956421678	(=+C18*E20+C20^2)
inv:sc	0.753249384	(=+D19*E20+D20^2)
sct:sct	1.876526899	(=+E20^2)

Figure 1- Single-Indicator LISREL8 Specification Code

```
NEGLECT with nonlinears (Full Structural Model)
DA NI=53 NO=222
LA
 sa1 sa2 sa3 sa4 sa5 sa6 sa7 al1 al2 al3 al4 al5 al6
 in1 in2 in3 in4 in5 in6 sc1 sc2 sc3 sc4 sc5
 ne1 ne2 ne3 ne4 ne5 ne6 ne7
 x1 x2 x3 x4 x5 x6 x7 x8 x9 x10
 ssc x12 x13 asc x14 isc scsc x15 x16 x17 x18 x19
RA FI=d:\lisrel8w\neg\neg1.dat FO
(f9.6,7f10.6/f9.6,7f10.6/f9.6,7f10.6/f9.0,6f10.0,f10.4/f9.4,7f10.4/f9.4,7f10.6/f9.6,4f10.6)
SE
 ne2 ne5 ne6 ne7
 sa2 sa4 sa5 sa6 sa7 al2 al3 al4 al5 in1 in3 in4 in5 sc2 sc3 sc4 sc5
 ssc asc isc scsc
/
MO NY=4 NX=21 ne=1 nk=8 ap=8
LE
NEG
LK
SAT ALT INV SWC
SSC ASC ISC SCSC
pa ly
*
1
1
1
0
ma ly
*
 .69415
 .73074
 .87250
1.00000
pa te
*
1 1 1 1
ma te
*
 .19060 .10722 .06301 .14102
pa lx
*
1 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0
0 1 0 0 0 0 0 0
0 1 0 0 0 0 0 0
0 1 0 0 0 0 0 0
0 0 1 0 0 0 0 0
0 0 1 0 0 0 0 0
0 0 1 0 0 0 0 0
0 0 1 0 0 0 0 0
0 0 0 1 0 0 0 0
0 0 0 1 0 0 0 0
0 0 0 1 0 0 0 0
0 0 0 1 0 0 0 0
0 0 0 0 1 0 0 0
0 0 0 0 0 1 0 0
0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 1
```

Figure 1- Single-Indicator LISREL8 Specification Code (Continued)

```
ma lx
*
 .57047  0    0    0    0    0     0    0
 .63797  0    0    0    0    0     0    0
 .71961  0    0    0    0    0     0    0
 .63319  0    0    0    0    0     0    0
 .67682  0    0    0    0    0     0    0
    0 .85382  0    0    0    0     0    0
    0 .83309  0    0    0    0     0    0
    0 .92148  0    0    0    0     0    0
    0 .72202  0    0    0    0     0    0
    0    0 .69181  0    0    0     0    0
    0    0 .76535  0    0    0     0    0
    0    0 .77603  0    0    0     0    0
    0    0 .75915  0    0    0     0    0
    0    0    0 .88871  0    0     0    0
    0    0    0 .96722  0    0     0    0
    0    0    0 .95239  0    0     0    0
    0    0    0 .98422  0    0     0    0
    0    0    0    0 .61402  0     0    0
    0    0    0    0    0 .79433   0    0
    0    0    0    0    0    0  .70928  0
    0    0    0    0    0    0     0 .89896
pa td
*
21*1
ma td
*
 .16702 .13028 .10938 .11966 .10219
 .27195 .24934 .07797 .24416
 .45436 .12037 .08948 .12429
 .29309 .21006 .17536 .20934
 .04726 .03345 .05733 .20569
pa ga
*
!sa al in sc ssc asc isc scsc
  1  1  1  1  1   1   1   1 !ne
pa ph
*
  0 !sa
  1  0 !al
  1  1  0 !in
  1  1  1  0 !sc
  0  0  0  0  0 !ssc
  0  0  0  0  0   0 !asc
  0  0  0  0  0   0   0 !isc
  0  0  0  0  0   0   0   0 !scsc
!sa al in sc ssc asc isc scsc
ma ph
*
  1 !sa
-.55972    1 !al
 .35660 -.28679   1 !in
 .26639 -.40420 .53819 1 !sc
   0       0      0    0  1 !ssc
   0       0      0    0  0   1 !asc
   0       0      0    0  0   0   1 !isc
   0       0      0    0  0   0   0   1 !scsc
! sa      al     in   sc ssc asc isc scsc
```

Figure 1- Single-Indicator LISREL8 Specification Code (Continued)

```
!par(1)=lsat, par(2)=lsct, par(3)=lalt, par(4)=linv
co par(1)=.2*lx(1,1)+.2*lx(2,1)+.2*lx(3,1)+.2*lx(4,1)+.2*lx(5,1)
co par(2)=.25*lx(14,4)+.25*lx(15,4)+.25*lx(16,4)+.25*lx(17,4)
co lx(18,5)=par(1)*par(2)
co par(3)=.25*lx(6,2)+.25*lx(7,2)+.25*lx(8,2)+.25*lx(9,2)
co lx(19,6)=par(3)*par(2)
co par(4)=.25*lx(10,3)+.25*lx(11,3)+.25*lx(12,3)+.25*lx(13,3)
co lx(20,7)=par(4)*par(2)
co lx(21,8)=par(2)*par(2)
!par(1)=lsat, par(2)=lsct, par(3)=lalt, par(4)=linv
!par(5)=tsat, par(6)=tsct, par(7)=talt, par(8)=tinv
co par(5)=.04*td(1,1)+.04*td(2,2)+.04*td(3,3)+.04*td(4,4)+.04*td(5,5)
co par(6)=.0625*td(14,14)+.0625*td(15,15)+.0625*td(16,16)+.0625*td(17,17)
!                psat*tsct*lsat^2   +  psct*tsat*lsct^2     + tsat*tsct
co td(18,18)=ph(1,1)*par(6)*par(1)^2+ph(4,4)*par(5)*par(2)^2+par(5)*par(6)
!                4psct*tsct*lsct^2    + 2*tsct*tsct
co par(7)=.0625*td(6,6)+.0625*td(7,7)+.0625*td(8,8)+.0625*td(9,9)
!                palt*tsct*lalt^2   +  psct*talt*lsct^2     + talt*tsct
co td(19,19)=ph(2,2)*par(6)*par(3)^2+ph(4,4)*par(7)*par(2)^2+par(7)*par(6)
co par(8)=.0625*td(10,10)+.0625*td(11,11)+.0625*td(12,12)+.0625*td(13,13)
!                pinv*tsct*linv^2   +  psct*tinv*lsct^2     + tinv*tsct
co td(20,20)=ph(3,3)*par(6)*par(4)^2+ph(4,4)*par(8)*par(2)^2+par(8)*par(6)
co td(21,21)=4*ph(4,4)*par(6)*par(2)^2+2*par(6)^2
OU xm nd=5 it=100 ad=off
```

Figure 2- Single-Indicator EQS Specification Code

```
/TITLE=NEGLECT with INV*SCT only (Structural Model);
/SPECIFICATIONS
 VARIABLES = 53; ME = ml,robust;
 cases=222;MA = raw;
 FO='(f9.6,7f10.6/f9.6,7f10.6/f9.6,7f10.6/
      f9.0,6f10.0,f10.4/f9.4,7f10.4/
      f9.4,7f10.6/f9.6,4f10.6)';
 DATA='c:\eqswin\neg\neg1.dat';
/labels
 v2=sa2;v4=sa4;v5=sa5;v6=sa6;v7=sa7;v9=al2;v10=al3;v11=al4;v12=al5;
 v14=in1;v16=in3;v17=in4;v18=in5;v21=sc2;v22=sc3;v23=sc4;v24=sc5;
 v26=ne2;v29=ne5;v30=ne6;v31=ne7;v42=ssc;v45=asc;v47=isc;v48=scsc;
/EQUATIONS
 V2 =  .9*f1           + e1;
 V4 =  .9*f1           + e2;
 v5 =  1.0f1           + e3;
 v6 =  .9*f1           + e4;
 v7 =  .9*f1           + e5;
 v9 =    .9*f2         + e6;
 v10 =   .9*f2         + e7;
 v11 =   1.0f2         + e8;
 v12 =   .9*f2         + e9;
 v14 =     .9*f3       + e10;
 v16 =     .9*f3       + e11;
 v17 =     1.0f3       + e12;
 v18 =     .9*f3       + e13;
 v21 =       .9*f4     + e14;
 v22 =       .9*f4     + e15;
 v23 =       .9*f4     + e16;
 v24 =       1.0f4     + e17;
 v26 =         .9*f5   + e18;
 v29 =         .9*f5   + e19;
 v30 =         .9*f5   + e20;
 v31 =         1.0f5   + e21;
!v42 =   0.865307292f6 + e22;
!v45 =   0.870706444f7 + e23;
 v47 =   0.928009505f8 + e24;
!v48 =   0.928062490f9 + e25;
 f5 = -.29*f1 + .13*f2 + -.17*f3 + -.004*f4 + -.13*f8 + d1;
/VARIANCES
 F1 = .51985*;
 f2 = .8487*;
 f3 = .603*;
 f4 = .96864*;
!f5 =
!F6 = 0.539276064*;
!F7 = 0.956421678*;
 F8 = 0.753249384;
!F8 = 0.753249384*;
!F9 = 1.876526899*;
 e1 to e21 = .1*;
!e22 = 0.047264018;
!e23 = 0.033451716;
 e24 = 0.057338752;
!e25 = 0.205695870;
 d1 = .5*;
```

Figure 2- Single-Indicator EQS Specification Code (Continued)

```
/COVARIANCE
 f1,f2 = -.37*;
 f1,f3 = .19*;
 f1,f4 = .18*;
 f2,f3 = -.20*;
 f2,f4 = -.36*;
 f3,f4 = .41*;
 f1,f8 = *;f2,f8 = *;f3,f8 = *;f4,f8 = *;
/print
 dig=5;
/END
```

Appendix- Calculated Interaction and Quadratic Variances and Covariances

The following presents the corrections for unadjusted variances and covariances (e.g., SAS or SPSS values) involving interactions and quadratics.

An estimate of the variance of the latent variable $>_X$ using the variance of the observed variable X = $(x_1 + x_2)/2$, where x_1 and x_2 are the observed indicators of $>_X$ (i.e., $x_1 = 8_{X1}>_X + ,_{x1}$ and $x_2 = 8_{X2}>_X + ,_{x2}$), x_1 and x_2 are independent of $,_{x1}$ and $,_{x2}$, $,_{x1}$ and $,_{x2}$ are independent of each other, and x_1 and x_2 are multivariate normal with zero means, is given by the following. Let $\text{э}_a = (8_{a1} + 8_{a2})/2$. Then

$$\begin{aligned} Var(X) &= Var[(x_1 + x_2)/2] \\ &= Var[\text{э}_X>_X + (,_{x1} + ,_{x2})/2] \\ &= \text{э}_X{}^2 Var(>_X) + [Var(,_{x1}) + Var(,_{x2})]/2^2 \\ &= \text{э}_X{}^2 Var(>_X) + 2_X, \end{aligned} \tag{i}$$

where Var(a) is the variance of a, Var(X) is the observed variance of X, and $2_X = [Var(,_{x1}) + Var(,_{x2})]/2^2$. As a result, an estimate of $Var(>_X)$ is given by

$$Var(>_X) = (Var(X) - 2_X)/\text{э}_X{}^2.$$

For $Cov(>_X>_Z)$, where Cov(a,b) is the covariance of a and b,

$$\begin{aligned} Cov(X,Z) &= Cov[(x_1 + x_2)/2, (z_1 + z_2)/2] \\ &= [Cov(8_{X1}>_X+,_{x1},8_{Z1}>_Z+,_{z1}) + Cov(8_{X1}>_X+,_{x1},8_{Z2}>_Z+,_{z2}) \\ &\qquad + Cov(8_{X2}>_X+,_{x2},8_{Z1}>_Z+,_{z1}) + Cov(8_{X2}>_X+,_{x2},8_{Z2}>_Z+,_{z2})]/2^2 \\ &= Cov(>_X,>_Z)\text{э}_X\text{э}_Z , \end{aligned} \tag{ii}$$

and an estimate of $Cov(>_X>_Z)$ is given by

$$Cov(>_X,>_Z) = Cov(X,Z)/\text{э}_X\text{э}_Z ,$$

where $Z = (z_1 + z_2)/2$.

Off-diagonal terms comprised of an interaction and a linear variable that does not appear in the interaction such as $Cov(>_V,>_W>_X)$ are estimated as follows:

$$Cov(V,WX) = Cov(\text{э}_V>_V + E_V,[\text{э}_W>_W + E_W][\text{э}_X>_X + E_X]) ,$$

where $E_a = (,_{a1} + ,_{a2})/2$. Hence

$$Cov(V,WX) = Cov(>_V,>_W>_X)\text{э}_V\text{э}_W\text{э}_X ,$$

and

$$Cov(>_V,>_W>_X) = Cov(V,WX)/\text{э}_V\text{э}_W\text{э}_X.$$

The covariance of two interactions with no common linear variables is given by

$$Cov(VW,XZ) = Cov(V,X)Cov(W,Z) + Cov(V,Z)Cov(W,X) , \tag{iii}$$

(Kendall & Stewart, 1958), and

$$\begin{aligned} Cov(VW,XZ) &= Cov(>_V,>_X)\text{э}_V\text{э}_X Cov(>_W,>_Z)\text{э}_W\text{э}_Z \\ &\quad + Cov(>_V,>_Z)\text{э}_V\text{э}_Z Cov(>_W,>_X)\text{э}_W\text{э}_X \\ &= Cov(>_V>_W,>_X>_Z)\text{э}_V\text{э}_W\text{э}_X\text{э}_Z , \end{aligned}$$

by equality ii. An estimate of $Cov(>_V>_W,>_X>_Z)$ is therefore given by

$$Cov(>_V>_W,>_X>_Z) = Cov(VW,XZ)/\text{э}_V\text{э}_W\text{э}_X\text{э}_Z. \tag{iv}$$

By equality iv the covariance of two quadratics such as $Cov(>_X>_X,>_Z>_Z)$ is

$$Cov(>_X>_X,>_Z>_Z) = Cov(XX,ZZ)/\text{э}_X{}^2\text{э}_Z{}^2.$$

For the variance of an interaction

$$\begin{aligned} Var(XZ) &= Cov(XZ,XZ) \\ &= Var(X)Var(Z) + Cov(X,Z)^2 , \end{aligned}$$

using equality iii. Hence

$$\begin{aligned} Var(XZ) &= [\text{э}_X{}^2 Var(>_X) + 2_X][\text{э}_Z{}^2 Var(>_Z) + 2_Z] + [Cov(>_X,>_Z)\text{э}_X\text{э}_Z]^2 \\ &= Cov(>_X>_Z,>_X>_Z)\text{э}_X\text{э}_Z\text{э}_X\text{э}_Z + Var(>_X)\text{э}_X{}^2 2_Z \\ &\quad + Var(>_Z)\text{э}_Z{}^2 2_X + 2_X 2_Z , \end{aligned}$$

using i and ii, and

$$Var(>_X>_Z) = (Var(XZ) - Var(>_X)\text{э}_X{}^2 2_Z - Var(>_Z)\text{э}_Z{}^2 2_X - 2_X 2_Z)/\text{э}_X{}^2\text{э}_Z{}^2$$

= (Var(XZ) - 2$_Z$Var(X) - 2$_X$Var(Z) + 2$_X$2$_Z$)/э$_X^2$э$_Z^2$.

The corrected estimate of a quadratic such as Var(XX) is similar:

Var(XX) = 2Var(X)^2

= 2[э$_X^2$Var(>$_X$) + 2$_X$]^2

= Var(>$_X$>$_X$)э$_X^4$ + 4Var(>$_X$)э$_X^2$2$_X$ +22$_X^2$,

by equality iii, and

Var(>$_X$>$_X$) = (Var(XX) - 4Var(>$_X$)э$_X^2$2$_X$ - 22$_X^2$)/э$_X^4$

= (Var(XX) - 4Var(X)2$_X$ + 22$_X^2$)/э$_X^4$.

For the covariance of a quadratic and an interaction that has a common linear variable such as Cov(XX,XZ),

Cov(XX,XZ) = 2Var(X)Cov(X,Z)

= 2[э$_X^2$Var(>$_X$) + 2$_X$]Cov(>$_X$,>$_Z$)э$_X$э$_Z$

= Cov(>$_X$>$_X$,>$_X$>$_Z$)э$_X^2$э$_X$э$_Z$ + 2Cov(>$_X$,>$_Z$)э$_X$э$_Z$2$_X$,

by equalities ii and iii, and

Cov(>$_X$>$_X$,>$_X$>$_Z$) = (Cov(XX,XZ) - 2Cov(>$_X$,>$_Z$)э$_X$э$_Z$2$_X$)/э$_X^3$э$_Z$

= (Cov(XX,XZ) - 2Cov(X,Z)2$_X$)/э$_X^3$э$_Z$.

For a combination of interactions with a common linear variable such as Cov(VW,VZ)

Cov(VW,VZ) = Var(V)Cov(W,Z) + Cov(V,Z)Cov(W,V)

= [э$_V^2$Var(>$_V$) + 2$_V$]Cov(>$_W$,>$_Z$)э$_W$э$_Z$

+ Cov(>$_V$,>$_Z$)э$_V$э$_Z$Cov(>$_W$,>$_Z$)э$_W$э$_V$

= Cov(>$_V$>$_W$,>$_V$>$_Z$)э$_V^2$э$_W$э$_Z$ + Cov(>$_W$,>$_Z$)э$_W$э$_Z$2$_V$

by equalities ii and iii, and

Cov(>$_V$>$_W$,>$_V$>$_Z$) = (Cov(VW,VZ) - Cov(W,Z)2$_V$)/э$_V^2$э$_W$э$_Z$.

By induction these estimates can be generalized to latent variables with an arbitrary number of indicators, e.g., V = (v_1 + v_2 +...+ v_p)/p, where v_i are the observed indicators of >$_V$ (i.e., v_i = 8$_{vi}$>$_V$ + ,$_{vi}$). 2$_V$ is given by 2$_V$ = [Var(,$_{v1}$) + Var(,$_{v2}$) +...+ Var(,$_{vp}$)]/p^2.

PSEUDO LATENT VARIABLE REGRESSION: AN ASSESSABLE ESTIMATION TECHNIQUE FOR LATENT VARIABLE INTERACTIONS

(An earlier, but revised, version of Ping 2003, *2003 Academy of Marketing Science Conference Proceedings*, Miami: Academy of Marketing Science)
(Updated October, 2003)

ABSTRACT

Because estimating interactions involving unobserved or latent variables in survey data has been difficult for substantive researchers (Aiken and West 1991), the paper proposes pseudo latent variable regression-- an Ordinary Least Squares Regression approach that uses reliabilities to adjust the regression input covariance matrix. Using simulated data, the proposed approach performed adequately. A pedagogical example is provided to illustrate the use of the proposed technique.

Unobserved or latent variables interactions, for example *XZ* in

$$Y = b_1X + b_2Z + b_3XZ + \zeta, \quad (1$$

where *X*, *Z*, *XZ*, and *Y* each have multiple indicators measured with error, have been difficult for substantive researchers to estimate (Aiken and West 1991). For example, Podsakoff, Todor, Grover and Huber (1984) examined 576 interactions involving moderators of leadership behaviors (i.e., interactions) using ordinary least squares (OLS) regression and survey data, and found 72 were significant, an incidence rate only slightly above that of chance. Later, McClelland and Judd (1993) demonstrated that interactions are inherently difficult to detect in survey data, and they suggested that experiments should be used instead.

Bohrnstedt and his colleagues (Bohrnstedt and Goldberger 1969, Bohrnstedt and Marwell 1978), among others, shed additional light on these difficulties when OLS regression is used. They demonstrated that regression is unreliable for estimating interaction coefficients when the interaction's constituent variables (e.g., *X* and *Z* in Equation 1) are measured with error: The resulting regression coefficients are biased (i.e., regression coefficient averages across many data sets do not approximate the population value) and inefficient (i.e., coefficient estimates vary widely across data sets from the same population).

Fortunately, other techniques for estimating interactions when there are errors in the constituent variables have been proposed (e.g., Cohen and Cohen 1975; Bohrnstedt and Marwell 1978; Feucht 1989; Fuller and Hidiroglu 1978; Hayduk 1987; Heise 1986; Kenny and Judd 1984; Ping 1995, 1996a, 1996b; Wong and Long 1987). Unfortunately, these techniques are difficult to use for reasons that include being limited to single indicator latent variables, the technique is inaccessable, it is tedious, or it requires structural equation analysis.

This paper proposes an accessible technique for jointly estimating several multiple indicator latent variable interactions using OLS regression software available in popular statistics software such as SAS and SPSS. The proposed technique, which we will term >pseudo latent variable regression,= uses sample based reliabilities to adjust the covariance matrix that is used in OLS regression. Thus, it could be used as an alternative to a structural model in structural equation analysis when estimating latent variable interactions. The technique may also be used for interactions involving >formative= latent variables (i.e., unobserved variables defined by their items rather than their items being observed instances of unobserved variables-- see Fornell and Bookstein 1982). In addition, the technique allows the joint investigation of multiple (or all possible) interactions, to either aid in the interpretation of significant effects as is routinely done in ANOVA studies, or to probe hypothesized but nonsignificant associations (i.e., to determine if the nonsiginificant associations are conditionally significant, or significant in subsets of the data).

The paper begins with a brief review of latent variable regression (Ping 1996b), upon which pseudo latent variable regression is based. Then it proposes substituting sample based reliabilities for latent variable regression=s requirement for measurement model parameter estimates. Next, using simulated data sets, the paper evaluates the performance of this proposed substitution. It concludes with a pedagogical example illustrating the proposed technique.

LATENT VARIABLE REGRESSION

To estimate interactions involving unobserved or latent variables with multiple indicators Ping (1996b) suggested using an input covariance matrix, which is adjusted for measurement error using structural equation analysis parameter estimates, and OLS regression. Using simulated data the proposed technique, latent variable regression, performed adequately by producing unbiased regression coefficients under various conditions. The technique uses the sample covariance matrix (e.g., available from SAS, SPSS, etc.), which is adjusted for measurement error, as input to OLS regression. The adjustments to the sample covariance matrix involve loadings and measurement errors from structural equation analysis (i.e., λ_{xi} and ε_{xi} respectively in $x_i = \lambda_{xi}X + \varepsilon_{xi}$, where λ_{xi} is the loading or path coefficient between x_i and X, and ε_{xi} is the measurement error of x_i). For example, for latent variables X, Z, XZ, and Y meeting the Kenny and Judd (1984) normality assumptions (i.e., indicators are multivariate normal with zero means and independent of their measurement errors, and measurement errors are independent of each other), a measurement model for X, Z and Y is estimated (i.e., a structural equation analysis model in which X, Z and Y are specified with their indicators, and X, Z and Y are specified as intercorrelated). Then the error-adjusted variance, Var(X), for X (= $x_1+x_2+...+x_n$) is estimated as follows:

$$\mathrm{Var(X)} = \mathrm{Var}(\lambda_{x1}x_1 + \varepsilon_{x1} + \lambda_{x2}x_2 + \varepsilon_{x2} + ... + \lambda_{xn}x_n + \varepsilon_{xn})$$
$$= \Lambda_X{}^2\,\mathrm{Var}(X) + \theta_X$$

and $$\mathrm{Var}(X) = [\mathrm{Var(X)}-\theta_X]/[\Lambda_X{}^2], \qquad (2$$

where Var(X) is the sample variance of X (available from SAS, SPSS, etc.), θ_x is the sum of the measurement errors (= $\mathrm{Var}[\varepsilon_{x1}] +...+ \mathrm{Var}[\varepsilon_{xn}]$) provided by the measurement model, and Λ_X is the sum of the loadings (= $\lambda_{x1} +...+ \lambda_{xn}$) from that measurement model. The error-adjusted variances of Z and Y are computed in a similar manner.

The error-adjusted covariance of X and Z is estimated using

$$\mathrm{Cov}(X,Z) = \mathrm{Cov(X,Z)}/[\Lambda_X\Lambda_Z], \qquad (3$$

where Cov(X,Z) is the attenuated covariance of X and Z (=$z_1+...+z_m$). The covariances of Y with X, and Z are estimated similarly.

The adjusted covariance of XZ with Y is estimated using

$$\mathrm{Cov}(XZ,Y) = \mathrm{Cov(XZ,Y)}/[\Lambda_X\Lambda_Z\Lambda_Y]. \qquad (4$$

The covariances of XZ with X and Z are estimated similarly.

Finally, the adjusted variance of XZ is estimated using

$$\mathrm{Var}(XZ) = (\mathrm{Var(XZ)} - \Lambda_X{}^2\mathrm{Var}(X)\theta_Z - \Lambda_Z{}^2\mathrm{Var}(Z)\theta_X - \theta_X\theta_Z)/[\Lambda_X{}^2\Lambda_Z{}^2]\ . \qquad (5$$

The adjustments for more than one interaction are shown in Ping (1996b).

PSEUDO LATENT VARIABLE REGRESSION

As an alternative to using measurement model parameter estimates, we propose using sample based reliabilities to estimate the loadings (e.g., Λ_X) and measurement errors (e.g., θ_X) in the adjustment equations for latent variable regression (e.g., equations 2-5). Werts, Linn and Jöreskog (1974) suggested the latent variable reliability (ρ) of a measure of the unidimensional latent variable X (i.e., the measure has only one underlying latent variable) is

$$\rho_X = \Lambda_X{}^2\mathrm{Var}(X)/[\Lambda_X{}^2\mathrm{Var}(X) + \theta_X] \ . \qquad (6$$

Using the definition of reliability, θ_X can be estimated by

$$\theta_X = \mathrm{Var(X)}(1 - \rho_X) \ . \qquad (7$$

Authors defin the reliability of a unidimensional indicator as the square of the loading between the indicator and its latent variable (e.g., Bollen, 1989). Thus, the square root of ρ could be used to estimate Λ (Kenny, 1979), and

$$\Lambda_X = \rho_X{}^{1/2} \ . \qquad (8$$

Anderson and Gerbing (1988) pointed out that for unidimensional constructs there is little practical difference between coefficient alpha (α) and ρ. Thus for unidimensional constructs an estimate of θ_X is

$$\theta_X \nabla \mathrm{Var(X)}(1 - \alpha_X) \ , \qquad (9$$

and for reliable X an estimate of Λ_X is

$$\Lambda_X = \alpha_X{}^{1/2} \ . \qquad (10$$

Thus for unidimensional latent variables, the adjustment equations used in latent variable regression (e.g., equations 2-5) could utilize the equations 9 and 10 estimates of loadings and measurement errors. In the balance of this section the paper evaluates the performance of this suggestion using simulated data.

Simulated Data Sets

Table 1, which shows the results of recovering known population structural coefficients (i.e., b=s in $Y = b_0 + b_1X + b_2Z + b_3XZ + b_4W + \zeta_Y$) with reliabilities of .7 and .9, structural coefficient size corresponding to R^2 's of .10 and .50, and sample sizes of 100 and 300 cases (see Appendix A for details), suggests that the suggested technique performed adequately. Table 1 also shows structural equation analysis estimation results for comparison. The pseudo latent variable regression coefficient averages (Column 4 in Table 1) were within a few points of the population values, and thus the biases were small. This suggests the suggested technique is unbiased. In addition, the variations of these coefficients around the population values (RMSE=s in Table 1) were comparatively small (and equivalent to those from structural equation analysis). This suggests the proposed technique is at leaset as efficient as structural equation analysis.

Table 2 shows the performance of the latent variable regression standard error term proposed by Ping (2001) when used with pseudo latent variable regression. These results also suggest the proposed technique performed adequately. The average standard errors for pseudo latent variable regression (Column 4 in Table 2) were within a few points of the Root Mean Standard Errors of the coefficients, and as a result the biases were comparatively small (and also equivalent to those from structural equation analysis). This suggests the standard error term is unbiased in this application.

AN EXAMPLE

For pedagogical purposes we will reanalyze data reported in Ping (1993). There exiting (E) in interfirm economic and social exchange relationships between firms was argued to be associated with relationship satisfaction (S), alternative relationship attractiveness (A), investment in the relationship (I), and the cost to switch to an alternative relationship (C). To illustrate the use of pseudo latent variable regression we will investigate several of the nonsignificant associations reported in Ping (1993) by adding SxA and AxI interactions to the model

$$\mathrm{E} = b_1\mathrm{S} + b_2\mathrm{A} + b_3\mathrm{I} + b_4\mathrm{C} + \zeta, \qquad (11$$

and estimating the model

$$E = b_1'S + b_2'A + b_3'I + b_4'C + b_5SxA + b_6AxI + \zeta \qquad (11'$$

(adding interactions changes the equation 11 regression coefficients-- see Aiken and West 1991).

First the unidimensionality of the measures for S, A, I, C, and E was assessed. Next the indicators for S, A, I, C, and E were zero centered. This was accomplished by subtracting the mean of an indicator from each case value of that indicator. Zero centering produces indicators with means of zero, which was assumed in deriving equations 2-5. It also reduces the otherwise high collinearity of an interaction (e.g., SxA) with its constituent variables (e.g., S and A). After zero centering, the indicators for each latent variable were summed to form the variables S, A, I, C, and E, then these variables were added to the data set. For emphasis, the indicators were summed, not averaged (averaging the indicators changes equations 2-5, and 7-10). Then the SxA and AxI interactions were formed by computing the products of S with A and A with I, and adding the results to each case in the data set. Next the sample (unadjusted) covariance matrix of S, A, I, C, E, SxA and AxI was obtained using SPSS (see Table 3).

Then, the coefficient alpha reliabilities for S, A, I, C, and E were produced by SPSS (see Table 3), and the sample covariance matrix was adjusted using EXCEL and equations 2-5 with $\Lambda_! = n_!(\alpha_!)^2$ and $\Lambda_{/}{}^2 = n_!\alpha_!$, where ! is the latent variable *S*, *A*, *I*, *C*, or *E*, and $n_!$ is the number of indicators of ! ($n_!$ is required to make $\Lambda_!$ and $\Lambda_{/}{}^2$ commensurate with Var(!)). Next the resulting adjusted covariance matrix (see Table 4) was used as input to OLS regression in SPSS, and this produced the pseudo latent variable regression results shown in Table 4.

The coefficient standard errors (SE=s) for the pseudo latent variable regression coefficients were then computed as follows. The Table 3 unadjusted covariance matrix was used to obtain unadjusted regression coefficient SE=s and the unadjusted regression standard error of the estimate (SEE) (SEE = $\Sigma[y_i - \{_i]^2$, where y_i and $\{_i$. are observed and estimated y=s respectively). To obtain an unadjusted regression SEE commensurate with that from latent variable regression, the unadjusted regression SEE was divided by the number of indicators of E (7). The adjusted pseudo latent variable regression SE=s were then calculated by multiplying each unadjusted regression coefficient SE by the ratio of the unadjusted regression SEE to the pseudo latent variable regression SEE. This procedure was suggested in Ping (2001) for Latent Variable Regression, and it is similar to the procedure used in two stage least squares to produce correct coefficient SE=s for the two stage least squares coefficient estimates.

The results are shown in Table 4, along with unadjusted regression results and structural equation estimates for comparison. Because the reliabilities of the variables were high (see Table 4), the unadjusted regression coefficient estimates were similar to those from pseudo latent variable regression and the structural equation analysis. However, with lower reliabilities the unadjusted regression coefficient estimates quickly diverge from the pseudo latent variable regression and the structural equation analysis estimates (not shown).

The SxA interaction can be interpreted as follows. Table 5 shows the contingent A association resulting from the significant SxA interaction. At the Table 5 average value of S in the study, which b_3 in equation 11 represents, the A-E association was significant. However when S was very high, the A-E association was nonsignificant. As S decreased, this association became stronger, and for lower levels of S it was significant. A substantive interpretation would be that when satisfaction was high, changes in alternative attractiveness had no association with exiting, as Rusbult and Buunk (1993) predicted, but as satisfaction decreased this association was significant and positive. As an aside, this illustrates the importance of post-hoc probing of significant associations for interactions. Although the unmoderated A-E association was significant, for very high values of S the association was nonsignificant.

The latent variable A also moderated the S-E association and it could also be interpreted. As shown in Table 5, when A was lower the S-E association was significant and negative, and as it increased the S-E association became stronger. Thus the S-E association is contingent on the level of A, and at lower values of A S has a weaker association with E than at higher values of A. A substantive interpretation would be that with low alternative attractiveness satisfaction has a weaker association with exiting that when alternative attractiveness is higher.

DISCUSSION

Several observations may be of interest. As the Table 5 survey data results suggest, pseudo latent variable regression can produce results that are interpretationally equivalent to structural equation analysis (i.e., they provide coefficient estimates with equivalent interpretations of significance). However, interpretational equivalence may not always hold for an association with a significance close to |t| = 2 (i.e., one technique may suggest the association is significant while the other may suggest it is nonsignificant), or when Maximum Likelihood (ML) estimates are produced by structural equation analysis (instead of GLS estimates). Remedies include bootstrapping the covariance matrix of the items and X, Z, XZ, and Y to lessen the effects of sampling variation and clarify the interpretation of an association with a t-value close to ∀2. Bootstrapping a covariance matrix is accomplished by averaging the covariance matrices that result from taking a large number of subsamples of the cases (e.g., several hundred subsamples each with 10-20% of the cases randomly deleted) (see Bentler 1989:76, and Jöreskog and Sörbom 1996:173,185).

To obtain ML estimates for pseudo latent variable regression, a LISREL, EQS, AMOS, etc. structural model with single summed indicators for X, Z, XZ, and Y (= Σx_i, Σz_i, $(\Sigma x_i)(\Sigma z_i)$, and Σy_i, respectively), and loadings and errors fixed at the equations 9 and 10 values could be used. The reliability of XZ is $\rho_{XZ} = (r_{XZ}{}^2 + \rho_X\rho_Z)/(r_{XZ}{}^2 + 1)$, where ρ denotes reliability and $r_{XZ}{}^2$ is the correlation of X and Z.

In addition, interpretational equivalence may decline with reliabilities below .7. We briefly investigated reliabilities of .6 and .5 (not reported), and the pseudo latent variable regression results appeared to diverge from those using structural equation analysis as reliability declined. Specifically, they became increasingly more biased than the structural equation analysis results, especially in samples of 100 cases. Thus pseudo latent variable regression should be used with caution in preliminary research, pilot tests, etc. where measures may can have reliabilities below .7. Similarly, we briefly investigated sample sizes below 100 (not reported). Again, pseudo latent variable regression results appeared to diverge from those of structural equation analysis as sample size declined, with pseudo latent variable regression becoming increasingly more biased as sample size declined. Thus pseudo latent variable regression results should also be interpreted with caution if the sample size is below 100.

Finally, the Tables 1 and 2 results for the coefficients of X, Z and W suggest that pseudo latent variable regression might be used to estimate equation 11 (i.e., in models with no interaction(s) specified). In fact, we have used pseudo latent variable regression in models with no second-order terms to produce error-adjusted forecasting equations. This is accomplished by "stepping in" the variables, based on their R^2, to produce a forecast equation that contains only the "important" forecast variables. This can also be accomplished using structural equations analysis, but the process is tedious.

REFERENCES

Aiken, Leona S. and Stephen G. West (1991), *Multiple Regression: Testing and Interpreting Interactions*, Newbury Park, CA: SAGE Publications.

Anderson, James C. and David W. Gerbing (1984), AThe Effect of Sampling Error on Convergence, Improper Solutions, and Goodness of Fit Indices for Maximum Likelihood Confirmatory Factor Analysis,≅ *Psychometrika*, 49, 155-173.

_______ and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 103 (May), 411-23.

Bentler, Peter M. (1989), *EQS Structural Equations Program Manual*, Los Angeles: BMDP Statistical Software.

Bohrnstedt, G. W. and A. S. Goldberger (1969), AOn the Exact Covariance of Products of Random Variables,≅ *Journal of the American Statistical Society*, 64, 1439-1442.

Bohrnstedt, G. W. and G. Marwell (1978), AThe Reliability of Products of Two Random Variables,≅ in *Sociological Methodology*, K. F. Schuessler ed., San Francisco: Jossy Bass, 254-273.

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*. New York: Wiley.

Browne, Michael W. and Robert Cudeck (1993), AAlternative Ways of Assessing Model Fit,≅ in *Testing Structural Equation Models*, K. A. Bollen et al. eds, Newbury Park CA: SAGE.

Cohen, J. and P. Cohen (1975), *Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences*, Hillsdale, NJ: Lawrence Erlbaum.

Feucht, Thomas E. (1989), AEstimating Multiplicative Regression Terms in the Presence of Measurement Error,≅ *Sociological Methods and Research*, 17 (February), 257-282.

Fornell, Claes Fred L. Bookstein (1982), ATwo Structural Equation Models: LISREL and PLS Applied to Exit-Voice Theory,≅ *Journal of Marketing Research*, 19 (November), 440-452.

Fuller, W. A. and M. A. Hidiroglou (1978), Aregression Estimation After Correcting for Attenuation,≅ *Journal of the American Statistical Association*, 73, 99-104.

Hayduk, Leslie A. (1987), *Structural Equation Modeling with LISREL: Essential and Advances*, Baltimore, MD: Johns Hopkins Press.

Heise, David R. (1986), AEstimating Nonlinear Models Correcting for Measurement Error,≅ *Sociological Methods and Research*, 14 (May), 447-472.

Jöreskog, Karl G.(1993), "Testing Structural Equation Models," In *Testing Structural Equation Models*, Kenneth A. Bollen and J. Scott Long eds., Newbury Park, CA: SAGE.

________ Dag Sörbom (1996), *Prelis 2 User=s Reference Guide*, Chicago: Scientific Software International, Inc.

Kenny, David (1979), *Correlation and Causality*. New York: Wiley.

Kenny, David and Charles M. Judd (1984), "Estimating the Nonlinear and Interactive Effects of Latent Variables," *Psychological Bulletin*, 96 (July), 201-10.

McClelland, G. H. and C. M. Judd (1993), AStatistical Difficulties of Detecting Interactions and Moderator Effects,≅ *Psychological Bulletin*, 114 (2), 376-390.

Ping, R. (1993), "The Effects of Satisfaction and Structural Constraints on Retailer Exiting, Voice, Loyalty, Opportunism, and Neglect," *Journal of Retailing*, 69 (Fall), 320-352.

________ (1995), "A Parsimonious Estimating Technique for Interaction and Quadratic Latent Variables," *The Journal of Marketing Research*, 32 (August), 336-347.

________ (1996a), "Latent Variable Interaction and Quadratic Effect Estimation: A Two-Step Technique Using Structural Equation Analysis, *The Psychological Bulletin*, 119 (January), 166-175.

________ (1996b), "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients,≅ A *Multivariate Behavioral Research*, 31 (1), 95-120.

_______ (2001), "A Suggested Standard Error for Interaction Coefficients in Latent Variable Regression," *2001 Academy of Marketing Science Conference Proceedings*, Miami: Academy of Marketing Science.

Podsakoff, P. M., W. D. Todor, R. A. Grover and V. L. Huber (1984), ASituational Moderators of Leader Reward and Punishment Behaviors: Fact or Fiction?≅ *Organizational Behavior and Human Performance*, 34, 21-63.

Rusbult, Caryl E. and Bram P. Buunk (1993), ACommitment Process in Close Relationships: An Interdependence Analysis,≅ *Journal of Social and Personal Relationships*, 10, 175-204.

Werts, C. E., R. L. Linn and K. G. Jöreskog (1974), AIntraclass Reliability Estimates: Testing Structural Assumptions,≅ *Educational and Psychological Measurement*, 34, 25-33.

Wong, S. K. and J. S. Long (1987), "Reparameterizing Nonlinear Constraints in Models with Latent Variables," Technical Report, Washington State University.

Table 1-- Simulated Data Sets Coefficient Estimates

			Pseudo LV Regression			Structural Equation Analysis		
Coefficient	Population Value[a]	Sample Size	Coefficient Average[b]	Bias[c]	RMSE[d]	Coefficient Average[b]	Bias[c]	RMSE[d]
$b_{Y,XZ} \neq 0$								
$\rho = .7$[e]								
$b_{Y,X}$	-0.15	100	-0.174	0.024	0.1490	-0.170	0.020	0.1510
		300	-0.163	0.013	0.0867	-0.162	0.012	0.0847
$b_{Y,Z}$	0.17	100	0.203	0.033	0.1340	0.196	0.026	0.1350
		300	0.191	0.021	0.0781	0.186	0.016	0.0801
$b_{Y,W}$	0.25	100	0.219	0.019	0.1810	0.205	0.005	0.1790
		300	0.211	0.011	0.1059	0.202	0.002	0.1069
$b_{Y,XZ}$	0.12	100	0.103	-0.017	0.1438	0.093	-0.027	0.1458
		300	0.110	-0.010	0.0916	0.109	-0.011	0.0896
$\rho = .9$[e]								
$b_{Y,X}$	-0.15	100	-0.137	-0.013	0.1221	-0.139	-0.011	0.1211
		300	-0.142	-0.008	0.0754	-0.140	-0.010	0.0764
$b_{Y,Z}$	0.17	100	0.159	-0.011	0.1085	0.157	-0.013	0.1075
		300	0.163	-0.007	0.0674	0.162	-0.008	0.0664
$b_{Y,W}$	0.25	100	0.190	-0.010	0.1483	0.193	-0.007	0.1493
		300	0.193	-0.007	0.0919	0.197	-0.003	0.0929
$b_{Y,XZ}$	0.12	100	0.131	0.011	0.0922	0.134	0.014	0.0912
		300	0.125	0.005	0.0649	0.139	0.019	0.0659
$\rho = .7$[e]								
$b_{Y,X}$	-0.35	100	-0.332	-0.018	0.1078	-0.331	-0.019	0.1088
		300	-0.340	-0.010	0.0639	-0.339	-0.011	0.0649
$b_{Y,Z}$	0.37	100	0.349	-0.021	0.0965	0.350	-0.020	0.0975
		300	0.358	-0.012	0.0570	0.360	-0.010	0.0560
$b_{Y,W}$	0.40	100	0.409	0.009	0.1313	0.407	0.007	0.1303
		300	0.401	0.001	0.0777	0.400	0.000	0.0787
$b_{Y,XZ}$	0.30	100	0.315	0.015	0.1056	0.316	0.016	0.1066
		300	0.308	0.008	0.0704	0.310	0.010	0.0694
$\rho = .9$[e]								
$b_{Y,X}$	-0.35	100	-0.355	0.005	0.0842	-0.354	0.004	0.0849
		300	-0.353	0.003	0.0520	-0.353	0.003	0.0512
$b_{Y,Z}$	0.37	100	0.377	0.007	0.0754	0.378	0.008	0.0762
		300	0.374	0.004	0.0471	0.375	0.005	0.0479
$b_{Y,W}$	0.40	100	0.398	-0.002	0.1016	0.404	0.004	0.1024
		300	0.400	0.000	0.0647	0.401	0.001	0.0640
$b_{Y,XZ}$	0.30	100	0.298	-0.002	0.0668	0.294	-0.006	0.0675
		300	0.300	0.000	0.0454	0.296	-0.004	0.0462

[a] Population coefficient values correspond to R^2's of .10 and .50 .
[b] Average over 100 replications.
[c] Difference between population coefficient value and coefficient average (Columns 4 or 7).
[d] Root mean squared error or >average= difference across 100 data sets between the coefficient estimates and the population value.
[e] Reliability.

Table 1 (Continued)-- Simulated Data Sets Coefficient Estimates

Coefficient	Population Value[a]	Sample Size	Pseudo LV Regression: Coefficient Average[b]	Pseudo LV Regression: Bias[c]	Pseudo LV Regression: RMSE[d]	Structural Equation Analysis: Coefficient Average[b]	Structural Equation Analysis: Bias[c]	Structural Equation Analysis: RMSE[d]
$b_{Y,XZ}$ = 0								
ρ = .7[e]								
$b_{Y,X}$	-0.15	100	-0.172	0.022	0.1470	-0.172	0.022	0.1450
		300	-0.165	0.015	0.0847	-0.164	0.014	0.0827
$b_{Y,Z}$	0.17	100	0.204	0.034	0.1350	0.197	0.027	0.1360
		300	0.193	0.023	0.0771	0.185	0.015	0.0761
$b_{Y,W}$	0.25	100	0.217	-0.033	0.1830	0.207	-0.043	0.1850
		300	0.212	-0.038	0.1079	0.204	-0.046	0.1099
$b_{Y,XZ}$	0.12	100	0.013	0.013	0.0469	0.012	0.012	0.0459
		300	0.009	0.009	0.0325	0.011	0.011	0.0345
ρ = .9[e]								
$b_{Y,X}$	-0.15	100	-0.138	-0.012	0.1231	-0.138	-0.012	0.1241
		300	-0.141	-0.009	0.0764	-0.139	-0.011	0.0774
$b_{Y,Z}$	0.17	100	0.158	-0.012	0.1075	0.156	-0.014	0.1065
		300	0.162	-0.008	0.0684	0.163	-0.007	0.0694
$b_{Y,W}$	0.25	100	0.191	-0.009	0.1493	0.194	-0.006	0.1503
		300	0.194	-0.006	0.0909	0.196	-0.004	0.0899
$b_{Y,XZ}$	0.12	100	0.009	0.009	0.0317	0.010	0.010	0.0327
		300	0.007	0.007	0.0226	0.008	0.008	0.0236
ρ = .7[e]								
$b_{Y,X}$	-0.35	100	-0.331	-0.019	0.1088	-0.330	-0.020	0.1098
		300	-0.339	-0.011	0.0629	-0.340	-0.010	0.0619
$b_{Y,Z}$	0.37	100	0.350	-0.02	0.0975	0.351	-0.019	0.0985
		300	0.357	-0.013	0.0560	0.359	-0.011	0.0550
$b_{Y,W}$	0.40	100	0.408	0.008	0.1323	0.408	0.008	0.1333
		300	0.402	0.002	0.0787	0.401	0.001	0.0797
$b_{Y,XZ}$	0.30	100	0.010	0.01	0.0362	0.011	0.011	0.0372
		300	0.007	0.007	0.0225	0.006	0.006	0.0215
ρ = .9[e]								
$b_{Y,X}$	-0.35	100	-0.354	0.004	0.0849	-0.353	0.0033	0.0857
		300	-0.354	0.004	0.0527	-0.352	0.0023	0.0535
$b_{Y,Z}$	0.37	100	0.378	0.008	0.0762	0.379	0.0087	0.0769
		300	0.375	0.005	0.0464	0.374	0.0043	0.0456
$b_{Y,W}$	0.40	100	0.399	-0.001	0.1024	0.405	0.0047	0.1031
		300	0.399	-0.001	0.0640	0.400	0.0003	0.0632
$b_{Y,XZ}$	0.30	100	0.006	0.006	0.0215	0.005	0.0053	0.0208
		300	0.004	0.004	0.0159	0.005	0.0048	0.0166

[a] Population coefficient values correspond to R^2's of .10 and .50 .
[b] Average over 100 replications.
[c] Difference between population coefficient value and coefficient average (Columns 4 or 7).
[d] Root mean squared error or >average= difference across 100 data sets between the coefficient estimates and the population value.
[e] Reliability.

Table 2-- Simulated Data Sets Coefficient Standard Error Results

			Pseudo LV Regression			Structural Equation Analysis		
Coeffic-ient	Population Value[a]	Sample Size	Average RMSE[b]	SE[c]	Bias[d]	Average RMSE[b]	Se[c]	Bias[d]
$b_{Y,XZ} = 0$								
$\rho = .7$[e]								
$b_{Y,X}$	-0.15	100	0.1490	0.1315	0.882	0.1510	0.1329	0.880
		300	0.0867	0.0813	0.938	0.0847	0.0813	0.938
$b_{Y,Z}$	0.17	100	0.1340	0.1180	0.881	0.1350	0.1190	0.882
		300	0.0781	0.0733	0.939	0.0801	0.0753	0.941
$b_{Y,W}$	0.25	100	0.1810	0.1589	0.878	0.1790	0.1575	0.880
		300	0.1059	0.0998	0.942	0.1069	0.1008	0.943
$b_{Y,XZ}$	0.12	100	0.1438	0.1211	0.842	0.1458	0.1226	0.841
		300	0.0916	0.0753	0.822	0.0896	0.0738	0.824
$\rho = .9$[e]								
$b_{Y,X}$	-0.15	100	0.1221	0.1122	0.919	0.1211	0.1114	0.920
		300	0.0754	0.0739	0.981	0.0764	0.0750	0.982
$b_{Y,Z}$	0.17	100	0.1085	0.0998	0.919	0.1075	0.0987	0.918
		300	0.0674	0.0661	0.981	0.0664	0.0651	0.980
$b_{Y,W}$	0.25	100	0.1483	0.1366	0.921	0.1493	0.1376	0.922
		300	0.0919	0.0900	0.979	0.0929	0.0911	0.980
$b_{Y,XZ}$	0.12	100	0.0922	0.0866	0.939	0.0912	0.0857	0.940
		300	0.0649	0.0579	0.891	0.0659	0.0588	0.892
$\rho = .7$[e]								
$b_{Y,X}$	-0.35	100	0.1078	0.0950	0.881	0.1088	0.0960	0.882
		300	0.0639	0.0600	0.939	0.0649	0.0610	0.940
$b_{Y,Z}$	0.37	100	0.0965	0.0850	0.881	0.0975	0.0860	0.882
		300	0.0570	0.0536	0.939	0.0560	0.0526	0.938
$b_{Y,W}$	0.40	100	0.1313	0.1154	0.879	0.1303	0.1146	0.880
		300	0.0777	0.0731	0.941	0.0787	0.0741	0.942
$b_{Y,XZ}$	0.30	100	0.1056	0.0877	0.831	0.1066	0.0887	0.832
		300	0.0704	0.0569	0.809	0.0694	0.0561	0.808
$\rho = .9$[e]								
$b_{Y,X}$	-0.35	100	0.0842	0.0775	0.921	0.0849	0.0783	0.922
		300	0.0520	0.0510	0.981	0.0512	0.0502	0.980
$b_{Y,Z}$	0.37	100	0.0754	0.0695	0.921	0.0762	0.0702	0.922
		300	0.0471	0.0462	0.979	0.0479	0.0469	0.980
$b_{Y,W}$	0.40	100	0.1016	0.0936	0.921	0.1024	0.0943	0.922
		300	0.0647	0.0634	0.979	0.0640	0.0626	0.979
$b_{Y,XZ}$	0.30	100	0.0668	0.0621	0.931	0.0675	0.0628	0.930
		300	0.0454	0.0400	0.881	0.0462	0.0407	0.882

[a] Population coefficient values correspond to R^2's of .10 and .50 .
[d] Root mean squared error or >average= difference across 100 data sets between the coefficient estimates and the population value.
[c] Average coefficient standard error over 100 replications.
[d] Average SE divided by RMSE. Values less than 1 indicate the SE is biased downward.
[e] Reliability.

Table 2 (Continued)-- Simulated Data Sets Coefficient Standard Error Results

Coefficient	Population Value[a]	Sample Size	Pseudo LV Regression RMSE[b]	Pseudo LV Regression Average SE[c]	Pseudo LV Regression Bias[d]	Structural Equation Analysis RMSE[b]	Structural Equation Analysis Average SE[c]	Structural Equation Analysis Bias[d]
$b_{Y,XZ}$ = 0								
ρ = .7[e]								
$b_{Y,X}$	-0.15	100	0.1470	0.1379	0.938	0.1450	0.1363	0.940
		300	0.0847	0.0836	0.988	0.0827	0.0815	0.986
$b_{Y,Z}$	0.17	100	0.1350	0.1270	0.941	0.1360	0.1281	0.942
		300	0.0771	0.0765	0.992	0.0761	0.0754	0.991
$b_{Y,W}$	0.25	100	0.1830	0.1724	0.942	0.1850	0.1739	0.940
		300	0.1079	0.1069	0.991	0.1099	0.1091	0.993
$b_{Y,XZ}$	0.12	100	0.0469	0.0445	0.949	0.0459	0.0437	0.951
		300	0.0325	0.0299	0.918	0.0345	0.0318	0.920
ρ = .9[e]								
$b_{Y,X}$	-0.15	100	0.1231	0.1207	0.981	0.1241	0.1216	0.980
		300	0.0764	0.0764	1.001	0.0774	0.0775	1.002
$b_{Y,Z}$	0.17	100	0.1075	0.1053	0.979	0.1065	0.1042	0.978
		300	0.0684	0.0684	0.999	0.0694	0.0694	1.000
$b_{Y,W}$	0.25	100	0.1493	0.1464	0.981	0.1503	0.1476	0.982
		300	0.0909	0.0910	1.001	0.0899	0.0899	1.000
$b_{Y,XZ}$	0.12	100	0.0317	0.0315	0.991	0.0327	0.0324	0.990
		300	0.0226	0.0211	0.931	0.0236	0.0220	0.932
ρ = .7[e]								
$b_{Y,X}$	-0.35	100	0.1088	0.1013	0.931	0.1098	0.1024	0.932
		300	0.0629	0.0617	0.981	0.0619	0.0607	0.980
$b_{Y,Z}$	0.37	100	0.0975	0.0908	0.931	0.0985	0.0918	0.932
		300	0.0560	0.0549	0.979	0.0550	0.0538	0.978
$b_{Y,W}$	0.40	100	0.1323	0.1232	0.931	0.1333	0.1240	0.930
		300	0.0787	0.0772	0.981	0.0797	0.0782	0.982
$b_{Y,XZ}$	0.30	100	0.0362	0.0341	0.941	0.0372	0.0350	0.942
		300	0.0225	0.0204	0.909	0.0215	0.0195	0.908
ρ = .9[e]								
$b_{Y,X}$	-0.35	100	0.0849	0.0833	0.981	0.0857	0.0841	0.982
		300	0.0527	0.0527	0.999	0.0535	0.0535	1.000
$b_{Y,Z}$	0.37	100	0.0762	0.0747	0.981	0.0769	0.0755	0.982
		300	0.0464	0.0464	1.001	0.0456	0.0456	1.000
$b_{Y,W}$	0.40	100	0.1024	0.1004	0.981	0.1031	0.1012	0.982
		300	0.0640	0.0639	0.999	0.0632	0.0631	0.999
$b_{Y,XZ}$	0.30	100	0.0215	0.0211	0.979	0.0208	0.0203	0.980
		300	0.0159	0.0146	0.921	0.0166	0.0153	0.922

[a] Population coefficient values correspond to R^2's of .10 and .50 .

[d] Root mean squared error or >average= difference across 100 data sets between the coefficient estimates and the population value.

[c] Average coefficient standard error over 100 replications.

[d] Average SE divided by RMSE. Values less than 1 indicate the SE is biased downward.

[e] Reliability.

Table 3-- Unadjusted Covariances for S, A, I, C, and E; with Reliabilities, Estimated Loadings (Λ=s), and Estimated Measurement Errors (θ)

	S	A	I	C	E	SxA	AxI
S	16.78816	-9.80047	7.58572	6.57148	-20.73284	53.51358	-7.73302
A	-9.80047	19.65213	-4.71621	-8.50311	20.44497	-55.59004	7.34272
I	7.58572	-4.71621	22.22471	14.73352	-8.08200	-7.733023	-2.69775
C	6.57148	-8.50311	14.73352	25.13758	-8.86009	5.52795	-3.685302
E	-20.73284	20.44497	-8.08200	-8.86009	66.64125	-117.17370	1.22814
SxA	53.51358	-55.59004	-7.73302	5.52795	-117.17370	857.81926	113.15626
AxI	-7.73302	7.34272	-2.69775	-3.68530	1.22814	113.15626	386.20813

	S	A	I	C	E
Reliabilities:	0.9325	0.9262	0.9271	0.9494	0.9622
Estimated Loadings (Λ=s)[a]	6.75962	5.77435	5.77716	4.87185	9.80917
Estimated Measurement Errors (θ)[a]	1.13320	1.45032	1.62018	1.27196	2.51903

[a] See p. 6.

Table 4-- Adjusted Covariances for S, A, I, C, and E, with Coefficient Estimates

	S	A	I	C	E	SxA	AxI
S	0.34261	-0.25108	0.19424	0.19954	-0.31268	0.20282	-0.03429
A	-0.25108	0.54589	-0.14137	-0.30225	0.36095	-0.24664	0.03811
I	0.19424	-0.14137	0.61735	0.52347	-0.14261	-0.03429	-0.01399
C	0.19954	-0.30225	0.52347	1.00550	-0.18540	0.02906	-0.02267
E	-0.31268	0.36095	-0.14261	-0.18540	0.66641	-0.30603	0.00375
SxA	0.20282	-0.24664	-0.03429	0.02906	-0.30603	0.55758	0.04549
AxI	-0.03429	0.03811	-0.01399	-0.02267	0.00375	0.04549	0.33604

Pseudo Latent Variable Regression Coefficient Estimates:

	S	A	I	C	SxA	AxI	
b_i	-0.571	0.335	-0.027	0.048	-0.192	-0.057	(Line 1
SE[a]	0.0969	0.0728	0.0726	0.0557	0.0751	0.0798	
t-value	-5.89	4.59	-0.37	0.86	-2.55	-0.71	
SEE[b]	0.59024						

OLS (Unadjusted) Regression Coefficient Estimates:

	S	A	I	C	SxA	AxI
b_i	-0.526	0.314	-0.037	0.034	-0.235	-0.018
SE[a]	0.0932	0.0700	0.0698	0.0535	0.0722	0.0767
t-value	-5.64	4.48	-0.53	0.64	-3.26	-0.24
SEE[b]	0.56737					
K[c]	1.04030					

LISREL 8 Coefficient Estimates:

	S	A	I	C	SxA	AxI	χ^2/df	GFI[d]	AGFI[d]	CFI[e]	RMSEA[f]
b_i	-0.553	0.333	-0.071	0.076	-0.160	-0.040					
SE[a]	0.1096	0.0767	0.0798	0.0624	0.0644	0.0823					
t-value	-5.04	4.34	-0.90	1.22	-2.49	-0.49	1519/575	.721	.677	.883	.086

---

[a] Coefficient standard error.
[b] Standard error of the estimate-- see p. 7.
[c] K is the adjustment factor used to obtain the pseudo latent variable coefficient SE=s, and is equal to the ratio of the SEE=s (i.e., 0.59024/ 0.56737) (see p. 7).
[d] Shown for completeness only-- GFI and AGFI may be inadequate for fit assessment in larger models (see Anderson and Gerbing 1984).
[e] .90 or better indicates acceptable fit (see McClelland and Judd 1993).
[f] .05 suggests close fit, .051-.08 suggests acceptable fit (Brown and Cudeck 1993, Jöreskog 1993).

Table 5-- SxA Interaction Significance

A-E Association				S-E Association			
S Value[a]	A Coef-ficient [b]	SE of A Coef-ficient[c]	t-value	A Value[d]	S Coef-ficient[e]	SE of S Coef-ficient[f]	t-value
1.20	0.90	0.24	3.76	1	-0.27	0.13	-2.04
2	0.75	0.18	4.02	2	-0.46	0.07	-6.15
3	0.56	0.12	4.42	2.56[g]	-0.57	0.06	-9.14
4	0.36	0.09	3.99	3	-0.65	0.07	-9.27
4.17[g]	0.33	0.09	3.67	4	-0.84	0.12	-6.79
5	0.17	0.11	1.58	5	-1.04	0.19	-5.37

---

[a] S ranged from 1.2 (=low) to 5 in the study.
[b] The coefficient of A is given by (.335-.192S) with S zero centered (i.e., S = Col. 1 - 4.16) (see Line 1 in Table 4).
[c] The Standard Error (SE) of the A coefficient is given by

$\sqrt{Var(b_A+b_{SxA}S)} = \sqrt{(Var(b_A)+S^2Var(b_{SxA})+2SCov(b_A,b_{SxA})}$,
where Var(a) is the squares of the Standard Errors (SE) of a at Line 1 of Table 4, and $Cov(b_A,b_{SxA}) = K*bcov(b_A,b_{SxA})$
where K is the SE adjustment factor (see Line 1 of Table 4) and bcov is the unadjusted covariance of b_A and b_{SxA} available in SPSS.
[d] A ranged from 1 (=low) to 5 in the study .
[e] The coefficient of S is given by (-.160+.276A) with A zero centered (i.e., A = Col. 5 - 2. 56) (see Line 5 of Table 2).
[f] The Standard Error (SE) of the S coefficient is given by

$\sqrt{Var(b_S+b_{SxA}A)} = \sqrt{Var(b_S)+A^2Var(b_{SxA})+2ACov(b_S,b_{SxA})}$,
where Var(a) is the squares of the Standard Errors (SE) of a at Line 5 of Table 2, and $Cov(b_S,b_{SxA}) = r*SE_{bS}*SE_{bSxA}$,
where r is the covariance of b_S and b_{SxA} available in SPSS.
[g] Mean value.

Appendix A-- Simulation Details

To produce the Table 1 and 2 results, the model

(A1) $$Y = b_1X + b_2Z + b_3W + b_4XZ + \zeta_Y$$

was estimated using simulated data sets that met the Kenny and Judd (1984) normality assumptions (indicators are multivariate normal with mean zero and independent of their measurement errors, and measurement errors are independent of each other), using the population parameters shown in Table A1. These parameters represent the original Kenny and Judd (1984) values for the variances of *X*, *Z*, and *W*, and polar but plausible values for model validation studies. For example, the loadings and measurement errors produced reliabilities of .7 (the minimum acceptable reliability in model validation studies) and .9, and the structural parameters (i.e., b=s and ζ=s) corresponded to R^2 's of .10 and .50.

X, *Z*, *W*, and their indicators x_1, ..., x_4, z_1, ..., z_4, w_1, ..., w_4 were created in data sets using PRELIS, and its normal random number generator. Each data set contained 100 or 300 cases and was replicated 100 times. Next, the values for x_1, ..., x_4 were summed (not averaged-- see p. 7) to form X in each case. vales for Z and W were added similarly, and the value of XZ (= X*Z) was added to each case. Y was determined using equation A1, the Table A1 population values, and PRELIS= random number generator (for ζ_Y). Then the Table A1 population parameters were used to generate the indicators of Y, y_1, ..., y_4, again using PRELIS= normal random number generator.

For each of the resulting data sets, the sample (unadjusted) covariance matrix for X, Z, W, Y and XZ was generated. Then these sample covariances were imported to an EXCEL spreadsheet, and the coefficient alphas for X, Z, W, and Y were calculated usung SPSS. Next these coefficient alphas were used in equations 2-5 with $\Lambda_! = n_!(\alpha_!)^2$ and $\Lambda_!^2 = n_!^2\alpha_!$, where ! is the latent variable *X*, *Z*, *W*, or *Y*, and $n_!$ is the number of indicators of ! (= 4 in this case-- $n_!$ is required to make $\Lambda_!$ and $^2\Lambda_!$ commensurate with Var(!)), to adjust the attenuated covariance matrix. Then this adjusted covariance matrix was exported to SPSS= matrix regression procedure.

The SPSS matrix regression procedure produced pseudo latent variable regression structural coefficients, coefficient standard errors, and a standard error of the estimate (SEE) (SEE = $\Sigma[y_i - \hat{y}_i]^2$, where y_i and $\hat{y}_i$. are observed and estimated y=s respectively). Then the unadjusted covariance matrix was input to SPSS=s matrix regression procedure to produce unadjusted coefficient standard errors and an unadjusted Standard Error of the Estimate (uSEE). Next the coefficient standard errors for the pseudo latent variable regression coefficients were computed as described on p. 7. Finally the raw data was input to LISREL 8 to produce structural coefficient estimates for comparison purposes.

Table A1-- Population Parameters for Simulated Data Sets

Parameter[a]	Population Variance	Population Coefficient
All Data Sets:		
X	2.15	
Z, Y	1.60	
W	1.00	
Corr(X,Z)		0.20
Corr(X,W)		0.20
Corr(Z,W)		0.20
b_0		0.00[b]
High Reliability Samples (ρ = .9):		
λ_{x1}		1.00
λ_{x2}-λ_{x4}		0.90
λ_{z1}	1.00	
λ_{z2}-λ_{z4}		0.90
λ_{w1}		1.00
λ_{w2}-λ_{w4}		0.90
ε_{x1}-ε_{x4}	0.82	
ε_{z1}-ε_{z4}	0.61	
ε_{w1}-ε_{w4}	0.38	
Low Reliability Samples (ρ = .7):		
λ_{x1}		1.00
λ_{x2}-λ_{x4}		0.70
λ_{z1}	1.00	
λ_{z2}-λ_{z4}		0.70
λ_{w1}		1.00
λ_{w2}-λ_{w4}		0.70
ε_{x1}-ε_{x4}	2.21	
ε_{z1}-ε_{z4}	1.65	
ε_{w1}-ε_{w4}	1.03	
Small Coefficients (R^2 = .10)		
ζ_Y	1.6	
$b_{Y,X}$		-0.15
$b_{Y,Z}$		0.17
$b_{Y,W}$		0.20
$b_{Y,XZ}$		0.12
Large Coefficients (R^2 = .50):		
ζ_Y	0.8	
$b_{Y,X}$		-0.35
$b_{Y,Z}$		0.37
$b_{Y,W}$		0.40
$b_{Y,XZ}$		0.30

[a] In $Y = b_{Y,X}X + b_{Y,Z}Z + b_{Y,W}W + b_{Y,XZ}XZ + \zeta_Y$ $x_i = \lambda_{xi}X + \varepsilon_{xi}$ $z_i = \lambda_{zi}Z + \varepsilon_{zi}$ $w_i = \lambda_{wi}W + \varepsilon_{wi}$.

[b] The indicators, including those for Y were zero centered.

Improving the Detection of Interactions in Selling and Sales Management Research

(An earlier, but revised, version of Ping 1996, "Improving the Detection of Interactions in Selling and Sales Management Research," *Journal of Personal Selling and Sales Management*, 16 (Winter), 53-64)
(Updated October, 2003)

Abstract

Because there is comparatively little guidance for substantive researchers in detecting interactions involving unobserved or latent variables in theory tests, the paper addresses these matters. After examining situations where including interactions might be appropriate, the paper describes detection techniques for these variables. Since structural equation analysis and errors-in-variables techniques are less accessible than regression in this application, the detection capabilities of several regression-based techniques are evaluated using Monte Carlo simulations.

Perhaps surprisingly, some of these techniques performed adequately in detecting interactions involving unobserved variables that were present in the population model, rejecting interactions that were not present in the population model, and not mistaking a quadratic in the population model for an interaction. Overall, product-term regression, and saturated product-term regression, followed by subgroup analysis and dummy variable regression detected true interactions better than ANOVA or the Chow test. These techniques also rejected spurious interactions better than the Chow Test. Overall, product-term regression, saturated - product-term regression, subgroup analysis, and dummy variable regression performed best at both tasks.

The paper also discusses characteristics of the data that appear to influence the detection of interactions involving unobserved variables and regression. These data characteristics include the presence of a quadratic in the population model and its being mistaken for an interaction, which has received no empirical attention to date. The effects of several data set characteristics are illustrated in the detection of an interaction between Role Clarity and Closeness of Supervision in their association with sales rep Satisfaction using a survey data set. The paper concludes with suggestions to improve the detection of interactions involving unobserved variables and regression that include mean centering, reporting multiple studies, and the use of a combination of detection techniques.

Introduction

In studies involving categorical independent variables (i.e., ANOVA studies), interactions are routinely estimated to aid in interpreting significant main effects. In studies involving continuous variables, interaction variables are also specified, although not routinely, and not to aid interpretation as they are in ANOVA. Typically continuous interactions are specified in response to theory that proposes their existence.

Researchers in the social sciences have called for the inclusion of interactions in models involving continuous variables (Aiken & West, 1991; Blalock, 1965; Cohen, 1968; Cohen & Cohen, 1975, 1983; Howard, 1989; Jaccard, Turrisi & Wan, 1990; Kenny, 1985). However, for variables measured with error such as unobserved variables, the options for detecting interactions have drawbacks. Regression is known to produce biased and inefficient coefficient estimates for variables measured with error (Bohrnstedt & Carter, 1971; Busemeyer & Jones, 1983). As a result, interaction detection techniques such as product-term regression and regression-based techniques involving sample splitting, such as subgroup analysis, will produce biased and inefficient coefficient estimates for unobserved variables.

While there have been several proposals to solve these problems (e.g., Warren, White & Fuller, 1974; Heise, 1986; Ping, 1995a) (see Feucht, 1989 for a summary), the proposed techniques lack significance testing statistics (Bollen, 1989), and are therefore inappropriate for theory tests. Nonlinear structural equation analysis (e.g., Kenny & Judd, 1984; Ping, 1995b,c) also shares this limitation for popular estimators such as Maximum Likelihood and Generalized Least Squares (Bollen, 1989; Kenny & Judd, 1984).[1]

This paper addresses this discouraging situation. After examining the influence that interactions can have on the interpretation of a model test, the paper summarizes the available approaches for detecting interactions. It then reports the results of an investigation, using Monte Carlo simulations, of the ability of regression-based approaches to detect true interactions, and reject spurious interactions involving unobserved variables. The paper discusses the effects the characteristics of the data have on the detection of these interactions using regression, and illustrates several of these effects using survey data. The paper concludes with suggestions for improved detection of these interactions.

We begin with a summary of the influence interactions can have on the interpretation of model tests involving continuous variables.

Interactions in Model Tests

Researchers include interactions in theory tests under several circumstances. The first arises when theory proposes the existence of interactions. The second occurs as part of the researcher's effort to improve the interpretation of significant main effects. Theories that propose interactions are ubiquitous in the Marketing literature (see for example Walker, Churchill & Ford 1977; and Weitz 1981 in the personal selling literature; Ajzen & Fishbein 1980, Engel, Blackwell & Kollat 1978, Howard 1977, and Howard & Sheth 1969 in the consumer behavior literature; Dwyer, Schurr & Oh 1987, and Stern & Reve 1980 in the channel literature; and Sherif & Hovland 1961 in the advertising literature). Researchers in Marketing have tested these and other proposed interactions involving continuous variables (see for example Batra & Ray 1986, Heide & John 1992, Kohli 1989, Laroche & Howard 1980, and Teas 1981). However, examples of the inclusion of continuous interactions reported as part of a researcher's efforts to reduce interpretational errors are rare.

Researchers who call for the investigation of continuous interaction and quadratic variables have argued that failing to do so increases the risk of false negative and conditional positive research findings. To demonstrate this, consider a model with linear terms only,

$$Y = b_0 + b_1X + b_2Z \; . \qquad (1$$

The Z-Y association in this model may be over- or understated because of the influence of an interaction of Z with X in the population model. When an XZ interaction is present in the population model, the actual coefficient of Z in equation (1) is given by

$$Y = b'_0 + b'_1X + b'_2Z + b_3XZ \; , \qquad (2$$

$$= b'_0 + b'_1X + (b'_2 + b_3X)Z \; . \qquad (3$$

In equation (3) the relationship between Z and Y varies with the values of X. For X values at one end of its range, it is possible for Z in equation (3) to have a stronger association with Y than it does in equation (1) (e.g., $b'_2 + b_3X$ is larger than b_2). For X values at the other end, it is possible for Z to have a weaker association with Y that it does in equation (1). It is also possible for Z to have a negative association with Y (i.e., $b'_2 + b_3X$ is negative) for X values near one end of its range, a positive association near the other end, and no association in between.

Perhaps more important for theory testing, the significance of the b_2 coefficient of Z in equation (1) could be different from the significance of the $b'_2 + b_3X$ coefficient of Z in equation (3). In particular,

o b_2 could be nonsignificant, while $b'_2 + b_3X$ could be significant over part(s) of the range of X, or

o b_2 could be significant while $b'_2 + b_3X$ could be nonsignificant over part of the range of X.

In the first situation, interpreting equation (1) could lead to a false disconfirmation of the Z-Y association. A nonsignificant linear variable in equation (1) may actually be significantly associated with the dependent variable over part of the range of an interacting variable in the population model. In the second situation, interpreting equation (1) could produce a misleading picture of the contingent Z-Y association. A significant linear effect (e.g., b_2) could actually be conditional in the population model, and nonsignificant for certain values of an interacting variable.

We now turn to the detection of interactions among unobserved variables.

Interaction Detection Techniques

Because studies in Marketing frequently involve unobserved variables with multiple observed variables measured with error, our discussion will involve variables in equations (1) and (2) that consist of sums of observed variables x_i and z_j, i.e.,

$$X=\sum_{i=1}^{k} x_i, \quad Z=\sum_{i=1}^{m} z_i, \quad Y=\sum_{i=1}^{n} y_i, \quad XZ=X*Z=(\sum_{i=1}^{k} x_i)(\sum_{i=1}^{m} z_i) \; , \qquad (4$$

or are specified as $V(x_i)= \lambda_i^2 V(X)+V(\varepsilon_i)$, $V(z_j)= \lambda_j^2 V(Z)+V(\varepsilon_j)$, or $V(x_i z_j)= \lambda_i^2\lambda_j^2 V(XZ)+\lambda_i^2 V(X)V(\varepsilon_j) +\lambda_j^2 V(Z)V(\varepsilon_i)$, where V(a) is the variance of a, and λ's and ε's are loadings and errors. The quadratic variable ZZ (= Z*Z) can be added to equation (1) or (2), and will be of interest later.

Approaches to detecting interactions among unobserved variables can be grouped into several general categories[2]: product indicator approaches, errors-in-variables approaches, product-term regression, and subgroup analysis. Product indicator approaches involve structural equation analysis, while errors-in-variables approaches typically involve regression using a moment matrix adjusted for measurement error. In product-term regression the dependent variable is regressed on variables comprised of summed observed variables and products of these summed variables (e.g., equations 2 and 4). Subgroup analysis involves splitting the sample and assessing differences in model coefficients when the model is restricted to the resulting subsets. Estimating these coefficient differences can be accomplished using regression, structural equation analysis, ANOVA, dummy variable regression, and the Chow test (Chow, 1960). We will discuss each of these approaches next.

Structural Equation Analysis

In product indicator/structural equation approaches, an interaction variable is specified using all possible products of the observed variables for the unobserved variables that comprise the interaction. For example if the unobserved variables X and Z have the observed variables x_1, x_2, z_1, and z_2, the indicators of the interaction XZ would be x_1z_1, x_1z_2, x_2z_1, and x_2z_2.[3] Structural coefficients (i.e., γ's and β's) can be estimated directly using the Kenny and Judd (1984) (see Jaccard & Wan, 1995) or Ping (1995b) techniques and software such as COSAN (available in SAS), or LISREL 8.[4] They can also be estimated indirectly using techniques such as the Hayduk (1987), Ping (1995c), or Wong and Long (1987) approaches and software such as CALIS (also available in SAS), EQS or LISREL 7.[5] However, these product indicator approaches produce model fit and structural coefficient significance statistics with Maximum Likelihood and Generalized Least Squares estimators that should be used with caution (Bollen, 1989; Jaccard & Wan, 1995; Kenny & Judd, 1984).

Regression Techniques

Typical of the errors-in-variables approaches are the Warren, White and Fuller (1974), Heise (1986), and Ping (1995a) proposals for adjusting the regression moment matrix to account for the errors in the variables (see Feucht, 1989 for a summary). The moment matrix (e.g., covariance matrix) produced by the sample data is adjusted using estimates of the errors. Regression estimates are then produced using this adjusted moment matrix in place of the customary unadjusted matrix. However, these approaches lack significance testing statistics (Bollen, 1989), and are not useful in theory tests.

In product-term regression (Blalock, 1965; Cohen, 1968) the dependent variable is regressed on the linear independent variables and one or more interactions formed as cross products of these linear independent variables (e.g., equation 2). The significance of the regression coefficient for the interaction variable (e.g., b_3) suggests the presence of an interaction between the components of this cross product variable (e.g., X and Z).

Subgroup analysis involves dividing the sample into subsets of cases based on different levels of a suspected interaction variable (e.g., low and high). The coefficients of the linear model (e.g., equation 1) are then estimated in each subset of cases using regression or structural equation analysis[6] (see Jöreskog, 1971). Finally, these coefficients are tested for significant differences using a coefficient difference test. A significant coefficient difference for a variable suggests an interaction between that variable and the variable used to create the subgroups.

Variations on this subgroup analysis theme include dummy variable regression and ANOVA. The ANOVA approach to detecting an interaction among continuous variables typically involves dicotomizing the independent variables in equation (1), frequently at their medians. This is accomplished by creating categorical variables that represent two levels of each independent variable (e.g., high and low), then analyzing these categorical independent variables using an ANOVA version of equation (2).

To use dummy variable regression (Cohen, 1968) to detect an interaction between X and Z in for example equation (1), the X (or Z) term of equation (1) is dropped, and dummy variables are added to create the regression model

$$Y = b''_0 + a_0 d + b''_2 Z + a_1 DZ \,, \qquad (5$$

where the dummy variable is defined as

$$D = 0 \text{ if } X_i < \text{the median of the values for X}$$
$$= 1 \text{ otherwise, } (i= 1,\ldots,\text{the number of cases})$$

and

$$DZ = D*Z .$$

The add and adz terms measure any difference in b''_0 and b''_2Z, respectively, when X is "high" and when it is "low." A significant coefficient for a dummy variable corresponding to an independent variable (e.g., a_1) suggests an interaction between that independent variable (e.g., Z) and the variable that produced the subsets (e.g., X).

Because of the potential drawbacks involving significance testing of product indicator/structural equation approaches and errors-in-variables techniques, we will restrict our attention to product-term regression and variations of subgroup analysis in the balance of the paper.

Population Models

For model tests there are several substantive matters that we have suggested should be addressed. One is the effect of failing to consider the possibility of an interaction in the population model. Others include failing to detect an interaction that is present in the population model (a true interaction), or mistakenly detecting an interaction that is absent in the population model (a spurious interaction).

These problems involving the detection of interactions could occur in several ways. An interaction could be detected using equation (2), when the population model contains no interaction and the population model is actually given by equation (1). In addition, the estimation of equation (2) could also produce a significant interaction coefficient (e.g., b_3) when the population model is given by

$$Y = b'''_0 + b''_1X + b'''_2Z + b_4ZZ . \qquad (6$$

This mistaking of a quadratic (e.g., ZZ) as an interaction has received no empirical attention to date, and was observed by Lubinski and Humphreys (1990). Finally, the estimation of equation (2) could produce a nonsignificant interaction coefficient (e.g., b_3) when there is an interaction in the population model and it is actually of the equation (2) form.

These matters will be examined next. We begin with the ability of the ANOVA approach, product-term regression, dummy variable regression, subgroup analysis, and the Chow test to detect an interaction that is actually present in the population model.

Detecting True Interactions

To gauge the ability of these regression techniques to detect an interaction that is present in the population model, we generated 100 data sets each containing 100 cases. The data sets were generated using the population model

$$Y = .5 -.15X + .35Z + .15XZ + e_Y , \qquad (7$$

and the population parameters shown in Table 1. These parameters produced variables that were normally distributed, and involved small interaction effects. The linear variables in these data sets (i.e., X and Z) were moderately correlated, and each had moderate reliability (ρ_x= .81 and ρ_z= .76). These characteristics were repeated in the other data sets used in this investigation, and the resulting data sets represent a somewhat average (i.e., neither favorable nor unfavorable) set of data characteristics for the detection of an interaction involving unobserved variables.

The population interaction term -.15XZ in equation (7) was estimated in each of the 100 data sets just described using each of the regression-based techniques of interest, beginning with the ANOVA approach.

ANOVA

Researchers have received little encouragement to use an ANOVA approach to detecting interactions between continuous variables. The approach is criticized in the Psychometric literature for its reduced statistical power that increases the likelihood of Type II (false negative) errors (Cohen, 1978; Humphreys & Fleishman, 1974; Maxwell, Delaney & Dill, 1984). Maxwell and Delaney (1993) showed that this approach can also produce Type I (false positive) errors. To gauge its false negative propensity we estimated equation (7) using the ANOVA approach and the 100 data sets just described. We expected the small population coefficient of XZ in equation (7), and the reduced statistical power of the ANOVA approach to combine to produce interaction detections at a chance level (e.g., 10%) for this technique.

X and Z in each of the 100 data sets were dicotomized at their medians. This was accomplished by resetting each observation for X, for example, to 0 if the observed value was less than the median of its data set values, and 1 otherwise. A two-way analysis of the main and XZ interaction effects of each of these 100 data sets using the ANOVA equivalent of equation (2) identified 50 of the 100 data sets in which the interaction effect was significant (see Table 2 line 1, column 1). These results will be discussed shortly.

Product-Term Regression

Regression involving variables measured with error produces coefficient estimates that are biased and inefficient (Bohrnstedt & Carter, 1971). Because product-term regression is based on regression, coefficient estimates for equation (2) using product-term regression are also biased and inefficient (Busemeyer & Jones, 1983). Since this bias is known to produce attenuated coefficient estimates, we expected that the weak -.15XZ interaction in the population model would be detected at a chance level only.

To test this anticipated result we added the cross product term XZ to each of the 100 data sets and estimated equation (2) using ordinary least squares regression. An R^2 difference test of XZ's incremental explained variance identified 81 of the 100 data sets in which there was a significant interaction (Table 2 column 1 shows this result as 100 - 81= 19, the number of data sets in which no significant interaction was identified).

Dummy Variable Regression

Dummy variable regression does not suffer from reduced statistical power as the ANOVA approach does, but it is a regression technique and it should therefore detect a weak population interaction such as -.15XZ at a chance level only. To test this, each of the 100 data sets was split at the median of X to create the dummy variable D in equation (5). Estimating the coefficients for equation (5) in each of the 100 data sets produced significant interactions in 69 data sets (see Table 2 column 1 for the number of nonsignificant interactions).

Subgroup Analysis

Turning to subgroup analysis, it too is criticized for its reduction of statistical power and increased likelihood of Type II error (Cohen & Cohen, 1983; Jaccard, Turrisi & Wan, 1990). We therefore expected that it would detect the weak -.15XZ population interaction by chance only. To test this expectation each data set was split at the median of X to produce two subsets, and X was dropped from equation (1) to create

$$Y = b_0 + b_2Z . \qquad (8$$

The coefficient of Z was then estimated in each subset, and coefficient difference tests (see Jaccard, Turrisi & Wan, 1990) for the Z coefficients between the pairs of subsets identified 75 of the 100 data sets in which there was a significant interaction (see Table 2 column 1 for the number of nonsignificant interactions).

Chow Test

A Chow test is used with dummy variable regression (see Dillon & Goldstein, 1984) and subgroup analysis to detect the presence of an interaction. Detecting an interaction using the Chow test involves comparing the total of the sum of squared errors associated with the estimation of equation (8)[7] in each subset, and the sum of squared error associated with estimating equation (8) using the full data set. If the Chow test suggests a significant sum of squared error difference, this in turn suggests that the Z coefficient for equation (8) in the full group is different from those in the subsets. Since these subsets were created by median splits of the cases using X, this is considered to be evidence of an interaction between X and Z.

Because it relies on splitting the sample, we also expected the Chow test to detect a weak population interaction such as -.15XZ by chance only. To test this expectation each of the 100 data sets was split at the median of X, and a Chow test was performed using equation (8) to determine if there were differences in the Z coefficients between the two subsets. The Chow test indicated significant interactions in 3 of the 100 data sets (see Table 2 column 1 for the number of nonsignificant interactions).

This completes the detection of the equation (7) population interaction using the regression-based techniques. All the techniques detected the weak -.15XZ population interaction at a rate that was higher than chance, except Chow test. We will complete the remaining tests and then discuss the results.

We now turn to rejecting interactions that are not present in the population model.

Rejecting Spurious Interactions

Detecting an interaction that is not in the population model can occur at least two ways. One involves the detection of a significant interaction using for example equation (2), when the population model contains no interaction variable and is of the form of equation (1). A second situation involves the presence of a quadratic term in the population model (e.g., equation 6) that is mistakenly detected as an interaction using for example equation (2) (see Lubinski & Humphreys, 1990).

No Population Interaction

We investigated both these possibilities, beginning with the first in which an interaction is specified using equation (2), for example, but an interaction is not present in the population model (e.g., the population model is of the equation 1 form).

We are aware of no theoretical or practical reason the above techniques should fail to reject an interaction that is not present in the population model. Accordingly we expected these approaches to detect a spurious interaction at approximately a chance level.

To test this expectation we generated 100 more data sets of 100 cases each, using the population model

$$Y = .5 + .35X + .35Z + e_Y \qquad (9$$

and the population parameters shown in Table 3, and tested for an interaction using an equation (2) model.

Repeating the procedures just described, we tested for the presence of an interaction in each of the 100 equation (9) data sets using the ANOVA approach, product-term regression, dummy variable regression, subgroup analysis, and the Chow test. We obtained the results shown in column 2 of Table 2. In summary, all the techniques except the Chow test detected spurious interactions at or below a chance rate. These techniques detected spurious interactions in from 3 to 8% of the samples, except for the Chow Test, which detected spurious interactions in 70 of the data sets.

A Population Quadratic

Next we investigated the detection of an interaction term using an equation (2) model in data sets generated using a population model that contained a quadratic term but no interaction (e.g., equation 6). We generated 100 more data sets of 100 cases each, using the population model

$$Y = .5 -.15X + .35Z + .15ZZ + e_Y \qquad (10$$

and the population parameters shown in Table 4, then tested for an interaction using an equation (2) model.

Lubinski and Humphreys (1990) suggested that product-term regression might mistake a quadratic variable in the population model for an interaction (see Cortina 1993). However, this possibility has not been investigated empirically. As a result, since their regression foundation relates all the detection techniques, we expected that the detection techniques would all mistake a quadratic in the population model for an interaction, and produce a significant interaction using an equation (2) model more frequently than by chance.[8]

We repeated the procedures described above using the additional equation (10) data sets, and the equation (2) model, along with the ANOVA approach, product-term regression, dummy variable regression, subgroup analysis, and the Chow test. We obtained the results shown in column 3 of Table 2. The ANOVA approach detected a spurious interaction in 10 of the 100 data sets. The Chow test detected no spurious interactions. The other techniques detected a spurious interaction in 19 to 30 of the 100 data sets.

These results are discussed next.

Interpretation

To help interpret these results we used the Table 2 detection frequencies to produce conditional probabilities of a true interaction, or lack of it, given the results in Table 2. The Table 5 entries for the ANOVA approach, for example, were calculated as follows. The "Significant" line for ANOVA shows the probability of a true interaction, given ANOVA produced a significant interaction effect. It was calculated by dividing the frequency that ANOVA detected a true interaction (100 - Column 1) by the frequency that it detected any interaction (100 - Column 1 + Column 2 + Column 3) (= 50/65= .77). Similarly, the "Non-Signif." line, the probability of no interaction in the population model given that ANOVA produced a nonsignificant interaction effect, was calculated by dividing the frequency that ANOVA rejected a spurious interaction (100 - Column 2 + 100 - Column 3) by the frequency that ANOVA rejected all interactions (Column 1 + 100 -Column 2 + 100 - Column 3) (= 185/235= .79).

Based on the conditional probabilities shown in column 2 of Table 5, product-term regression, followed by dummy variable regression, subgroup analysis, and the ANOVA approach, attained the highest probability of no interaction in the population model given none was detected (.90 to .79). The Chow Test performed the worst. It produced a probability of no interaction in the population model given none was detected of .57.

However, none of these techniques detected a true interaction as well as they rejected a spurious interaction (see column 1 of Table 5). The ANOVA approach, product-term regression, and dummy variable regression performed about the same in detecting a true interaction, and produced probabilities of a true interaction given one was detected of .77 to .76. Again, the Chow test did not perform well. It produced a probability of a true interaction given one was detected of .04.

Saturated Approaches

Examining Table 2 more closely, the low true interaction detection rates in column 1 of Table 5 were caused by the misdetections of a population quadratic as an interaction. Were it not for these misdetections, the Table 5 column 1 probabilities would have been in the .9 range, except for the Chow test. Following Lubinski and Humphrey's (1990) suggestion, we investigated adding X and Z quadratic terms (i.e., XX and ZZ) to equation (2) to create a saturated second order equation for estimation, i.e.,

$$Y = b''''_0 + b'''_1X + b''''_2Z + b'_3XZ + b'_4ZZ + b_5XX \ . \qquad (11$$

We re-ran the product-term regression procedure described earlier using an equation (11) model and the equation (10) data sets, and the results are shown in Table 6. The resulting conditional probabilities are shown in Table 7.

The resulting .88 probability of a true interaction, given one was detected using saturated product-term (equation 11) regression, was higher than the .76 probability using the equation (2) approach (see Table 5). In addition, saturated product-term regression produced a probability of no interaction in the population model, given none was observed, of .88.

Efficacy

Combining the results shown in Tables 5 and 7, saturated product-term regression, followed by ANOVA analysis, product-term regression, dummy variable regression, and subgroup analysis detected true interactions better than the Chow test. Product-term regression, followed by saturated product-term regression, subgroup analysis, dummy variable regression, and ANOVA analysis rejected spurious interactions better than the Chow Test.

Overall, product-term regression, saturated product-term regression, and dummy variable regression performed best. They exhibited the highest percentage of correct detections (see Table 8).

However, the Chow test performed the worst. It produced a probability of detecting a true interaction of .04, and rejecting a spurious interaction of .57. The Chow test appears to be unable to detect weak nonlinear variables in general. When we created 200 case versions of the 100 case samples, we observed the same lack of sensitivity to nonlinear variables with relatively small coefficients.

We now turn to other factors that influence the detection of interactions involving unobserved variables and regression.

Conditions that Influence Interaction Detection

The number and size of the continuous interactions reported in the social sciences have been small. In a single study of management variables Podsakoff, Todor, Grover and Huber (1984) examined 576 interactions and found 72 of them significant (12.5%). This detection frequency is only slightly above that of chance. Literature reviews in Marketing and Psychology report that observed interactions are typically small--accounting for 3 to 9% of the variance explained in Marketing studies (see Churchill, Ford, Hartley & Walker, 1985), and 1 to 3% of the variance explained in Psychological studies (see Aiken & West, 1991).

Research to date suggests that several characteristics of the data used to test a model can have deleterious effects on the detection of a true interaction. For example, McClelland and Judd (1993) showed that because field studies are similar to an ANOVA model with unbalanced data, field studies are less efficient than ANOVA models with balanced data in detecting interactions (see also Stone-Romero, Alliger & Aguinis, 1994). Authors have argued or shown that other characteristics of the data, such as reliability and multicollinearity between an interaction and its constituent variables can also affect the detection of an interaction. We will discuss these data characteristics and their impact on the detection of interactions involving unobserved variables next.

Reliability of the Independent and Dependent Variables

Low reliability in the independent variables reduces the observed size of the coefficient of a true interaction. Aiken and West (1991) observed that when reliabilities of the independent variables X and Z, and the dependent variable Y, all drop from 1 to .7, the observed interaction effect size for XZ is 33% of its true size. As a result, reduced reliability in first order variables also attenuates the R^2 contribution of an interaction containing those variables (Busemeyer & Jones, 1983).

In addition, low interaction reliability attenuates the standard errors of observed interactions, which can reduce the power of the test for an interaction (Aiken & West, 1991). However Dunlap and Kemery (1987) reported that for small samples (N=30) and reliabilities in X and Z of .8 or above, reasonable power of the test for XZ is maintained. In addition, larger samples can offset the loss of power of the test for XZ induced by low interaction reliability (Jaccard, Turrisi & Wan, 1990).

The reliability of XZ is a function of the reliability of X times the reliability of Z (see Bohrnstedt & Marwell, 1978; Busemeyer & Jones, 1983). For an uncorrelated X and Z that are bivariate normal, and have zero means, the reliability of XZ is the product of the reliabilities of X and Z. As the correlation between this X and Z increases, the reliability the interaction XZ increases (see Table 9). However, as the shaded area of Table 9 suggests, for X and Z with correlations in a range typical of survey data (0 to .5), the reliabilities of X and Z should generally be .7 or above to produce interaction reliabilities of .6 or more.

Systematic Error and Sample Size

Correlated/systematic error between the independent and dependent variables also attenuates the observed coefficient sizes of interactions (Evans, 1985). As a result, survey data gathered using a single questionnaire and/or the use of scales that are identical for the independent and dependent variables could produce an attenuated coefficient estimate for a true interaction.

As with first order variables, sample size and power are directly related in the detection of interactions. As a result in order to detect a weak population interaction, a relatively large sample size is required. For example in a model with an R^2 of .20 or .30, samples of 119 or 103 are required to detect an interaction that raises the model R^2 by 5 points to .25 or .35 (Jaccard, Turrisi & Wan, 1990) (see also McClelland & Judd, 1993).

Research Design

In an exploration of the difficulties of detecting predicted interactions using survey data and regression, McClelland and Judd (1993) noted that interactions are frequently detected in ANOVA studies. Comparing field studies and experiments, optimal experiments for detecting interactions can be described as requiring a balanced polar distribution for the independent variables. For two independent variables this distribution has polar treatment combinations with equal cell sizes. The most efficient of these McClelland and Judd characterized as a "four-cornered" data model (which has a three-dimensional frequency distribution that looks like the four legs on an upside-down kitchen table), and an "X-model" (which has a three-dimensional frequency distribution that resembles a bas-relief X anchored on the four polar cells). They showed that the interaction variance that remains after the main effects have been partialed out depends on the joint distribution of the linear variables comprising the interaction, and suggested the most efficient joint distributions were produced by four cornered or X models.

Because field studies produce mound shaped joint distributions for the independent variables they are similar to ANOVA models with unbalanced data, which are not as efficient at detecting interactions as balanced data models. As a result, McClelland and Judd argued that field studies may have observed comparatively few interactions to date because field studies are relatively less able to do so. For example comparing a four-cornered ANOVA data model involving two independent variables with typical field study bivariate distributions, they argued that the typical field study data distributions are only 6% to 10% as efficient at detecting interactions as the four-cornered ANOVA data model.

As a result in order to detect a population interaction, they suggested using field studies that were designed to oversample the extremes of the scales. Based on their results, a stratified sample that produces an approximately uniform frequency distribution for the two independent variables increases the efficiency of the interaction detection by a factor of between 2.5 and 4.

Intercorrelation

Interactions are usually highly correlated with their components (Blalock, 1979). This collinearity produces inflated standard errors for the linear variables (e.g., X and Z in equation 2) (Jaccard, Turrisi & Wan, 1990; see Aiken & West, 1991 for a demonstration). As a result for scales that have an arbitrary zero point, mean centering[9] is recommended to reduce this correlation (Aiken & West, 1991; Cronbach, 1987; Jaccard, Turrisi & Wan, 1990). This is accomplished by subtracting the mean of X, for example, from the value for X in each case. The result is a zero mean for X. Most rating scales used in the social sciences such as Likert scales would have an arbitrary zero point. As an aside, the independent variables in the data sets used in the study were mean centered.

Aiken and West (1991) showed that if X, Z and Y are multivariate normal with zero means, the covariance between XZ and Y is zero, and there is no interaction between XZ and Y. As a result, one view of an interaction is that it is the result of nonnormality in the data. This suggests that scale construction could influence nonnormality and the detection of interactions. Scales that produce data sets with distributions that are skewed, truncated (censored) (i.e., the frequency distributions are mound shaped with one end cut off), or have excess kurtosis (a peaked or flattened distribution) are nonnormal. As a result, using pretests it should be possible to design scale items that will alter the skewness, truncation or kurtosis in the resulting data and thereby increase the likelihood of detecting an interaction.

In an investigation of spurious interactions Lubinski and Humphreys (1990) pointed out that interactions and quadratics are usually correlated. As the correlation between X and Z approaches 1, the correlation between XX (or ZZ) and XZ also approaches 1. This is seen most easily for an X and Z that are bivariate normal with zero means:

$$
\begin{aligned}
r_{XX,XZ} &= \frac{C(XX,XZ)}{(V(XX)V(XZ))^{1/2}} \\
&= \frac{2V(X)C(X,Z)}{(2V(X)^2V(XZ))^{1/2}} \qquad \text{(see Kenny and Judd 1984)} \\
&= \frac{2^{1/2}C(X,Z)}{V(XZ)^{1/2}} \\
&= r_{x,z}\,\frac{2^{1/2}}{\left(\frac{V(XZ)}{V(X)V(Z)}\right)^{1/2}} \qquad (12 \\
&= K\, r_{x,z}\ ,
\end{aligned}
$$

where $r_{a,b}$ is the correlation between a and b, V(a) is the variance of a, C(a,b) is the covariance of a and b, and K is a constant equal to the terms following $r_{x,z}$ in equation (12). As a result they argued, and we have seen, that a population quadratic can be mistaken for an interaction. The Monte Carlo results for saturated product-term regression reported above support their argument that the quadratic combinations of the linear variables comprising an interaction should be entered before the interaction is entered and tested for significant incremental variance explained.

Spurious interactions have not been shown to be induced by reliability/measurement error or correlated errors between the predictor and criterion variables (see Aiken & West, 1991). In addition, the use of standardized coefficients does not affect the observed significance of the interaction (Jaccard, Turrisi & Wan, 1990).

We now illustrate the use of the detection techniques and the effects of several of the data characteristics using a survey data set.

A Survey Example

As part of a larger study of sales rep reactions to dissatisfaction with their organization (see Rusbult, Farrell, Rogers & Mainous, 1988), data were gathered concerning overall Satisfaction, Role Clarity, and Closeness of Supervision. Of interest were the relationships between Satisfaction and these antecedents (see Comer & Dubinsky, 1985), and the possibility that the reported variability in these associations may be due to an interaction between Role Clarity and Closeness of Supervision in their association with Satisfaction.

Since this is an illustration of the use of the detection techniques and data characteristics, the study will simply be sketched. Satisfaction (SAT), Role Clarity (RC), and Closeness of Supervision (CS) were measured using multiple item 5 point rating scales. The 204 survey responses were used to create summed variables for SAT, RC, and CS that were then mean centered. To test for the presence of an interaction between RC and CS, saturated product-term regression, product-term regression, subgroup analysis, dummy variable regression, ANOVA analysis, and the Chow test were used. Table 10 shows the zero-order correlations of the variables, and Table 11 shows the results of these analyses.

The product-term regression techniques and subgroup analysis suggested the presence of an RC-CS interaction (see Table 11), while dummy variable regression, ANOVA, and the Chow test suggested the absence of this interaction. Based on the Monte Carlo results reported earlier, we are inclined to discard the Chow test results as unreliable. The remaining 3 positive and 2 negative tests plus the positive product-term regression results suggest that there is an interaction between Closeness of Supervision and Role Clarity.

Several comments on these results seem warranted. The divergence of the results of the detection techniques underscores the desirability of using multiple techniques to detect interactions involving unobserved variables and regression. In addition, comparing the saturated product-term and the product-term regression results, the inclusion of a significant quadratic term in the model increased the size and the significance of the interaction coefficient. This suggests that improper specification of nonlinear effects (i.e., ignoring quadratics) may also attenuate the detection of a population interaction.

We varied several of the conditions in this data set to observe the effects on the interaction coefficient. We reran the product-term regression analysis using uncentered data, less reliable indicators, fewer cases, and a more balanced data model. The results are shown in Table 12. For example, the product-term regressions were rerun using uncentered data for CS and RC.[10] Comparing the product-term regression results using uncentered data with those in Table 11, the collinearity between the interaction and the linear variables severely attenuated the significance of the linear variables in the uncentered data, and the efficacy of mean centering is demonstrated.

Table 12 also shows the results of reduced reliability in the independent and dependent variables on the detection of an interaction involving unobserved variables with product-term regression. To obtain these results a random amount of error was added to each indicator of SAT, RC and CS for each case, to produce reliabilities for these variables of .6. Comparing Tables 11 and 12, reducing the reliability of SAT, RC and CS from approximately .8 to .6 produced an interaction coefficient that was not significant. The significance of the coefficients for CS and RC, however, were relatively unaffected by this reliability reduction. This illustrates the multiplicative effect of the reliability of linear variables on an interaction comprised of these variables. CS and RC had reliabilities of .6 while RCxCS had a reliability of .38 ($\rho_{XZ} = (r^2_{X,Z} + \rho_X \rho_Z)/(r^2_{X,Z} + 1)$, where ρ_a is the reliability of a and $r_{X,Z}$ is the correlation between X and Z (=.207); see Busemeyer and Jones 1983).

To gauge the effect of a small sample, we drew a subsample of 50 cases from the data set. Comparing Tables 11 and 12, the small number of cases adversely affected the standard error of the interaction, and produced an interaction that was not significant.

We also altered the weights on the cases to approximate the results of an experiment with a balanced number of cases in each cell. The result was a more-nearly balanced data model in which the distributions of CS and RC were adjusted to approximate bivariate uniform distributions. First the ranges of CS and RC were divided into five intervals or "cells," and the resulting 5x5 matrix of cells was used to increase the weights of the polar cells (i.e., cells (1,1) (1,5), (5,1) and (5,5)) . Next the cases in all the cells were initially weighted by 1, and the size of the polar cell with the most cases was determined. Then the cases in the other three polar cells were weighted by the ratio of the largest polar cell size to their (smaller) cell size. The resulting heavier weights on the cases in the smaller polar cells approximated an oversampling of these polar cells, and produced a more nearly balanced data model.[11]

The results of this polar weighting for the reduced reliability and small sample data sets are shown in Table 12. In both situations the effects of unreliability or sample size on the significance of the RCxCS interaction was reversed. The heavier weights on the polar cases increased the interaction coefficient and decreased its standard error. However, we tried this polar weighting approach on the small sample, adding reduced reliability, and the interaction was not significant.

In summary these survey data results were generally consistent with the predicted results of the detection technique used, data reliability, mean centering, small data sets, and balanced data on the detection of an interaction involving unobserved variables and regression. In addition, the observed interaction was small, and its incremental R^2 contribution (1.7%) was consistent with the incremental variance explained by interactions observed in previous studies (see Table 11).

Implications

In order to increase the likelihood of detecting a true interaction in a field study, the number of cases in the data set should be relatively large. Based on Jaccard, Turrisi and Wan's (1990) results and R^2's typical of survey research, data sets of 100 or more may be appropriate. In addition, the reliability of the independent and dependent variables should be high. Based on Dunlap and Kemery's (1987) results and Table 9, reliabilities below .7 should be avoided.[12] Further, the independent variables should be mean centered. Finally, the study should use a different method to measure the independent variables from that used to measure the dependent variable. Perhaps at a minimum, a different scale should be used for the independent variables from that used for the dependent variables.

As the study results suggested, the techniques used to analyze the data are also important for detecting interactions. The field survey results, the greater-than-chance error rates implied by Table 8, and the frequency of misdetection of a population quadratic as an interaction suggest that the use of a combination of detection techniques may be appropriate for detecting true interactions and rejecting spurious interactions. The Monte Carlo results suggested that saturated product-term regression and product-term regression, and dummy variable regression, subgroup analysis, and ANOVA may perform best in this application. The field study results suggested that a combination of the two product-term techniques could be used, followed by a combination of dummy variable regression, a subgroup analysis, and ANOVA, particularly if the product-term regression tests provide inconsistent results.

Based on McClelland and Judd's (1993) results, to more nearly approximate a balanced data model in a field study and thereby improve its ability to detect an interaction, polar responses could be oversampled. As they point out, this suggestion is controversial because it creates interpretational difficulties.[13] In order to show the presence or absence or an interaction, an alternative to oversampling would be to report two studies: the field study, and an additional study involving an experiment using the survey instrument and a balanced data design.[14] If the independent variables are measured with the survey instrument in this experiment, their distributions would have properties that more nearly approximated a four-cornered data model. This in turn would increase the efficiency of a product-term regression analysis using the experimental data, and presumably the efficiency of the other regression-based detection techniques investigated in this paper.

Finally, while the Tables 5 and 7 detection probabilities were not unity, they were not unacceptably low. However, a study's detection frequency for interactions involving unobserved variables could be lower than those observed in this study if it combines a poor choice of detection technique, low reliability, fewer cases, omission of mean centering, and few polar responses. But with care, especially regarding the choice of detection technique, the use of multiple techniques, mean centering, reliability, and a more-nearly balanced data model, the detection frequency a researcher actually experiences in a study involving these variables could be higher than those observed in this study.

Summary

The paper has addressed interactions involving unobserved variables. It began with the case for including interactions in theory tests, even when theory is silent on their existence, to avoid false negative and conditional positive interpretations of the theory test results.

The paper commented on the efficacy of the available detection techniques for interactions involving unobserved variables. It then observed the actual performance of popular regression-based techniques in detecting interactions involving unobserved variables, with several interesting results. Of the five techniques studied, dummy variable regression, subgroup analysis, ANOVA analysis, and product-term regression mistook a weak population quadratic variable for an interaction at a level higher than chance, as Lubinski and Humphreys' (1990) results implied. However, saturated product-term regression, product-term regression, dummy variable regression, subgroup analysis, and ANOVA performed acceptably in detecting a true interaction and rejecting a spurious interaction. The Chow test performed the worst in this task. There was evidence that the Chow test is insensitive to small nonlinear effects.

The paper discussed the characteristics of the data used to test a model, such as the existence of a quadratic in the population model, reliability, the number of cases, mean centering, and systematic error, nonnormality, and their effect on the detection of interactions involving unobserved variables. Several of these effects were illustrated in an analysis of a survey data set.

In order to improve the likelihood of detecting a true interaction involving unobserved variables and field survey data, the paper suggested that a combination of detection techniques should be used, beginning with saturated product-term regression. It also suggested reporting regression results from an experiment with balanced data, along with the field survey results. In addition, the independent variables should be mean centered, the study should use a different method to measure the independent variables from that used to measure the dependent variable, the reliability of the independent and dependent variables should be high, and the number of cases to be analyzed should be relatively large.

References

Aiken, L.S. & S.G. West (1991), *Multiple Regression: Testing and Interpreting Interactions*, Newbury Park, CA: SAGE Publications.

Ajzen, I. & M. Fishbein (1980), *Understanding Attitudes and Predicting Social Behavior*, Englewood Cliffs, NJ: Prentice Hall.

Anderson, T.W. & Y. Amemiya (1985), "The Asymptotic Normal Distribution of Estimators in Factor Analysis under General Conditions," Technical Report 12, Econometric Workshop, Stanford U.

Anderson, T.W. & Y. Amemiya (1986), "Asymptotic Distribution in Factor Analysis and Linear Structural Relations, " Technical Report 18, Econometric Workshop, Stanford U.

Anderson, J. C., & Gerbing, D. W. (1982), "Some Methods for Respecifying Measurement Models to Obtain Unidimensional Construct Measurement," *Journal of Marketing Research*, 25, 186-92.

Batra, R. & M.L. Ray (1986), "Affective Responses Mediating Acceptance of Advertising," *Journal of Consumer Research*," 13 (September), 249.

Blalock, H.M., Jr. (1965), "Theory Building and the Concept of Interaction," *American Sociological Review*, 30, 374-81.

Blalock, H.M., Jr. (1979), *Social Statistics*, New York: McGraw Hill.

Bohrnstedt, G.W. & Carter T.M. (1971), "Robustness in Regression Analysis," in *Sociological Methodology*, H.L. Costner ed., San Francisco: Jossey-Bass 118-46.

Bohrnstedt, G.W. & G. Marwell (1978), "The Reliability of Products of Two Random Variables," in K.F. Schuesser ed., *Sociological Methodology*, San Francisco: Jossey-Bass.

Bollen, K.A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.

Boomsma, A. (1983), *On the Robustness of LISREL (Maximum Likelihood Estimation) Against Small Sample Size and Nonnormality*, Unpublished Dissertation, University of Groningen.

Browne, M.W. (1987), "Robustness of Statistical Inference in Factor Analysis and Related Models," *Biometrika*, 74, 375-84.

Busemeyer, J.R. & L.E. Jones (1983), "Analysis of Multiplicative Combination Rules When the Causal Variables are Measured With Error," *Psychological Bulletin*, 93, 549-62.

Chow, G.C. (1960), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, Vol. 28, No. 3.

Churchill, G.A., Jr. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research*, 16 (February), 64-73.

Churchill, G.A., Jr. & J.P. Peter (1984), "Research Design Effects on the Reliability of Rating Scales: A Meta Analysis," *Journal of Marketing Research*, 21, (November), 360-375.

Churchill, G.A., Jr., N.M. Ford, S.W. Hartley & O.C. Walker, Jr. (1985), "The Determinants of Salesperson Performance: A Meta-Analysis," *Journal of Marketing Research*, 22 (May), 103-118.

Cohen, J. (1968), "Multiple Regression as a General Data-Analytic System," *Psychological Bulletin*, 70, 426-43.

Cohen, J. (1978), "Partialed Products Are Interactions; Partialed Powers Are Curve Components," *Psychological Bulletin*, 85, 858-66.

Cohen, J. & P. Cohen (1975), *Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences*, Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J. & P. Cohen (1983), *Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences*, Hillsdale, NJ: Lawrence Erlbaum.

Comer, J.M. & A. J. Dubinsky (1985), *Managing the Successful Sales Force*, Lexington Press, Inc., Lexington, MA.

Cortina, J.M. (1993), "Interaction, Nonlinearity, and Multicollinearity: Implications for Multiple Regression," *Journal of Management*, 19 (4), 915-922.

Cronbach, L.J. (1987), "Statistical Tests for Moderator Variables: Flaws in Analysis Recently Proposed," *Psychological Bulletin*, 102, 414- 417

Dillon, W.R. & M. Goldstein (1984), *Multivariate Analysis Methods and Applications*, New York: Wiley.

Dunlap, W.P. & E. Kemery (1987), "Failure to Detect Moderating Effects: Is Multicollinearity the Problem?" *Psychological Bulletin*, 102, 418-420.

Dwyer, F.R., P.H. Schurr, & S. Oh 1987, "Developing Buyer-Seller Relationships," *Journal of Marketing*, 51 (April): 11-27.

Engel, J.F., R.D. Blackwell & D.T. Kollat (1978), *Consumer Behavior*, 3rd. ed., Hinsdale, IL: Dryden.

Evans, M.T. (1985), "A Monte Carlo Study of the Effects of Correlated Methods Variance in Moderated Multiple Regression Analysis," *Organizational Behavior and Human Decision Processes*, 36, 305-323.

Feucht, T.E (1989), "Estimating Multiplicative Regression Terms in the Presence of Measurement Error," *Sociological Methods and Research*, 17 (3), 257-82.

Fornell, C. & D. F. Larker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.

Gerbing, D.W. & J.C. Anderson (1985), "The Effects of Sampling Error and Model Characteristics on Parameter Estimation for Maximum Likelihood Confirmatory Factor Analysis," *Multivariate Behavioral Research*, 20, 255-271.

Gerbing, D. W. & J. C. Anderson (1988), "An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment," *Journal of Marketing Research*, 25 (May), 186-192.

Harlow, L.L. (1985), *Behavior of Some Elliptical Theory Estimators with Nonnormal Data in a Covariance Structures Framework: A Monte Carlo Study*, Unpublished Dissertation, University of California, Los Angeles.

Hayduk, L.A. (1987), *Structural Equation Modeling with LISREL: Essentials and Advances*, Baltimore, MD: Johns Hopkins Press.

Hedges, L.V. & I. Olkin (1985), *Statistical Methods for Meta Analysis*, NY: Academic Press.

Heide, J.B. & G. John (1992), "Do Norms Matter in Marketing Relationships?" *Journal of Marketing*, 56 (April), 32-44.

Heise, D.R. (1986), "Estimating Nonlinear Models Correcting for Measurement Error," *Sociological Methods and Research*, Vol. 14 No. 4 (May), 447-72.

Howard, J.A. (1977), *Consumer Behavior: Application of Theory*, New York: McGraw Hill.

Howard, J.A. (1989), *Consumer Behavior in Marketing Strategy*, Englewood Cliffs, NJ: Prentice Hall.

Howard, J.A. & J.N. Sheth (1969), *The Theory of Buyer Behavior*, New York: Wiley.

Hu, L., P.M. Bentler & Y. Kano (1992), "Can Test Statistics in Covariance Structure Analysis Be Trusted?" *Psychological Bulletin*, 112, 351-362.

Humphreys, L.G. & A. Fleishman (1974), "Pseudo-orthogonal and other Analysis of Variance Designs Involving Individual-Differences Variables," *Journal of Educational Psychology*, 66, 464-72.

Jaccard, J. & C.K. Wan (1995), "Measurement Error in the Analysis of Interaction Effects Between Continuous Predictors Using Multiple Regression: Multiple Indicator and Structural Equation Approaches," *Psychological Bulletin*, 117 (2), 348-357.

Jaccard, J., R. Turrisi & C.K. Wan (1990), *Interaction Effects in Multiple Regression*, Newbury Park, CA: SAGE Publications.

Jöreskog, K.G. (1971), "Simultaneous Factor Analysis in Several Populations," *Psychometrika*, 57, 409-26.

Jöreskog, K.G. (1993), "Testing Structural Equation Models," in *Testing Structural Equation Models*, K.A. Bollen and J.S. Long eds., Newbury Park, CA: SAGE.

Kenny, D.A. (1985), "Quantitative Methods for Social Psychology," in *Handbook of Social Psychology*, 3rd Ed., Vol. 1, G. Lindzey and E. Aronson eds., New York: Random House.

Kenny, D. & C.M. Judd (1984), "Estimating the Nonlinear and Interactive Effects of Latent Variables," *Psychological Bulletin*, 96, 201-10.

Kohli, A.K. (1989), "Effects of Supervisory Behavior: The Role of Individual Differences Among Salespeople," *Journal of Marketing*, 53 (October), 40-50.

Lubinski, D. & L.G. Humphreys (1990), "Assessing Spurious 'Moderator Effects': Illustrated Substantively with the Hypothesized ('Synergistic') Relation Between Spatial and Mathematical Ability," *Psychological Bulletin*, 107, 385-93.

Maxwell, S.E., Delaney, H.D. & C.A. Dill (1984), "Another Look at ANCOVA Versus Blocking," *Psychological Bulletin*, 95, 136-47.

Maxwell, S.E. & H.D. Delaney (1993), "Bivariate Median Splits and Spurious Statistical Significance," *Psychological Bulletin*, 113, 181-90.

McClelland, G.H. & C.M. Judd (1993), "Statistical Difficulties of Detecting Interactions and Moderator Effects," *Psychological Bulletin*, 114 (2), 376-390.

Netemeyer, R.G., M.W. Johnson, & S. Burton (1990), "Analysis of Role Conflict and Role Ambiguity in a Structural Equations Framework," *Journal of Applied Psychology*, 75 (2), 148-157.

Ping, R.A. (1995a), "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients," Conditionally accepted for the *Journal of Multivariate Behavioral Research*.

Ping, R.A. (1995b), "A Parsimonious Estimating Technique for Interaction and Quadratic Latent Variables," *Journal of Marketing Research*, forthcoming.

Ping, R.A. (1995c), "Latent Variable Interaction and Quadratic Effect Estimation: A Two-Step Technique Using Structural Equation Analysis," Conditionally accepted for the *Psychological Bulletin*.

Podsakoff, P.M., W.D. Todor, R.A. Grover & V.L. Huber (1984), "Situational Moderators of Leader Reward and Punishment Behaviors: Fact or Fiction?" *Organizational Behavior and Human Performance*, 34, 21-63.

Rusbult, C.E., D. Farrell, G. Rogers and A.G. Mainous III (1988), "Impact of Exchange Variables on Exit, Voice, Loyalty, and Neglect: An Integrative Model of Responses to Declining Job Satisfaction," *Academy of Management Journal*, 31 (September), 599-627.

Sharma, S., S. Durvasula & W.R. Dillon (1989), "Some Results on the Behavior of Alternative Covariance Structure Estimation Procedures in the Presence of Nonnormal Data," *Journal of Marketing Research*, 26 (May), 214-21.

Sherif, M. & C.I. Hovland (1961), *Social Judgement: Assimilation and Contrast Effects in Communication and Attitude Change*, New Haven: Yale University Press.

Stone-Romero, E.F., Alliger, G.M. & H. Aguinis (1994), "Type II Error Problems in the Use of Moderated Multiple Regression for the Detection of Moderating Effects of Dichotomous Variables," *Journal of Management*, 20 (1), 167-178.

Stern, L.W. & T. Reve (1980), "Distribution Channels as Political Economies: A Framework for Comparative Analysis," *Journal of Marketing*, 44 (Summer), 52-64.

Tanaka, J.S. (1984), *Some Results on the Estimation of Covariance Structure Models*, Unpublished Dissertation, University of California, Los Angeles.

Tate, R.L. (1984), "Limitations of Centering for Interactive Models," *Sociological Methods and Research*, 13, 251-271.

Teas, R.K. (1981), "An Empirical Test of Models of Salespersons' Job Expectancy and Instrumentality Perceptions," *Journal of Marketing Research*, 18 (May), 209-26.

Walker, O.C., G.A. Churchill & N.M. Ford (1977), "Motivation and Performance in Industrial Selling: Present Knowledge and Needed Research," *Journal of Marketing Research*, 14 (May), 156-68.

Warren, R.D., J.K. White & W.A. Fuller (1974), "Errors in Variables Analysis of Managerial Role Performance," *Journal of the American Statistical Association*, 69, 886-93.

Weitz, B.A. (1981), "Effectiveness in Sales Interactions: A Contingency Framework," *Journal of Marketing*, 45 (Winter), 85-103

Wong, S.K. & J.S. Long (1987), "Reparameterizing Nonlinear Constraints in Models with Latent Variables," Technical Report, Washington State University.

Table 1--Population Parameters

	Population	
Parameter[a]	Variance	Coefficient
X	2.15	
Z	1.60	
Corr(X,Z)		0.20
e_{x1} 0.36		
e_{x2} 0.81		
e_{z1} 0.49		
e_{z2} 0.64		
e_Y	0.64	

[a] $Y = .5 - .15X + .35Z + .15XZ + e_Y$,
$X = (x_1 + x_2)/2$,
$x_1 = .9*X + e_{x1}$,
$x_2 = .6*X + e_{x2}$,
$Z = (z_1 + z_2)/2$,
$z_1 = .8*Z + e_{z1}$,
$z_2 = .7*Z + e_{z2}$, and
$XZ = X*Z$.

Table 2--Detection Results

Technique	Frequency: Interaction in the Population Equation, and No Interaction Detected	Interaction Not in the Population Equation, and an Interaction Detected	Quadratic in the Population Equation, and an Interaction Detected
ANOVA Analysis	50	5	10
Product-Term Regression	19	3	22
Dummy Variable Regression	31	3	19
Subgroup Analysis	25	8	30
Chow Test	97	70	0

Table 3--No Interaction Term Population Parameters

	Population	
Parameter[a]	Variance	Coefficient
X	2.15	
Z	1.60	
Corr(X,Z)		0.20
e_{x1} 0.36		
e_{x2} 0.81		
e_{z1} 0.49		
e_{z2} 0.64		
e_Y	0.16	

[a] $Y = .5 + .35X + .35Z + e_Y$,
$X = x_1 + x_2$,
$x_1 = .9*X + e_{x1}$,
$x_2 = .6*X + e_{x2}$,
$Z = z_1 + z_2$,
$z_1 = .8*Z + e_{z1}$, and
$z_2 = .7*Z + e_{z2}$.

Table 4--Quadratic Term Population Parameters

	Population	
Parameter[a]	Variance	Coefficient
X	2.15	
Z	1.60	
Corr(X,Z)		0.20
e_{x1} 0.36		
e_{x2} 0.81		
e_{z1} 0.49		
e_{z2} 0.64		
e_Y	0.64	

[a] $Y = .5 - .15X + .35Z + .15ZZ + e_Y$,
$X = (x_1 + x_2)/2$,
$x_1 = .9*X + e_{x1}$,
$x_2 = .6*X + e_{x2}$,
$Z = (z_1 + z_2)/2$,
$z_1 = .8*Z + e_{z1}$,
$z_2 = .7*Z + e_{z2}$, and
$ZZ = Z*Z$.

Table 5--Conditional Probabilities

Technique and Test Result	P(Interaction |Test Result)	P(No Interaction |Test Result)
ANOVA Analysis:		
Significant	.77	
Non-Signif.		.79
Product-Term Regression		
Significant	.76	
Non-Signif.		.90
Dummy Variable Regression:		
Significant	.76	
Non-Signif.		.85
Subgroup Analysis:		
Significant	.66	
Non-Signif.		.87
Chow Test:		
Significant	.04	
Non-Signif.		.57

Table 6--Results for Saturated Regression Estimation

Technique	Frequency: Interaction in the Population Equation, and No Interaction Detected	Interaction Not in the Population Equation, and an Interaction Detected	Quadratic in the Population Equation, and an Interaction Detected
Saturated Product-Term Regression	25	0	10

Table 7--Conditional Probabilities for Saturated Model Estimation

Test Result	P(Interaction |Test Result)	P(No Interaction |Test Result)
Saturated Product-Term Regression		
Significant	.88	
Non Signif.		.88

Table 8--Percent Correct Detections

Technique	% Correct[a]
ANOVA Analysis	78
Product-Term Regression	85
Dummy Variable Regression	82
Subgroup Analysis	79
Chow Test	44
Saturated Product-Term Regression	80

[a] (300 - Table 2/6 Column 1 - Table 2/6 Column 2 - Table 2/6 Column 3)/300

Table 9[a]--Interaction Reliabilities[b] for Selected Constituent Variables[c] Reliabilities and Intercorrelations

	Reliability										
	X:	0.9				0.8			0.7		0.6
	Z:	0.9	0.8	0.7	0.6	0.8	0.7	0.6	0.7	0.6	0.6
$r_{X,Z}$											
0		0.81	0.72	0.63	0.54	0.64	0.56	0.48	0.49	0.42	0.36
0.1		0.81	0.72	0.63	0.54	0.64	0.56	0.49	0.50	0.43	0.37
0.15		0.81	0.73	0.64	0.55	0.65	0.57	0.49	0.50	0.43	0.37
0.2		0.82	0.73	0.64	0.56	0.65	0.58	0.50	0.51	0.44	0.38
0.25		0.82	0.74	0.65	0.57	0.66	0.59	0.51	0.52	0.45	0.40
0.3		0.83	0.74	0.66	0.58	0.67	0.60	0.52	0.53	0.47	0.41
0.35		0.83	0.75	0.67	0.59	0.68	0.61	0.54	0.55	0.48	0.43
0.4		0.84	0.76	0.68	0.60	0.69	0.62	0.55	0.56	0.50	0.45
0.45		0.84	0.77	0.69	0.62	0.70	0.63	0.57	0.58	0.52	0.47
0.5		0.85	0.78	0.70	0.63	0.71	0.65	0.58	0.59	0.54	0.49
0.55		0.85	0.79	0.72	0.65	0.72	0.66	0.60	0.61	0.55	0.51
0.6		0.86	0.79	0.73	0.66	0.74	0.68	0.62	0.63	0.57	0.53
0.65		0.87	0.80	0.74	0.68	0.75	0.69	0.63	0.64	0.59	0.55
0.7		0.87	0.81	0.75	0.69	0.76	0.70	0.65	0.66	0.61	0.57
0.75		0.88	0.82	0.76	0.71	0.77	0.72	0.67	0.67	0.63	0.59
0.8		0.88	0.83	0.77	0.72	0.78	0.73	0.68	0.69	0.65	0.61
0.85		0.89	0.84	0.79	0.73	0.79	0.74	0.70	0.70	0.66	0.63

[a] The shading depicts combinations of X and Z reliabilities that produce XZ reliabilities of .6 or above for the typical range of correlations observed in field studies (0 to .5).
[b] $_{XZ} = (r_{X,Z}^2 + \ _X\ \ _Z)/(r_{X,Z}^2 + 1)$, where $r_{X,Z}$ is the correlation between X and Z, and $_a$ is the reliability of a.
[c] X and Z are bivariate normal and have zero means.

Table 10--Zero Order Correlations for Field Survey Variables

	CS	RC	CS^2	RC^2	RCxCS	SAT
CS	1.0000	.2078*	-.2942**	-.0855	-.0623	.4004**
RC	.2078*	1.0000	-.0529	-.4009**	-.1224	.3379**
CS^2	-.2942**	-.0529	1.0000	.1238	.1883*	-.0998
RC^2	-.0855	-.4009**	.1238	1.0000	.3481**	-.2052*
RCxCS	-.0623	-.1224	.1883*	.3481**	1.0000	.0761
SAT	.4004**	.3379**	-.0998	-.2052*	.0761	1.0000

* $p < .01$
** $p < .001$

Table 11--Detection Techniques Results for Field Survey Data[a]

Dependent Variable	Predictor Variable	Coefficient b	p-value	R^2	p-value of difference
Saturated Product-Term Regression[b]					
SAT	CS	.267	.000		
	RC	.112	.001	.228	
	CS^2	.000	.995		
	RC^2	-.058	.046		
	RC x CS	.143	.009[c]	.261	.004[c]
Product-Term Regression[b]					
SAT	CS	.266	.000		
	RC	.138	.000	.228	
	RC x CS	.107	.038[c]	.245	.039[c]
Subgroup Analysis[b]					
Higher CS:					
SAT	CS	.389	.005		
	RC	.175	.005		
Lower CS:					
SAT	CS	.344	.001		
	RC	.091	.046		.000[c]
Dummy Variable Regression[d]					
SAT	CS	.332	.001		
	RC	.130	.000		
	D x CS	.076	.659[e]		
	D	1.851	.521		

ANOVA[b]		df	SS	MS	F	p-value
SAT	CS	1	2.02	2.02	12.79	.000
	RC	1	1.92	1.92	12.13	.001
	RC x CS	1	.14	.14	.91	.339[e]
	Residual	191	30.25	.15	9.62	.000

Chow Test[b]
SSE Full Data Set= 26.878
SSE Lower CS= 12.181
SSE Higher CS= 14.143
Q~ $F_{3,189}$= 1.325 (p-value= .267)[e]

[a] Cases= 204, centered CS and RC, and unreduced reliabilities: SAT= 79, CS= .81, RC= .83.
[b] SAT= $b_0+b_1RC+b_2CS+b_3RCxCS$.
[c] Suggests the existence of an interaction in the population equation.
[d] SAT= b_0+D+DxRC+RC+CS (D= 0 if CS   the median of the values for CS, D= 1 otherwise).
[e] Suggests the absence of an interaction in the population equation.

Table 12--Various Conditions Results

Dependent Variable	Predictor Variable	Coefficient b	p-value	R^2	p-value of difference
Uncentered, Unreduced Reliability, Unweighted, Cases= 204					
SAT	CS	-.273	.563		
	RC	.054	.815		
	CS^2	.000	.995		
	RC^2	-.058	.046	.234	
	RC x CS	.143	.009[b]	.261	.035[b]
SAT	CS	-.137	.487		
	RC	-.235	.189	.228	
	RC x CS	.107	.038[b]	.245	.035[b]
Centered, Reduced Reliability[a], Unweighted, Cases= 204					
SAT	CS	.198	.000		
	RC	.101	.006		
	CS^2	-.004	.936		
	RC^2	-.029	.318	.148	
	RC x CS	.071	.141[c]	.158	.124[c]
Centered, Reduced Reliability[a], Polar Weighted, Cases= 204					
SAT	CS	.351	.000		
	RC	.198	.000		
	CS^2	.060	.713		
	RC^2	.099	.065	.506	
	RC x CS	.116	.000[b]	.643	.000[b]
Centered, Unreduced Reliability, Unweighted, Cases= 50					
SAT	CS	.282	.016		
	RC	.162	.011		
	CS^2	-.087	.586		
	RC^2	-.052	.437	.275	
	RC x CS	.114	.357[c]	.289	.340[c]
Centered, Unreduced Reliability, Polar Weighted, Cases= 50					
SAT	CS	.388	.000		
	RC	.204	.000		
	CS^2	-.100	.389		
	RC^2	.029	.698	.527	
	RC x CS	.225	.003[b]	.615	.005[b]

[a] Reliabilities: SAT= .60, CS= .60, RC= .60.
[b] Suggests the existence of an interaction in the population equation.
[c] Suggests the absence of an interaction in the population equation.

Footnotes

1. While Maximum Likelihood (ML) and Generalized Least Squares (GLS) estimators are not formally appropriate for nonnormal variables (because ML and most implementations of GLS assume multivariate normality), their coefficient estimates appear to be robust against departures from normality (Anderson & Amemiya, 1985, 1986; Bollen, 1989; Boomsma, 1983; Browne, 1987; Harlow, 1985; Sharma, Durvasula & Dillon, 1989; Tanaka, 1984). However, their model fit and significance statistics are believed to be biased by nonnormal variables. Since the product indicators in nonlinear structural equation analysis are nonnormal (products of normally or nonnormally distributed variables are nonnormal), ML and GLS estimators are believed to be inappropriate for significance and model fit statistics in nonlinear structural equation analysis (Bollen, 1989; Kenny & Judd, 1984). The evidence to date (Jaccard & Wan, 1995; Kenny & Judd, 1984; Ping, 1995b,c) suggests that in nonlinear structural equation analysis these statistics from ML and possibly GLS estimators are robust to the addition of a few product indicators, involving linear indicators that are normal. However, the robustness of fit and significance statistics from these estimators to the addition of many product indicators (i.e. over four) or product indicators comprised of nonnormal linear indicators (typical of survey data) is unknown. While estimators that are less dependent on distributional assumptions hold some promise in this application, their empirical properties for smaller samples are generally unknown (see Hu, Bentler & Kano, 1992 for a summary).

2. There are also correlational approaches to detecting an interaction. These approaches include a subgrouping approach, and a case weighted regression approach (see Hedges & Olkin, 1985). Because these techniques are rarely seen, and the interested reader is directed to Jaccard, Turrisi and Wan (1990).

3. Ping (1995b) has proposed the use of a single indicator for XZ, $x{:}z = (x_1 + x_2)(z_1 + z_2)$.

4. Software such as LINCS (distributed by APTEC Systems), RAMONA (distributed by Professor Michael W. Browne at The Ohio State University), and MECOSA (distributed by SLI-AG, Frauenfeld Switzerland) could also be utilized.

5. LISREL 8 is available for microcomputers only, and according to individuals at SSI and SPSS when this paper was written (May, 1995), there were no plans to release a mainframe version of LISREL 8 in the near future. As a result mainframe LISREL 7 is likely to remain in use.

6. The sample size requirement of subgroup analysis using structural equation analysis limits its utility. Samples of 200 cases per group are usually recommended (Boomsma, 1983) (see Gerbing & Anderson, 1985 for an alternative view).

7. The Chow test can also be used with equation (5) and dummy variable regression. This test was not reported because the results of subgroup analysis and the Chow test are representative of the use of the Chow test with dummy variable regression.

8. Lubinski and Humphreys (1990) pointed out that as the correlation between X and Z approaches 1, the correlation between XX (or ZZ) and XZ also approaches 1. We therefore expected the .2 correlation between X and Z to increase the incidence of misdetecting a population quadratic as an interaction to a level above that of chance.

9. While it is a standard assumption in Structural Equation Modeling (see for example Bollen 1989), mean or zero centering has been the source of much confusion in regression analysis. The interested reader is directed to Aiken and West (1991) for an exhaustive and compelling demonstration of the efficacy of mean centering.

10. Mean centering is optional for the dependent variable, and SAT was uncentered for the Table 11 results (see Aiken and West 1991).

11. This strategy was generally representative of the results of the many variations on this theme. For example, the cases in the nonpolar cells could also have been weighted at some minimum level such as .1. This would have produced a four-cornered data model. Alternatively, the diagonal cells could have been weighted using a ratio of the largest polar cell size and their cell size, and the off-diagonal cells could have been minimally weighted. This would have produced an X data model.

12. As the Churchill and Peter (1984) results suggest, scale reliability can be improved by increasing the number of points in the scale. Churchill (1979) and Gerbing and Anderson (1988) proposed systematic approaches to improving scale reliability. In addition Anderson and Gerbing (1982) and Jöreskog (1993) proposed procedures to improve the unidimensionality of scales, which is related to their reliability. Finally see Netemeyer, Johnson and Burton (1990) for an example of their efforts to reduce measurement error in the unmeasured variables Role Conflict and Role Ambiguity using Average Variance Extracted (Fornell & Larker, 1981).

13. In addition we have experimented with this approach, and it appears to increase the likelihood of mistaking a quadratic in the population equation for an interaction.

14. Rusbult, Farrell, Rogers and Mainous (1988) reported multiple studies of employee reactions to dissatisfaction that included a Scenario Analysis. In this experiment students were instructed to read written scenarios in which they were to imagine they were the subjects of the experiment. The results of this research design and the other reported designs with considerably more internal and external validity were generally similar. Although no interactions were reported, the similarity of its results with those from other designs suggests that an experiment using Scenario Analysis might be appropriate as a easily executed second study.

INTERPRETING LATENT VARIABLE INTERACTIONS

(An earlier, but revised, version of Ping 2002, *American Marketing Association Winter Educator's Conference Proceedings,* Chicago: American Marketing Association)
(Updated June, 2002)

ABSTRACT

Using several examples, the paper proposes an easily implemented procedure for interpreting interactions and quadratics in unobserved or latent variables. The suggested procedure also sheds considerable light on these variables, and, overall, the paper is intended to help bridge what may be a gap in survey researchers' understanding of the nature of associations in survey data.

Interactions, e.g. XZ in

1) $$Y = b_0 + b_1X + b_2Z + b_3XZ + b_4XX + b_5ZZ + \zeta_Y ,$$

and quadratics (e.g., XX and ZZ in Equation 1) are likely to occur in experiments (McClelland and Judd 1993). Perhaps as a result, commercially available statistical packages (e.g., SAS, SPSS, etc.) estimate all possible ANOVA interactions by default, and experimental researchers routinely estimate all possible quadratics in ANOVA, to help them interpret significant main effects (e.g., b_1 and b_2 in Equation 1).

Interactions and quadratics are less likely to occur in survey data (see McClelland and Judd 1993, Podsakoff, et al 1984). Nevertheless, they may be more likely than their lack of investigation in survey research suggests (Busemeyer and Jones 1983; Birnbaum 1973, 1974) (however, in Marketing see Baumgartner and Bagozzi 1995; Lusch and Brown 1996; Osterhuis 1997; Ping 1994, 1999; Singh 1998).[1]

Authors have warned of the risks involved in ignoring interactions and quadratics in surveys (Blalock 1965; Cohen 1968; Cohen and Cohen 1975, 1983; Friedrich 1982; Kenny 1985; Howard 1989; Jaccard, Turrisi and Wan 1990; Aiken and West 1991): failure to consider interactions and quadratics is likely to lead to erroneous interpretations of the study's results.[2]

Yet estimating interactions and quadratics in survey data is rare for several reasons. Estimating these variables with regression is comparatively straightforward (see Aiken and West 1991). However, when variables are measured with error, regression coefficients (e.g., b_1, b_2, b_3, and b_4 in Equation 1) are biased (i.e., as more studies are done, the average of b_1, for example, across these studies does not converge to the population value for b_1), and inefficient (i.e., b_1 varies widely across studies) (Bohrnstedt and Carter 1971; Busemeyer and Jones 1983; Cochran 1968; Fuller 1987; Gleser, Carroll and Gallo 1987).

When variables are measured with error, unbiased and comparatively consistent structural equation estimation techniques for interactions and quadratics in are available (e.g., Bollen 1995; Hayduk 1987; Jaccard and Wan 1995; Jöreskog and Sörbom 1996; Kenny and Judd 1984; Ping, 1995, 1996b; Wong and Long 1987). However, these approaches are difficult to use (Aiken and West 1991-- see for example Jöreskog and Yang 1996). Further, despite numerous explanations of interactions and their interpretation (e.g., Aiken and West 1991; Darlington 1990; Denters and Van Puijenbroek 1989; Friedrich 1982; Hayduk 1987; Hayduk and Wonnacott 1980; Jaccard, Turissi and Wan 1990; Stolzenberg 1980), they may not be well understood by survey researchers (see for example Bedeian and Mossholder 1994; Denters and Van Puijenbroek 1989; Friedrich 1982; Lubinski and Humphreys 1990, Maxwell and Delaney 1993).

This paper aims to shed additional light on interactions and quadratics in survey data. It describes interactions and quadratics in this venue, and provides a straightforward approach to interpreting them that further illuminates their nature. We begin with a brief description of interactions and quadratics in survey data.

INTERACTIONS AND QUADRATICS IN SURVEY DATA

When there is no interaction between the variables X and Z in their association with the variable Y, their joint relationship with Y can be visualized as forming a plane in 3 dimensional space, the edges of which are the regression line of Y on Z (line 1 in Figure 1A), and the regression line of Y on X (line 2 in Figure 1A). The slope of regression line 1, for example, gives the strength, direction and significance of the association between Z and Y. For any values of X (e.g., x_a and x_b in Figure 1A) the regression lines of Y on Z at x_a and x_b (e.g., lines 1' and 1") are identical to line 1: they do not change orientation (i.e., they have the same slope). For all values of X the slope of the corresponding Z-Y regression line, or the association between Z and Y, at each of these x's is independent of the choice of x. Similarly, for all values of Z the slope of the corresponding X-Y regression line is independent of Z, and thus the association of X and Y is independent of Z.

When there is an interaction between X and Z, their joint relationship with Y can be visualized as a warped plane (see Figure 1B). In this case each value of x (e.g., x_c and x_d in Figure 1B) produces a different regression line of Y on Z (e.g., lines 3 and 4): each of these lines has a different orientation or slope. The equation for the slope of each of these lines, or the association between Z and Y, is the factored coefficient of Z in Equation 1 (e.g., the slope of line 3, or the association of Z and Y at x_c, is the number $b_2 + b_3x_c$; the association of Z and Y at x_d illustrated by line 4 is the number $b_2 + b_3x_d$; etc.). In this case, for each value of X the slope of the regression line, or the association between Z and Y, is different (i.e., the slope of regression line 3 is different from regression line 4, etc.), and thus the association between Z and Y (the slope) depends on the level of X.

Notice that in Equation 1 the factored coefficient of X is $b_1 + b_3Z$, and if the regression lines of Y on X were drawn in Figure 1 for different values of Z, each of them would also have a different orientation. Thus an XZ interaction produces a plane that is warped in both the Z-Y direction and the X-Y direction, and an XZ interaction affects the relationship or association between Z and Y, and the relationship between X and Y. In different words, a significant XZ changes the Z-Y and the X-Y associations from constants to variables (e.g., $b_2 + b_3X$ and $b_1 + b_3Z$ in Equation 1).

These comments apply regardless of how large the coefficient of the interaction XZ (b_3) is, and they have a perhaps surprising implication. There is always some amount of interaction between X and Z in Equation 1 with survey data, except in the unlikely event that the survey data is multivariate normal. Thus for proper interpretation of a study's results, the issue is how large is this interaction? If it is small (e.g., b_3 is not significant), it can safely be ignored because the Y relationships with X and Z, though they are nearly always variable in the sample, are not sufficiently variable to warrant special attention. However, if b_3 is not small, the almost always variable Y relationships with X and Z are sufficiently variable to warrant special attention.

Quadratics are visualized using tangent lines to a sufrace. When there is no quadratic in X (i.e., b_4 in Equation 1 is non significant), the joint relationship of X and Z with Y can be visualized as forming a plane as it did when there was no interaction. However, when there is a significant quadratic in X, the joint association plane is deformed into a trough (see Figure 1C). The shape of the trough is independent of Z (i.e., it is the same for all values of Z), but each different value of X produces a different tangent line to the trough (e.g., the tangent lines 3 and 4 for x_e and x_f, respectively, have different slopes), the slope of which is the association between X and Y. Stated differently, the association of X with Y at x_c is different from the association of X with Y at x_d because the tangent lines for each of these points have different slopes. The equation for the slope of each of these tangent lines (i.e., line 3 and 4) is the factored coefficient of X from Equation 1 (i.e., $b_1 + b_4x_e$

or b_1 + b_4x_f). Thus the size and direction of the slope, or the association between X and Y, in a small neighborhood of x (e.g., x_c) depends on the level of X (e.g, x_c): when X is low (e.g., x_c) the X-Y association (the slope if line 3) is negative, but when X is higher the X-Y association (the slope of line 4) is nearly zero.

Warps and troughs are not the only possible shapes of the Y-X-Z response surface. Other possibilities include shapes corresponding to third order terms (e.g., interactions between Z and XX or ZZ, cubics in Z such as ZZZ, etc.), fourth order terms (e.g., interactions between ZZ and XZ), and more. However, the strength of these higher order terms (i.e., their structural coefficients or their b's) depends directly on their reliability, which is a function of the product of powers of the reliabilities of the first order variables that make up these higher order terms (e.g., X and Z for ZXZ). Because reduced reliability in higher order variables reduces their structural coefficient strength in survey data (see Aiken and West 1991), higher order terms above second order are likely to have small structural coefficients (i.e., b's), and thus they are likely to have minimal effects on the primary associations in a study (e.g., b_1 and b_4).

For emphasis, Equation 1 is the correct specification of a model with the three variables X, Z and Y. Similarly, a model with four variables (e.g., X, Z, W, and a dependent variable Y) is correctly specified by including all possible interactions and quadratics among X, Z and W (i.e., XZ, XW, ZW, XX, ZZ, and WW). The fact that this is done automatically in ANOVA for interactions, or easily accomplished by the analyst using contrasts for quadratics, but it requires many additional software instructions in regression or structural equation analysis, may help explain why interactions and quadratics are seldom investigated in survey data. However, we suspect that even if the effort required to specify all possible interactions and quadratics in ANOVA were the same as that required in structural equation analysis, few experimental researchers would simply assume that interactions and quadratics are unimportant, and not bother to investigate them, as survey researchers appear to do.

INTERPRETING INTERACTIONS AND QUADRATICS

There have been several proposals for interpreting regression interactions in survey data (e.g., Aiken and West 1991; Darlington 1990; Denters and Van Puijenbroek 1989; Friedrich 1982; Hayduk 1987; Hayduk and Wonnacott 1980; Jaccard, Turissi and Wan 1990; Stolzenberg 1980). Some of these interaction approaches involve evaluating the coefficient of the interaction itself (e.g., b_3 in Equation 1), while others involve data plots similar to those employed in ANOVA. Others compare coefficients from median splits of the data, or they evaluate the factored coefficients (e.g., $b_2 + b_3X$ from Equation 1). However, there is little guidance for interpreting quadratics in survey data, and there is no guidance for interpreting latent variable interactions and quadratics. To help fill this gap, we will adapt an approach patterned after Friedrich's (1982) suggestions for interpreting interactions in regression (see also Darlington 1990, Jaccard, Turrisi and Wan 1990) because the result may be useful and it sheds additional light on the nature of interactions and quadratics.

To illustrate the proposed approach we will use examples that are based on disguised, but nevertheless real, data from actual surveys. The abbreviated results of the LISREL 8 (ML) estimation of a model similar to Equation 1 is shown in Figure 2A. In these results the XZ interaction is large enough to warrant interpretation (i.e., it is significant). The proposed interpretation approach for interactions relies on tables such as Table 1 and factored coefficients such as the factored coefficient of Z, $b_Z + b_{XZ}X$ in Figure 2A. Column 2 in Table 1, for example, shows the factored coefficient of Z from Figure 2A (.047 - .297X) at several Column 1 levels of X in the study that produced the Figure 2A results. Column 3 shows the standard errors of these factored coefficients of Z at the various levels of X, and Column 4 shows the resulting t-values. Footnotes a) through d) further explain the Columns 1-4 entries, and Footnote a) provides a verbal summary of the moderated Z-Y association produced by the significant XZ interaction in Figure 2A.

Because there are always two factored coefficients produced by an interaction, and the XZ interaction was large enough to warrant the Table 1 attention, Columns 5-8 are shown to interpret the factored coefficient of X, -.849 - .297Z. Column 6 shows this factored coefficient at several Column 5 levels of Z. Column 7 shows the standard errors of this factored coefficient at the various levels of Z, and Column 4 shows the resulting t-values. Again additional information regarding Columns 5-8 is provided in Footnotes e) through i), and Footnote e) provides a verbal summary of the moderated X-Y association produced by the significant XZ interaction in Figure 2A.

To provide an example of the interpretation of a quadratic, the abbreviated results of the LISREL 8 (ML) estimation of another model similar to Equation 1 is shown in Figure 2B. The interpretation of the significant VV quadratic shown there uses Table 2 and the factored coefficient of V, .348 - .159V. Column 2 of Table 2 shows the factored coefficient of V at several Column 1 levels of V in the study that produced the Figure 2B results. Column 3 shows the standard errors of the factored coefficient of V at the various levels of V, and Column 4 shows the t-values. Again, the Footnotes further explain the column entries, and Footnote a) in Table 2 provides a verbal description of the effect of V on the V-Y association.

The suggested interpretation procedure is:

a) Request that the structural equation output include the variances and the covariances of the structural coefficients (e.g., Var(b_Z)'s and Cov(b_Z,b_{XZ})) required to compute factored coefficient standard errors (see Footnotes d and h).

b) Create table(s) similar to Table 1 for a significant interaction, or Table 2 for a significant quadratic, using a spreadsheet. For emphasis, use mean centered X (Z) values in Column 2 (Column 6). Use the standard errors of the factored coefficients for an interaction that are shown in Footnotes d) and h) of Table 1, and the standard error of the factored coefficient for a quadratic that is shown in Footnote d) of Table 2. For the minimum, maximum, and average values shown in Column 1 (and 5) use those of the observed indicator whose loading was fixed at 1 (e.g., in Column 1 use the minimum, maximum, and average values of the indicator x with a structural model loading of 1, since this is the indicator that establishes the metric for X).[3]

c) Use Footnotes a) and e) as a guide to develop a verbal description(s) of the moderated association(s). These verbal description(s) should then be used in the discussion of the study results and their implications.

DISCUSSION

Several comments about Tables 1 and 2 may be of interest. As Columns 2 and 4 of Table 2 suggest, quadratics can behave like interactions. A factored coefficient for a quadratic can be negative at the high end of the range of the moderating variable (where it may or may not be significant), and it could be positive at the other end (where it may or may not be significant).

There are other possibilities as well. With different combinations of coefficients, interactions and quadratics could start out positive then become negative, or they could be consistently positive or consistently negative and simply become larger or smaller across the range of values of the moderating variable.

Notice that in Figure 1A b_Z was nonsignificant (t = 0.59), yet the factored coefficient of Z was significant at both ends of the range of X in the study (see Column 4 of Table 1). Z had a negative association with Y when X was high or above its study average for the sample, but its association with Y was positive when X was lower or below its study average. Similarly, b_X was significant in Figure 1A (t= -5.32), yet the coefficient of X moderated by Z was non significant when Z was very low (see Column 8 of Table 1). X was positively associated with Y for almost all levels of Z, except when Z was very low, where it was not associated with Y.

If standardized factored coefficients are desired (e.g., to compare the strength of the moderated coefficients to those of other variables, for example), they can be computed by multiplying the Column 2 factored coefficients in Table 1, for example, by the ratio of the variances of Z and Y (i.e., Column 2 value*Var(Z)/Var(Y), where Var is the dissattenuated variance available in the structural equation output).

As Aiken and West (1991) point out, the Column 2 (and 6) factored coefficient value at the average of the moderating variable is approximately the coefficient (and significance) of Z, for example, if XZ were omitted from Equation 1. Stated differently, if XZ were not specified in the Figure 2A equation, the observed Z-Y association would $b_Z' \approx .04$. Similarly, if XZ were not specified in the Figure 2A equation the X-Y association would be $b_X' \approx -.84$, and if VV were not specified in the Figure 2B equation b_V' would be approximately .34. Notice that because mean centered data was used, the Column 2 value (and significance) of the factored coefficient for Z, for example, at the mean value of the moderating variable X was equal to the b_Z value in Figure 2A (i.e., $b_Z + b_{XZ}X$ at X = 0, the mean of X, is $b_Z + b_{XZ}0 = .047 = b_Z$).[4]

This has several implications. Omitting a significant interaction or quadratic will not change the size, direction or significance of first order variables (e.g., X and Z in Equation 1), as long as mean centered variables are used. Omitting them simply reduces what we know about significant and nonsignificant first order variables and therefore clouds what we can say about them. For example in Figure 1A, b_X would still have been significant if XZ were omitted; thus any hypothesis involving X and Y would have been supported, and many would be tempted, for example, to recommend increasing X to decrease Y. However, this would be a poor recommendation because X actually had no effect on Y when Z was low (see Column 8 of Table 1).

Similarly if XZ were omitted from Figure 2A, any hypothesis involving Z-Y would have been disconfirmed, but we would not have known that the Z-Y hypothesis could have just as easily been confirmed if the study average of X were lower (or higher) (see Column 2 of Table 1) . Stated differently, a significant XZ interaction, for example, can mask or suppress the Z-Y association. Based on our experience, when interactions and quadratics are omitted in the analysis of survey data, an hypothesized association, such as b_Z for example, that turns up nonsignificant is frequently a signal that there is a significant XZ interaction or a significant XX or ZZ quadratic that is masking the hypothesized association (i.e., making it nonsignificant).

The invariance of b_x or b_z, for example, in Figure 2A to the presence or absence of significant interactions and quadratics also suggests that a significant interaction or quadratic adds to the explained variance of Y in Equation 1. In different words, omitting significant interactions and quadratics understates the impact of the first order variables X and Z in Equation 1, for example, on Y.

In addition, omitting a significant interaction or quadratic distorts any comparison of main effects (e.g., the comparison of β_X and β_Z, the standardized versions of b_x or b_z). For example, if the variances of X and Z were approximately equal (and thus Var(X)/Var(Y) □ Var(Z)/Var(Y)), X does not actually explain more variance in Y than the nonsignificant Z in Figure 2A. When X was low, X and Z explained about the same amount of variance in Y (see the bottom of Column 2 in Table 1).

For emphasis, mean or zero centered variables (i.e., deviation scores, where the mean of a variable is subtracted from each case value for that variable) were used in Tables 1 and 2. Mean centering is required for all variables, including Y, in order to use structural equation analysis with the available interaction and quadratic detection techniques (see Kenny and Judd 1984). However, there has been much confusion over the use of deviation scores and interactions in the regression literature. Nevertheless, as Aiken and West (1991) showed, when factored coefficients are used, mean centering has no effect on interpretation (i.e., the factored coefficients are scale invariant: the factored coefficients, their standard errors, and thus their t-values are identical with or without mean centering).

Explaining a quadratic usually requires some thought and care in wording. At first, it seems confusing to speak of the variable V, for example, as moderating itself. But when VV is significant, this is exactly what is implied: the variable V's association with Y depends on the existing level of V. For study respondents with a lower level of V in Table 2, increases in the variable V are associated with increased Y (see Column 2 of Table

2). For study respondents with a high level of V, changes in V are not associated with changes in Y.

Explaining both sides of an interaction can also be challenging. For example, a weakening and eventually non significant X-Y association (e.g., Columns 6 and 8 of Table 1) could be theoretically (or practically) plausible, but a positive then negative Z-Y association (e.g., Columns 2 and 4 of Table 1) could challenge existing theory. Associations that are significantly positive and significantly negative in one study can be difficult to explain, and it is natural to want to dismiss the result as faulty analysis. However, an implausible Z-Y association in Column 2 of Table 1, for example, can always be verified by restricting the model to cases where X (actually the indicator x whose loading was set to 1) is above 4.05 and below 3, and one may be left with a result that creates more questions than it answers.

The Columns 3 and 7 standard errors must be manually computed. While the output of variances and covariances of the structural coefficients (i.e., b's) required in Footnotes d) and h) can be requested, LISREL produces correlations instead. Nevertheless, the variance of a structural coefficient b is the square of its standard error, and standard errors are always available. In addition, the covariance of b_X and b_Z, for example, is $r_{X,Z}*SE_{bx}*SE_{bz}$), where $r_{X,Z}$ is the correlation of b_X and b_Z produced by LISREL.

It is also possible to calculate the point(s) at which the factored coefficients become significant (see Aiken and West 1991). For t = 2, the Column 1 (or 5) values (x_c) at which the factored coefficient t-value equals 2 are

$$\frac{-[2ab - 8Cov(a,b)] \forall \%\overline{[2ab - 8Cov(a,b)]^2 - 4[a - 4Var(a)][b^2 - 4Var(b)]}}{2[b^2 - 4Var(b)]},$$

where a and b are the coefficients in the factored coefficient a + bX (e.g., $b_2 + b_3X$, $b_1 + b_4X$, etc.); Var denotes the variance of a structural coefficient (i.e., the square of its standard error); and Cov denotes covariance. For emphasis a or b could be negative.

While it is possible to have significant pairs of second order variables that share a common first order variable (e.g., XZ and ZZ, XZ and XX, etc.), such occurrences appear to be rare with survey data in the social sciences (second order variables with no common first order variable, such as XZ and VV, are more likely). However, for completeness, the factored coefficient of Z, for example, for significant XZ and ZZ is ($b_2 + b_3X + b_5Z$), and the standard error of this factored coefficient is

% $\text{Var}(b_Z) + X^2\text{Var}(b_{XZ}) + 2X\text{Cov}(b_Z,b_{XZ}) + 2XZ\text{Cov}(b_3,b_5) + Z^2\text{Var}(b_5) + 2Z\text{Cov}(b_2,b_5)$.

The Table 1 is replaced by several interpretation tables, one for each level of X. The factored coefficient for X is similar, and it too has multiple interpretation tables.

REFERENCES

Aiken, Leona S. and Stephen G. West (1991), *Multiple Regression: Testing and Interpreting Interactions*, Newbury Park, CA: SAGE Publications.

Baumgartner, Hans and Richard P. Bagozzi (1995), ASpecification, Estimation, and Testing of Moment Structure Models Based on Latent Variates Involving Interactions Among the Exogenous Constructs,≅ *Sociological Methods and Research*, 24 (2) November, 187-213.

Bedeian, Arthur G. and Kevin W. Mossholder (1994), "Simple Question, Not So Simple Answer: Interpreting Interaction Terms in Moderated Multiple Regression," *Journal of Management*, 20 (1), 159-165.

Blalock, H. M., Jr. (1965), "Theory Building and the Concept of Interaction," *American Sociological Review*, 30, 374-81.

Bohrnstedt, G.W. & Carter, T.M. (1971). Robustness in regression analysis. In H.L. Costner (Ed.), *Sociological Methodology* (pp. 118-146). San Francisco: Jossey-Bass.

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.

Bollen, Kenneth A. and J. Scott Long (1993), *Testing Structural Equation Models*, Newbury Park, CA: SAGE.

Birnbaum, M.H. (1973), "The Devil Rides Again: Correlations as an Index of Fit," *Psychological Bulletin*, 79, 239-242.

_____ (1974), "Reply to the Devil's Advocates: Don't Confound Model Testing with Measurement," *Psychological Bulletin*, 81, 854-859.

Busemeyer, Jerome R. and Lawrence E. Jones (1983), "Analysis of Multiplicative Combination Rules When the Causal Variables are Measured With Error," *Psychological Bulletin*, 93 (May), 549-62.

Cochran, W.G. (1968). Errors in measurement in statistics. *Technometrics*, 10, 637-666.

Cohen, Jacob (1968), "Multiple Regression as a General Data-Analytic System," *Psychological Bulletin*, 70, 426-43.

_______ and Patricia Cohen (1975), *Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences*, Hillsdale, NJ: Lawrence Erlbaum.

_______ (1983), *Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences*, Hillsdale, NJ: Lawrence Erlbaum.

Darlington, R. B. (1990), *Regression and Linear Models*, New York: McGraw-Hill.

Denters, Bas and Rob A. G. Van Puijenbroek (1989), AConditional Regression Analysis,≅ *Quality and Quantity*, 23 (February), 83-108.

Friedrich, R. J. (1982), AIn Defense of Multiplicative Terms in Multiple Regression Equations,≅ *American Journal of Political Science*, 26, 797-833.

Fuller, W.A. (1987). *Measurement Error Models*. New York: Wiley.

Gleser, L.J., Carroll, R.J. & Gallo, P.P. (1987). The Limiting distribution of least squares in an errors-in-variables linear regression model. *The Annals of Statistics*, 15, 220-233.

Hayduk, Leslie A. and Tom Wonnacut (1980), " 'Effect Equations' or Effect Coefficients:' A Note on the Visual and Verbal Presentation of Multiple Regression Interactions," *Canadian Journal of Sociology*, 5, 399-404.

Hayduk, Leslie A. (1987), *Structural Equation Modeling with LISREL: Essential and Advances*, Baltimore, MD: Johns Hopkins Press.

Howard, John A.(1989), *Consumer Behavior in Marketing Strategy*, Englewood Cliffs, NJ: Prentice Hall.

Jaccard, James, Robert Turrisi and Choi K. Wan (1990), *Interaction Effects in Multiple Regression*, Newbury Park, CA: SAGE Publications.

_______ and C. K. Wan (1995), AMeasurement Error in the Analysis of Interaction Effects Between Continuous Predictors Using Multiple Regression: Multiple Indicator and Structural Equation Approaches,≅ *Psychological Bulletin*, 117 (2), 348-357.

Jöreskog, Karl G. and Dag Sörbom (1996), *LISREL 8 User=s Reference Guide*, Chicago: Scientific Software International, Inc.

_______ and F. Yang (1996), ANonlinear Structural Equation Models: The Kenny and Judd Model with Interaction Effects,≅ *Advances in Structural Equation Modeling Techniques*, G.A. Marcoulides, R.E. Schumacker, eds., Hillsdale, NJ: LEA.

Kenny, David A. (1985), "Quantitative Methods for Social Psychology," in *Handbook of Social Psychology*, Vol. 1, G. Lindsey et al. eds., New Your: Random House, 487-508.

_______ and C. M. Judd (1984), "Estimating the Nonlinear and Interactive Effects of Latent Variables," *Psychological Bulletin*, 96, 201-10.

Lubinski, D. & Humphreys, L.G. (1990). Assessing spurious Amoderator effects≅: illustrated substantively with the hypothesized (Asynergistic≅) relation between spatial and mathematical ability. *Psychological Bulletin*, 107, 385-393.

Lusch, Robert F. and James R. Brown (1996), AInterdependency, Contracting, and Relational Behavior in Marketing Channels,≅ *Journal of Marketing*, 60 (4) October, 19-38.

Maxwell, S.E. & Delaney, H.D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, 113, 181-190.

McClelland, G. H. and C. M. Judd (1993), AStatistical Difficulties of Detecting Interactions and Moderator Effects,≅ *Psychological Bulletin*, 114 (2), 376-390.

Osterhuis, Thomas L (1997), APro-social Consumer Influences Strategies: When and How Do They Work?≅ *Journal of Marketing*, 61 (4) October, 16-29.

Ping, R.(1994), "Does Satisfaction Moderate the Association Between Alternative Attractiveness and Exit Intention in a Marketing Channel?≅ *Journal of Retailing*, 22 (4) Fall, 364-371.

_______ (1995), "A Parsimonious Estimating Technique for Interaction and Quadratic Latent Variables," *The Journal of Marketing Research*, 32 (August), 336-347.

_______ (1996b), "Latent Variable Interaction and Quadratic Effect Estimation: A Two-Step Technique Using Structural Equation Analysis, *The Psychological Bulletin*, 119 (January), 166-175.

_______ (1999), "Unexplored Antecedents of Exiting in a Marketing Channel,≅ *Journal of Retailing*, 75 (2) Summer, 218-242.

Podsakoff, Philip M., William D. Tudor, Richard A. Grover and Vandra L. Huber (1984), ASituational Moderators of Leader Reward and Punishment Behaviors: Fact or Fiction?≅ *Organizational Behavior and Human Performance*, 34, 21-63.

Singh, J. (1998), AStriking a Balance in Boundary-spanning Positions: an Investigation of Some Unconventional Influences of Role Stressors and Job Characteristics on Job Outcomes of Salespeople,≅ *Journal of Marketing*, 62 (3) July, 69-86.

Stolzenberg, Ross M. (1980), "The Measurement and Decomposition of Causal Effects in Nonlinear and Nonadditive Models," in *Sociological Methodology*, Karl F. Schueller ed., San Francisco: Jossey-Bass, 459-488

Wong, S. K. and J. S. Long (1987), "Reparameterizing Nonlinear Constraints in Models with Latent Variables," Technical Report, Washington State University.

Figure 1- Response Surfaces for Y as a function of X and Z

(Click here to view Figure 1)

Figure 2- Equation 1 Structural Model Estimation Results

A) $Y = b_X X + b_Z Z + b_{XZ} XZ + b_{XX} XX + b_{ZZ} ZZ$

-.849	.047	-.297	.001	.004	(= b)
(-5.32)	(0.59)	(-4.00)	(0.10)	(0.09)	(= t-value)

B) $Y = b_V V + b_W W + b_{VW} VW + b_{VV} VV + b_{WW} WW$

.348	-.347	-.007	-.159	.010	(= b)
(5.11)	(-2.34)	(-0.09)	(-3.45)	(0.29)	(= t-value)

Table 1- Unstandardized Y Associations with Z and X Implied by the Figure 2A Results

Z-Y Association Moderated by X[a]				X-Y Association Moderated by Z[e]				
X Level[b]	Z Coef-ficient[c]	SE of Z Coef-ficient[d]	t-value of Z Coef-ficient	Z Level[f]	X Coef-ficient[g]	SE of X Coef-ficient[h]	t-value of X Coef-ficient	
5	-0.23	0.09	-2.48	5	-1.31	0.25	-5.19	
4.05[i]	0.04	0.08	0.59	4	-1.01	0.19	-5.33	
4	0.06	0.08	0.77	3.44[i]	-.84	0.16	-5.32	
3	0.36	0.12	2.92	3	-.71	0.14	-5.15	
2	0.65	0.18	3.52	2	-.42	0.11	-3.59	
1.2	0.89	0.24	3.70	1	-.12	0.13	-.94	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(Column Number)

[a] The table displays the variable association of X and Z with Y produced by the significant XZ interaction. In Columns 1-4 when the existing level of X was low in Column 1, small changes in Z were positively associated with Y (see Column 2). At higher levels of X however, Z was less strongly associated with Y, until near the study average for X, the association was nonsignificant (see Column 4). When X was above its study average, Z was negatively associated with Y.

[b] The value of X ranged from 1.2 (= low X) to 5 in the study.

[c] The coefficient of Z is (.047-.297X)Z with X mean centered. E.g., when X = 1.2 the coefficient of Z is .047-.297*(1.2 - 4.05) = .89.

[d] The Standard Error of the Z coefficient is:

$$\%\ \overline{Var(b_Z+b_{XZ}X)} = \%\ \overline{Var(b_Z) + X^2Var(b_{XZ}) + 2XCov(b_Z,b_{XZ})},$$

where Var and Cov denote variance and covariance, and b denotes unstandardized structural coefficients from Figure 2A.

[e] This portion of the table displays the association of X and Y moderated by Z. When Z was low in Column 5, the X association with Y was not significant (see Column 8). However, as Z increased, X=s association with Y quickly strengthened, until it was negatively associated with Y for most values of Z in the study.

[f] The value of Z ranged from 1 (= low Z) to 5 in the study.

[g] The unstandardized coefficient of X is (-.849-.297Z)X with Z mean centered. E.g., when Z = 1 the coefficient of X is -.849-.297*(1-3.44) = -.12.

[h] The Standard Error of the X coefficient is:

$$\%\ \overline{Var(b_X+b_{XZ}Z)} = \%\ \overline{Var(b_X) + Z^2Var(b_{XZ}) + 2ZCov(b_X,b_{XZ})},$$

where Var and Cov denote variance and covariance, and b denotes unstandardized structural coefficient from Figure 2A.

[i] Mean value in the study.

Table 2- Unstandardized V-Y Associations Implied by the Figure 2B Results

V-Y Association Moderated by the Level of V[a]			
V Level[b]	V Coef-ficient [c]	SE of V Coef-ficient[d]	t-value of V Coef-ficient
5	-0.07	0.11	-0.64
4	0.08	0.07	1.11
3	0.24	0.06	3.96
2.36[e]	0.34	0.06	5.11
2	0.40	0.07	5.31
1	0.56	0.11	5.15
(1)	(2)	(3)	(4)

(Column Number)

[a] The table displays the variable association of V with Y produced by the significant quadratic VV. When the existing level of V was low in Column 1, small changes in V were positively associated with changes in Y (see Column 2). As the Column 1 level of V increased in the study, V=s association with Y weakened (i.e., became smaller in Column 2), and it was nonsignificant when the level of V was high (see Column 4).

[b] The value of V ranged from 1 (= low V) to 5 in the study.

[c] The factored coefficient of V is (.348-.159V)V with V mean centered. E.g., when V = 1 the coefficient of V is .348 -.159*(1-2.36) = .56.

[d] The Standard Error of the V coefficient is:

$$\% \overline{Var(b_V+b_V V)} = \% \overline{Var(b_V)+V^2*Var(b_{VV})+2*V*Cov(b_V,b_{VV})} ,$$

where Var and Cov denote variance and covariance, and b denotes unstandardized structural coefficients from Figure 2B.

[e] The mean of V in the study.

ENDNOTES

1.Interactions and quadratics can be viewed as an artifact of data that is not multivariate normal, and continuous or ratio survey data, although typically nonnormal, is less so than data from experiments. However, because many surveys involve ordinal data (e.g., from Likert scales) that is per se nonnormal, interactions and quadratics should be more likely than their occurrence in published survey results suggest.

2. In Equation 1 the factored coefficient of Z is (b_2 + b_3X). When the coefficient of XZ (b_3) is significant, this factored coefficient of Z will be significant for some X values in the study, and this produces very different interpretations of the Z-Y association than does the coefficient of Z in Equation 1 without XZ (i.e., Y = b_0' + b_1'X + b_2'Z) (see Aiken and West 1991). If b_2' is not significant, failing to include a significant XZ in Equation 1 produces a misleading interpretation of the Z-Y association. A nonsignificant b_2' implies the Z-Y association is disconfirmed, but it is misleading to state that Z is not associated with Y in the study. This association simply depends on the level of X in the study, and this statement may be more important to some readers than the disconfirmation of the Z-Y association.

Alternatively, b_2' could be significant while (b_2 + b_3X) could be nonsignificant for some (but not all) values of X. In this case, not including a significant XZ will produce a misleading confirmation of the Z-Y association. The significant Z-Y association is actually nonsignificant for some values of X in the study, and this information may also be more important to some readers than a confirmation of the Z-Y association

The algebra and implications of not including quadratics in Equation 1 are similar, except the equation for the coefficient of Z, for example, changes to (b_2 + b_5Z).

3. If there are any indicators of X with loadings greater than one, the loadings of X should be respecified by fixing the loading of the indicator with largest loading to 1 and freeing the other indicators of X.

4. This is not true with non mean centered data. With non mean centered data (which should not be used to detect interactions or quadratics with structural equation analysis), the Column 2 factored coefficient value at the average of the moderating variable X, for example, will not equal b_Z in Figure 2A.

A PARSIMONIOUS ESTIMATING TECHNIQUE FOR INTERACTION AND QUADRATIC LATENT VARIABLES

(An earlier, but revised, version of Ping 1995, *Journal of Marketing Research*, 32 (August), 336-347)
(Revised July, 2001)

ABSTRACT

An alternative estimation technique is proposed for interaction and quadratic latent variables in structural equation models using LISREL, EQS, and AMOS. The technique specifies these variables with a single indicator. The loading and error terms for this single indicator can be specified as constants in the structural model. The technique is shown to perform adequately using synthetic data sets.

INTRODUCTION

Opportunities for investigating interactions and quadratic variables are ubiquitous in Marketing theory (e.g., Walker, Churchill, and Ford 1977 and Weitz 1981 in the personal selling literature; Ajzen and Fishbein 1980, Engel, Blackwell and Kollat 1978, Howard 1977, and Howard and Sheth 1969 in the consumer behavior literature; Dwyer, Schurr and Oh 1987 and Stern and Reve 1980 in the channel literature; and Sherif and Hovland 1961 in the advertising literature). Interactions and quadratics are often encountered by survey researchers in Marketing (Howard 1989) (see for example Batra and Ray 1986, Heide and John 1992, Kohli 1989, Laroche and Howard 1980, and Teas 1981). In addition, researchers have called for the investigation of interaction and quadratic variables in survey data to improve the interpretation of study results (see the citations in Aiken and West 1991, and Jaccard, Turrisi and Wan 1990 involving the social sciences, and Howard's 1989 remarks in Marketing). They point out that failing to consider the existence of interactions and quadratic variables in survey data increases the risk of misleading research findings, as it does in ANOVA studies.

However, researchers encounter major obstacles when they attempt to estimate interactions or quadratics in models involving latent variables. The popular estimation methods for these nonlinear latent variables have theoretical or practical limitations. For example, the most popular estimation technique, regression, is known to produce coefficient estimates that are biased and inefficient for variables measured with error such as latent variables (Busemeyer and Jones 1983). Approaches that involve sample splitting to detect these variables are criticized for their reduction of statistical power, and the resultant likelihood of false disconfirmation (Cohen and Cohen 1983, Jaccard, Turrisi and Wan 1990).

Structural equation analysis approaches are difficult to use (Aiken and West 1991), in part because, until recently, popular structural equation software packages (e.g., LISREL, EQS, etc.) were unable to properly specify interaction and quadratic latent variables.

This article proposes an estimation technique for interaction and quadratic latent variables that avoids many of these obstacles. The technique involves structural equation analysis, and it specifies an interaction or quadratic latent variable with a single indicator. The loading and error term for this single indicator need not be estimated in the structural model: they can be specified as constants in that model. The efficacy of this technique is investigated by recovering known coefficients, detecting known significant effects, and gauging known model-to-data fits in synthetic data sets.

ESTIMATING INTERACTION AND QUADRATIC VARIABLES

While there are many proposed approaches to detecting interactions and quadratics (see Jaccard, Turrisi and Wan 1990 for a summary), there are three general categories of approaches to estimating these variables involving latent variables: regression analysis, subgroup analysis, and indicator-product structural equation analysis.

To use regression analysis with unobserved or latent variables, a dependent variable is regressed on independent variables composed of summed indicators and products of these summed indicators (i.e, for the interactions or quadratics). Subgroup analysis involves dividing the study cases into subgroups of cases using, for example, the median of a suspected interaction or quadratic variable, estimating the model using each subgroup of cases and regression or structural equation analysis, and then testing for significant coefficient differences between the subgroups. To use structural equation analysis, quadratic and interaction latent variables are specified in a structural equation model using products of indicators. Coefficients are estimated either directly using software such as LISREL 8 or CALIS (available in SAS), or indirectly using software such as EQS AMOS, or earlier versions of LISREL.

Regression analysis, is generally recommended for continuous variables (Cohen and Cohen 1983, Jaccard, Turrisi and Wan 1990, Aiken and West 1991). However, for variables measured with error such as latent variables, regression is known to produce biased and inefficient coefficient estimates (Busemeyer and Jones 1983). While proposals to remedy this situation have been made (see Feucht 1989 for a summary), they are seldom seen in the social sciences, perhaps because they lack significance testing statistics (Bollen 1989). Authors have also commented on the loss of statistical power in regression as reliability declines (see Aiken and West 1991 for a summary). Finally, regression limits the researcher to investigating one dependent variable at time.

The second approach, subgroup analysis, is a preferred technique in some situations. Jaccard, Turrisi and Wan (1990) state that subgroup analysis may be appropriate on theoretical grounds: the model could be posited to be structurally different for different subject subgroups. They also point out that an interaction need not be of the form "X times Z" (the possibilities for the form of an interaction are infinite, Jaccard, Turrisi and Wan 1990), and that three group analysis may be more appropriate in these cases. Sharma, Durand and Gur-Arie (1981) recommend subgroup analysis to detect what they term a homologizer: a variable W, for example that affects the strength of the independent-dependent variable association, yet is not related to either of these variables.

However, the subgroup analysis approach of splitting the sample is criticized for its reduction of statistical power, and the resultant likelihood of false disconfirmation (Cohen and Cohen 1983, Jaccard, Turrisi and Wan 1990). This approach also reveals neither the magnitude nor the actual form of any significant interaction or quadratic.

The third approach, uses products of indicators to specify interaction and quadratic variables in a structural equation model (Kenny and Judd 1984). For example, in the Figure 1 model that involves the latent variables X, Z and Y, the latent variables XZ and XX have been added by specifying their indicators to be all possible unique products of the X and Z indicators. This product-indicator structural equation model is then analyzed using the full set of cases, and significant XZ or XX coefficients suggest the presence of an interaction or quadratic, respectively, and their form and magnitude.

However, this approach has theoretical and practical limitations. For example it is inappropriate when the form of the interaction is other than the product of independent variables (e.g., the interaction could be of the form X/Z, Jaccard, Turrisi and Wan 1990). In addition, it is complicated by the nonlinear form of the loadings and error terms of the product indicator (e.g., λ_{x1z1} and ε_{x1z1} in Figure 1). These nonlinear loadings and error terms cannot be specified in a straightforward manner in most structural equation analysis software (e.g., EQS, AMOS, etc.) (Hayduk 1987). For structural equation modeling software that provides straightforward specification of nonlinear loadings and error terms (e.g., LISREL 8), however, the mechanisms provided to accomplish are tedious to use. In addition, straightforward specification produces many additional variables that can produce problems with convergence and unacceptable solutions in larger models. Further, adding these nonlinear indicators (e.g., more than about 6) can produce model-to-data fit problems. Finally, significance tests and model fit statistics produced by popular estimators such as maximum likelihood are believed to be inappropriate for models involving interactions and quadratics (Bollen 1989, see Hu, Bentler and Kano 1992).

Fortunately these matters are beginning to be addressed for product-indicator structural equation analysis (see Bollen 1989 for a summary). Because product-indicator analysis avoids the limitations of regression and subgroup analysis, the balance of the article will discuss product-indicator analysis.

PRODUCT-INDICATOR ANALYSIS TECHNIQUES

There are two published implementations of product-indicator analysis (when this article was published).[1] The first was proposed by Kenny and Judd (1984), and the next was suggested by Hayduk (1987).

THE KENNY AND JUDD APPROACH

Kenny and Judd (1984) proposed that products of indicators would specify nonlinear latent variables. For example, in the Figure 1 model that involves the latent variables X, Z and Y, the latent variables XX and XZ have been added by specifying their indicators to be all possible unique products of the X and Z indicators. In addition, they showed that under certain conditions, the variance of a product of indicators is determined by the variance of their linear constituents.

They showed for latent variables X and Z, the variance of the indicator x_1z_1 is given by

$$Var(x_1z_1) = Var[(\lambda_{x1}X + \varepsilon_{x1})(\lambda_{z1}Z + \varepsilon_{z1})]$$

1) $$= \lambda_{x1}^2\lambda_{z1}^2Var(XZ) + \lambda_{x1}^2Var(X)\theta_{\varepsilon z1} + \lambda_{z1}^2Var(Z)\theta_{\varepsilon x1} + \theta_{\varepsilon x1}\theta_{\varepsilon z1} .$$

when X and Z are independent of the error terms ε_{x1} and ε_{z1}, the error terms are themselves mutually independent, the indicators (x_1 and z_1) have zero expected values, and X and Z along with ε_{x1} and ε_{z1} are normally distributed.

Then they specified latent variables such as XZ with indicators such as x_1z_1 by constraining the loading and the error term for x_1z_1 (λ_{x1z1} and $\theta_{\varepsilon x1z1}$) to be the following nonlinear combinations of linear-terms-only model parameters

2) $$\lambda_{x1z1} = \lambda_{x1}\lambda_{z1} ,$$

and

3) $$\theta_{\varepsilon x1z1} = \lambda_{x1}^2Var(X)\theta_{\varepsilon z1} + \lambda_{z1}^2Var(Z)\theta_{\varepsilon x1} + \theta_{\varepsilon x1}\theta_{\varepsilon z1} .$$

They specified these directly using the structural equation package COSAN (now available in SAS as a subprocedure in the procedure CALIS), that accepts nonlinear constraints such as the terms on the right-hand sides of equations (2) and (3).

THE HAYDUK APPROACH

Hayduk demonstrated that the product indicators that Kenny and Judd proposed could be specified indirectly by adding additional "convenience" variables to the Figure 1 model. For example, inserting a convenient latent variable η_1 on the path between XZ and x_1z_1 in Figure 1 will specify the first term of equation (1) when the loading of this variable ($\lambda_{XZ,\eta1}$) is set equal to λ_{x1}, its loading on x_1z_1 ($\lambda_{\eta1,x1z1}$) is set equal to λ_{z1}, and its variance is fixed at 1 (using the rules of path analysis: the variance of x_1z_1 is now the product of $\lambda_{XZ,\eta1}^2$ (= λ_{x1}^2), $\lambda_{\eta1,x1z1}^2$ (= λ_{z1}^2), and the variance of XZ). By creating additional paths to x_1z_1 using more such η's with parameters fixed at the equation (1) values, the remaining three terms in equation (1) can be specified (see Hayduk 1987 Chapter 7).

For a latent variable with many indicators, or for models with several interactions or quadratics, however, these approaches can become impractical: the volume of indicator product variables created using all pairwise products, and the number of "equation 2's and 3's" to be coded can create difficulties for the researcher, the computer estimation process, and model-to-data fit. This suggests the need for an approach that does not require the use of numerous additional variables or numerous equation (2) and (3) specifications.

The balance of this article describes an estimation approach that involves a single indicator per latent variable.

A PARSIMONIOUS ESTIMATION TECHNIQUE

In the regression literature Cohen and Cohen (1983) suggested the use of the product of summed indicators to estimate an interaction or quadratic variable. They proposed that, for example, the observed variables $x = x_1 + x_2$ and $z = z_1 + z_2$, when multiplied together as $(x_1 + x_2)(z_1 + z_2)$, would specify an XZ interaction. Similarly this article proposes that a single indicator, for example $x{:}z = (x_1 + x_2)(z_1 + z_2)$, could be used to specify the latent variable interaction XZ. In particular, the Figure 1 model could be respecified as the Figure 2 model in which the single indicators x:x (= $[x_1 + x_2][x_1 + x_2]$) and x:z (= $[x_1 + x_2][z_1 + z_2]$) are used in place of the product indicators shown in the Figure 1 model.

The loadings and errors for the indicators x:z and x:x in Figure 2 are given by

4) $$\lambda_{x:z} = (\lambda_{x1} + \lambda_{x2})(\lambda_{z1} + \lambda_{z2})$$

5) $$\theta_{\varepsilon x:z} = (\lambda_{x1}+\lambda_{x2})^2Var(X)(\theta_{\varepsilon z1}+\theta_{\varepsilon z2}) + (\lambda_{z1}+\lambda_{z2})^2Var(Z)(\theta_{\varepsilon x1}+\theta_{\varepsilon x2}) + (\theta_{\varepsilon x1}+\theta_{\varepsilon x2})(\theta_{\varepsilon z1}+\theta_{\varepsilon z2}),$$

6) $$\lambda_{x:x} = (\lambda_{x1} + \lambda_{x2})^2,$$

and

7) $$\theta_{\varepsilon x:x} = 4(\lambda_{x1}+\lambda_{x2})^2Var(X)(\theta_{\varepsilon x1}+\theta_{\varepsilon x2}) + 2(\theta_{\varepsilon x1}+\theta_{\varepsilon x2})^2$$

(see Appendix A for details).

With these formulas for $\lambda_{x:z}$, $\theta_{\varepsilon x:z}$, $\lambda_{x:x}$, and $\theta_{\varepsilon x:x}$, CALIS, or LISREL 8 could be used to estimate the Figure 2 model directly. However, since estimates of the parameters on the right-hand side of equations (4) through (7) are available in the measurement model for Figure 2, we could further simplify matters by using measurement model parameter estimates.

Anderson and Gerbing (1988) recommended the use of a measurement model to separate measurement issues from model structure issues. Many researchers view a latent variable model as the synthesis of two models: the measurement model that specifies the relationships between the latent variables and the observed variables, and the structural model that specifies the relationships among latent variables (Anderson and Gerbing 1988, Bentler 1989, Bollen 1989, Jöreskog and Sörbom 1989). Anderson and Gerbing (1988) proposed specifying these two models separately, beginning with the measurement model, and using the measurement model to ensure the unidimensionality of the latent variables. They argued that this avoids interpretational confounding (Burt, 1976), the interaction of the measurement and structural models, and the possibility of marked differences in the estimates of the parameters associated with the observed variables (i.e., λ's, θ_ε's, and latent variable variances) between the measurement and structural models.

Anderson and Gerbing (1988) added that with "acceptable unidimensionality" the measurement model parameter estimates should change trivially, if at all, when the measurement submodel and alternative structural submodels are simultaneously estimated (p. 418). As a result, I propose that as an alternative to specifying the interaction and quadratic parameters (e.g., $\lambda_{x:z}$, $\theta_{\varepsilon x:z}$, $\lambda_{x:x}$, and $\theta_{\varepsilon x:x}$ in Figure 2) as variables, they can be specified as constants in the structural model when X and Z are each acceptably unidimensional. Specifically, parameter estimates from a linear-terms-only measurement model (e.g., involving X and Z only) can be used to compute the values of $\lambda_{x:z}$, $\theta_{\varepsilon x:z}$, $\lambda_{x:x}$, and $\theta_{\varepsilon x:x}$ in equations (4) through (7), and these computed values can be specified as fixed loadings and errors for x:z and x:x in the Figure 2 structural model. The unidimensionality of X and Z in Figure 2 enables the omission of the nonlinear latent variables from the linear-terms-only measurement model: because X and Z are each unidimensional, their indicators are unaffected by the presence or absence of other latent variables in a measurement or structural model, in particular XX or XZ. Stated differently, this provides trivially dissimilar measurement parameter estimates between measurement and structural models, and enables the use of the equation (4) through (7) estimates as fixed values in the structural model.

To gauge the efficacy of this proposed approach with its two options for estimating equations (4) through (7) either as variables or as constants, known coefficients were recovered in synthetic data sets.

SYNTHETIC DATA SETS

Synthetic data sets were generated using known population parameters, and the proposed approach was used to estimate the population structural coefficients. Using a normal random number generator and the procedure described in Appendix B, data sets composed of 100 replications of samples of 100, 200 or 300 cases were created.[2]

We will describe the baseline simulation first: a data set containing 100 replications of a sample involving 200 cases. Each replication was generated using the Table 1 population characteristics for x_1, x_2, z_1, z_2, t_1, t_2 and y in the Figure 3 model.

This model was estimated using the proposed technique on each replication by (i) estimating the measurement model parameters,[3] (ii) calculating the equations (4) through (7) values for the loadings and error variances of x:z and x:x (i.e., $\lambda_{x:z}$, $\theta_{\varepsilon x:z}$, $\lambda_{x:x}$, and $\theta_{\varepsilon x:x}$) using the measurement model parameter estimates,[4] and (iii) estimating the Figure 3 structural model with fixed equation (4) through (7) values for $\lambda_{x:z}$, $\theta_{\varepsilon x:z}$, $\lambda_{x:x}$, and $\theta_{\varepsilon x:x}$ as follows.

For each replication, the linear-terms-only measurement model associated with the Figure 3 model was estimated using maximum likelihood (ML) and LISREL 8. This produced estimates of the λ's, θ_ε's, and latent variable variances required in equations (4) through (7). Then the structural model for Figure 3 was specified by fixing the values for the single indicator loadings ($\lambda_{x:z}$ and $\lambda_{x:x}$) and error variances ($\theta_{\varepsilon x:z}$ and $\theta_{\varepsilon x:x}$) to the appropriate equation (4) through (7) calculated values. The results of the subsequent structural model estimations of the Figure 3 β's using LISREL 8 and ML [5] are shown in Table 2 and titled "2 Step."

We also generated several additional estimates. These included LISREL 8 ML estimates of Figure 3 produced by specifying the equation (4) through (7) single indicator loadings ($\lambda_{x:z}$ and $\lambda_{x:x}$) and error variances ($\theta_{\varepsilon x:z}$ and $\theta_{\varepsilon x:x}$) using LISREL 8's constraint equations (i.e., the proposed approach with free instead of fixed single indicator loadings and error variances) (these estimates are titled "LISREL 8" in Table 2). In addition, Kenny and Judd estimates were produced using a product indicator version of Figure 3 with XX and XZ specified as they are in Figure 1, LISREL 8 with ML estimation, and constraint equation specifications for the loadings and errors of the indicator products (i.e., x_1z_1, x_1z_2, x_2z_1, z_2z_2, x_1x_1, x_1x_2, and x_2x_2). Finally, regression estimates were produced using ordinary least squares.

To gauge the effects of varying the simulation conditions, eight more data sets were generated. These variations in the simulation conditions reflected four indicators per linear latent variable, two levels of sample size (100 and 300), two levels of linear latent variable reliability (ρ=.6 and .9), and two levels of nonlinear coefficient size. Following the two step procedure described above using Figure 3, four indicators per linear latent variable, and the population parameters shown in Table 3; and 100 replications, ML estimates, and EQS instead of LISREL 8, the results shown in Table 4 were obtained.

In order to assess significance and model fit in these eight data sets, Tables 5 and 6 summarize the observed incidence of nonsignificant coefficients (i.e., coefficients with t-values less than 2) and lack of fit (i.e., a Comparative Fit Index (Bentler 1990) less than .9) produced by ML estimates, and two convenient less distributionally dependent estimators, the

Robust estimator (Bentler and Dijkstra 1985), and the asymptotic distribution free (ADF) estimator (Browne 1982). These results, along with those shown in Tables 2 and 4, will be discussed next.

RESULTS

Based on the estimation results, the detection of significant effects, and model-to-data fit the proposed approach performed adequately. For example, the proposed approach with fixed or free single indicator loadings and error variances produced average coefficient values (E(β)'s in Table 2) that were within a few points of the population values. It also had a bias or distance from the population value of these averages equivalent to the Kenny and Judd estimates,[6] and less than the regression estimates, except for T.[7]

The mean squared differences between the estimated coefficient and the population value (MSE in Table 2) for the proposed approach were equivalent overall to those from the Kenny and Judd approach, but larger than they were for regression for the linear variables.[8] For the nonlinear variables, the situation was reversed with respect to regression, and overall, the techniques all produced approximately the same average variation around the population value.

The variance of the coefficient estimates for the proposed approach was equivalent to the Kenny and Judd approach - (see Table 2). In comparison, the variances for the proposed approach were larger than those for regression and the linear variables (see Footnote 6), but smaller than regression for the nonlinear variables. Overall, the variances for the proposed approach were smaller than those for regression.

The eight additional simulation results shown in Table 4 paralleled the baseline simulation results shown in Table 2: the coefficient estimate averages were within a few points of the population values, the biases were small and they appeared to be random, and the variations of estimates around the population and average values were consistent with the baseline simulation. The low-reliability-100-case samples (SβLρ100 and LβLρ100) produced the worst effectiveness measures, while the high-reliability-300-case samples (SβHρ300 and LβHρ300) produced the best. For the low-reliability-100-case samples, however, the effectiveness measures did not appear to be unacceptable. For example, the distance of the coefficient averages from the population value (Bias) ranged from 2.29 to 8.67%. In addition, the mean squared differences between the estimated coefficient and the population value variations (MSE), and the variance of the coefficient estimates (Var(β)), ranged between .05 and .07 .

Turning to the detection of significant effects, the proposed approach performed acceptably (see Table 5). However, the incidence of false negative significance tests was sensitive to reliability, sample size, and population coefficient size. In general, smaller population coefficients in the 100 case samples with low (.6) reliabilities produced false negative significance tests at a level well above that of chance (10%). As coefficient and sample size increased and reliability improved, however, the incidence of false negatives declined to chance levels. The relative effects of increasing reliability and larger samples were slightly different: selectively raising reliability decreased the incidence of false negatives more than selectively increasing sample size. In particular, at low reliability the larger sample size reduced the incidence of false negatives to near chance levels. But at the smaller sample size, increased reliability reduced the incidence of false negatives to zero. Overall, these results were consistent with previous studies (see Aiken and West 1991 for a summary), which suggest that increased sample size, reliability and population coefficient size increase the likelihood of detecting a nonlinear effect.

The incidence of false negatives was generally the same using ML, ML-Robust, and ADF estimators (see Table 5). This result was unexpected because it is believed that standard errors associated with ML estimates are not robust to departures from normality (see for example Bentler 1989, Bollen 1989, and Jöreskog and Sörbom 1989). Since the simulations involved the specification of two nonlinear indicators in a model involving more than a dozen indicators that were generated to be normally distributed, it is possible that the proposed approach's addition of relatively few nonlinear indicators retains some robustness with ML estimation.

The proposed approach also performed satisfactorily in assessing model fit. Table 6 shows the incidence of replications in which the population model did not fit the data using the Comparative Fit Index (Bentler 1990) resulting from ML, ML-Robust, and ADF estimators. While there is little agreement on indices of model fit (see Bollen and Long 1993), the Comparative Fit Index is commonly reported in Marketing studies, it is available in EQS, LISREL 8, and CALIS, and it ranges in value between 0 and 1 (values above .9 suggest adequate fit (Bollen 1990)). Since each replication was generated using the model being fitted, lack of fit (a Comparative Fit Index value less than .9) should have occurred at a chance level (in 10% or fewer replications). For reference purposes, model fit is also shown for a structural model involving only the linear terms. Replications involving this linear-terms-only model were generated in an identical manner as the Figure 3 replications, except that XX and XZ were not generated, and Y did not depend on XX or XZ in the population. Since these linear-terms-only replications were multivariate normal, their model fit results serve as a baseline "best case" to which to compare the Figure 3 model fit results. The results suggest that the proposed approach failed to fit the population model with generally the same incidence as the linear-terms-only replications. This result was also unexpected because it is believed that the χ^2 statistic (upon which the Comparative Fit Index is based) associated with ML estimates is not robust to departures from normality (see Bentler 1989, Bollen 1989, and Jöreskog and Sörbom 1989), and as a result, the incidence of lack of fit for the ML estimates should be different between the linear-terms-only replications and the proposed approach. Again, since the Figure 3 model produced by the proposed approach involved comparatively few non normal indicators, it is possible that the proposed approach's addition of relatively few nonlinear indicators to an otherwise normal model retains some robustness when ML estimates are used.

There were, however, differences between the ML, ML-Robust and ADF estimator results for model fit (see Table 6). The ML estimator rejected the Figure 3 population model at a chance rate. The ADF estimates, on the other hand, rejected the Figure 3 population model at rates that were considerably above chance in all the simulation conditions. For the low reliability, small sample condition, the Robust estimator rejected the model at rates that were slightly above chance. These results are generally consistent with those reported in Hu, Bentler and Kano (1992), where the ML and ML-Robust estimators performed about the same, and ADF estimates appeared to be sensitive to smaller sample sizes (i.e., less than 1000).

DISCUSSION

The assumption of normality for the linear indicators made in the proposed and the Kenny and Judd approaches - enabled the simplification of the variance calculations in these approaches, and this assumption cannot be relaxed for either approach. Consequently, the assessment of linear indicator normality is an important step in the use of these approaches. A reviewer suggested using the skewness and kurtosis tests in discussed in Bollen (1989, p. 418) to assess linear indicator normality. These tests are available in PRELIS (available with LISREL) and EQS, and involve determining the degree of skewness and kurtosis in the linear indicators singly and jointly (see Jöreskog and Sörbom 1993 p. 23, Bentler 1989 p. 227). To correct for linear indicator nonnormality, Bollen (1989) suggests transformation of the data (p. 425) (see Neter, Wasserman and Kunter 1988 for alternatives to the $\log_e$ transformation), and Bentler (1989) discusses the deletion of cases that contribute to nonnormality (p. 228). Bollen (1995) suggested using 2 stage least squares estimation, which does not assume multivariate normality.

However, the assumption of independent error terms for the indicators could be relaxed. The derivations in Appendix A would be changed by the addition of $\mathrm{Cov}(3\varepsilon_{xi},3\varepsilon_{zj})$ and $\mathrm{Cov}(3\varepsilon_{xk},3\varepsilon_{xm})$ terms, and the resulting covariance parameters could be estimated in a measurement model specifying correlated error terms.

In addition, the assumption of unidimensionality in the latent variables could be relaxed.[9] For constructs that are not sufficiently unidimensional to produce measurement parameter estimates for the linear latent variables that are "trivially different" between the measurement and structural models (e.g., different in the third decimal place), an iterative approach could be used. In this approach, the nonlinear loadings and error variances are recomputed using the structural model estimates of the equation (4) through (7) parameters and equations (4) through (7), and the structural model is reestimated. One to three of these iterations involving recomputation of equation (4) through (7) values using the latest structural equation estimates of the equation (4) through (7) parameters should be sufficient to produce exact effect estimates (i.e., equal to direct LISREL 8 estimates).

The strongest limitation of the proposed approach is shared with the Kenny and Judd, and Hayduk approaches: nonlinear indicators are not normal (Kenny and Judd 1984, Bollen 1989). This renders popular estimators such as ML and generalized least squares formally inappropriate because these estimators assume indicator normality. However, ML (and generalized least squares) estimates appear to be robust against departures from normality by the indicators (see Footnote 5). The results of the present study support this: the simulations involved indicators that were formally not normal, yet both the proposed approach and the Kenny and Judd approach recovered the population coefficients (Kenny and Judd 1984 and Hayduk 1987 reported similar findings).

The strongest limitation of the article is that a pedagogical example was not provided. The web site provides a pedagogical example in "Interactions and Quadratics in Latent Variable Associations: a Sourcebook for Advanced Survey Researchers," and an EXCEL spreadsheet for calculating single indicator values is available on the web site.

For emphasis, the proposed technique could be utilized in two ways. The baseline simulations were estimated using the proposed product indicators and LISREL 8 with and without constraint equations. This suggests that equations (4) and (5), or (6) and (7) could be specified as constants in a structural model using the two step technique. Alternatively, using the nonlinear constraint capabilities available in LISREL 8 and CALIS, these equations could be specified as variables in a structural model and estimated directly. Specifically, the loading of a product indicator in a latent variable model could be specified as a variable to be estimated, rather than specified as fixed, using equations (4) or (6) and a nonlinear constraint equation. The error term could also be specified as a variable to be estimated in a similar manner using equation (5) or (7). In this way, the proposed technique could be used with or without the two step approach as a parsimonious specification alternative for interaction and quadratic latent variables in situations where researchers in Marketing desired fewer indicators and constraint equations than required by the Kenny and Judd technique.

APPENDIX A- Var(x:z) and Var(x:x) Derivations

The variance of $X = x_1 + x_2$, where X is independent of ε_{x1} and ε_{x2}, ε_{x1} and ε_{x2} are independent of each other, and X is multivariate normal with zero mean, is given by the following:

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{Var}(x_1 + x_2) = \mathrm{Var}[(\lambda_{x1}\mathit{X} + \varepsilon_{x1}) + (\lambda_{x2}\mathit{X} + \varepsilon_{x2})] \\ &= \mathrm{Var}((\lambda_{x1} + \lambda_{x2})\mathit{X}) + \mathrm{Var}(\varepsilon_{x1}) + \mathrm{Var}(\varepsilon_{x2}) \\ &= \mathrm{Var}(\Gamma_X \mathit{X}) + \theta_X = \Gamma_X^2\mathrm{Var}(\mathit{X}) + \theta_X, \end{aligned}$$

where $\Gamma_X = \lambda_{x1} + \lambda_{x2}$, $\theta_X = \mathrm{Var}(\varepsilon_{x1}) + \mathrm{Var}(\varepsilon_{x2})$. By induction, $\mathrm{Var}(X) = \Gamma_X^2\mathrm{Var}(\mathit{X}) + \theta_X$, where $\mathrm{Var}(\mathit{X})$ is the variance of the latent variable *X*, $\Gamma_X = \lambda_{x1} + \lambda_{x2} + \ldots + \lambda_{zm}$, $\theta_X = \mathrm{Var}(\varepsilon_{x1}) + \mathrm{Var}(\varepsilon_{x2}) + \ldots + \mathrm{Var}(\varepsilon_{xm})$, and m is the number of indicators of *X*.

Since $\mathrm{Var}(X*X) = 2\mathrm{Var}(X)^2$ and $\mathrm{Var}(X*Z) = \mathrm{Var}(X)\mathrm{Var}(Z) + \mathrm{Cov}(X,Z)^2$ under the above assumptions (Kenny and Judd 1984),

$$\begin{aligned} \mathrm{Var}(x{:}x) &= \mathrm{Var}(X*X) = 2\mathrm{Var}(X)^2 = 2[\Gamma_X^2\mathrm{Var}(\mathit{X}) + \theta_X]^2 \\ &= \Gamma_x^4\mathrm{Var}(\mathit{XX}) + 4\Gamma_x^2\mathrm{Var}(\mathit{X})_x\, \theta_X + 2\theta_X^2 \\ &= \lambda_{x:x}^2\mathrm{Var}(\mathit{X}^2) + 2\theta_{\varepsilon x:x}\ , \end{aligned}$$

where $\lambda_{x:x} = \Gamma_x^2$ and $\theta_{\varepsilon x:x} = 4\Gamma_x^2\mathrm{Var}(\mathit{X})_x\, \theta_X + \theta_X^2$.

Similarly for Z meeting the same assumptions as X,

$$\begin{aligned} \mathrm{Var}(x{:}z) &= \mathrm{Var}(X)\mathrm{Var}(Z) + \mathrm{Cov}(X,Z)^2 \\ &= [\Gamma_X^2\mathrm{Var}(\mathit{X})+\theta_X][\Gamma_Z^2\mathrm{Var}(\mathit{Z})+\theta_Z] + [\Gamma_X\Gamma_Z\mathrm{Cov}(\mathit{X},\mathit{Z})]^2 \\ &= \Gamma_X^2\Gamma_Z^2[\mathrm{Var}(\mathit{X})\mathrm{Var}(\mathit{Z})+\mathrm{Cov}(\mathit{X},\mathit{Z})^2] + \Gamma_X^2\mathrm{Var}(\mathit{X})\theta_Z + \Gamma_Z^2\mathrm{Var}(\mathit{Z})\theta_X + \theta_X\theta_Z \\ &= \Gamma_X^2\Gamma_Z^2\mathrm{Var}(\mathit{XZ}) + \Gamma_X^2\mathrm{Var}(\mathit{X})\theta_Z + \Gamma_Z^2\mathrm{Var}(\mathit{Z})\theta_X + \theta_X\theta_Z \\ &= \lambda_{x:z}^2\mathrm{Var}(\mathit{XZ}) + \theta_{x:z}\ , \end{aligned}$$

where $\lambda_{x:z} = \Gamma_X\Gamma_Z$, and $\theta_{x:z} = \Gamma_X^2\mathrm{Var}(\mathit{X})\theta_Z + \Gamma_Z^2\mathrm{Var}(\mathit{Z})\theta_X + \theta_X\theta_Z$.

APPENDIX B- Synthetic Data Set Creation

The data for equation (8) was generated as follows. Let **M** be an n X 1 vector of random normal variates with mean 0 and variance 1, where n is the number of cases. The n X 3 matrix **P** with columns that were the population values for the n X 1 vectors **X**, **Z**, and **T** were determined by **P** = **M**(1 1 1)**C**', where (1 1 1) is a 1 X 3 unit vector and **C** is a lower triangular matrix such that

$$\mathbf{CC}' = \begin{bmatrix} V_X & & \\ r(V_Z V_X)^2 & V_Z & \\ r(V_T V_X)^2 & r(V_T V_Z)^2 & V_T \end{bmatrix}$$

where V_* is the variance of *, and r is the correlation between **X**, **Z** and **T**. The n X 4 matrices of observed values **x**, **z** and **t** for the population vectors **X**, **Z** and **T**, respectively, (an n X 2 matrix was used in the baseline simulation) were given by **x** = (.6**P**(1 0 0) + **N**($0,\theta_{\varepsilon x}$))(1 1 1 1) , **z** = (.6**P**(0 1 0) + **N**($0,\theta_{\varepsilon z}$))(1 1 1 1) and **t** = (.6**P**(0 0 1) + **N**($0,\theta_{\varepsilon t}$))(1 1 1 1), where (1 0 0), (0 1 0), (0 0 1), and (1 1 1 1) are vectors of 1's, and the N($0,\theta_{\varepsilon *}$)'s are n X 1 vectors of random normal variates with mean 0 and variance $\theta_{\varepsilon *}$ ((1 1 1 1) was replaced by (1 1) in the baseline simulation). The values for the n X 1 vector for dependent variable **Y** was determined by

$$\mathbf{Y} = b_X\mathbf{X} + b_Z\mathbf{Z} + b_T\mathbf{T} + b_{XX}\mathbf{XX} + b_{XZ}\mathbf{XZ} + \boldsymbol{\zeta}_{\mathbf{Y}} \,,$$

where the b_*'s are the scalar effects of * on **Y**, and $\boldsymbol{\zeta}_{\mathbf{Y}}$ is an n X 1 vector of random normal variates with mean 0 and variance equal to .16 .

REFERENCES

Aiken, Leona S. and Stephen G. West (1991), Multiple Regression: Testing and Interpreting Interactions, Newbury Park, CA: SAGE Publications.

Ajzen, Icek and Martin Fishbein (1980), Understanding Attitudes and Predicting Social Behavior, Englewood Cliffs, NJ: Prentice Hall.

Anderson, James C. and David W. Gerbing (1982), "Some Methods for Respecifying Measurement Models to Obtain Unidimensional Construct Measurement," Journal of Marketing Research, 19 (November), 453-60.

Anderson, James C. and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," Psychological Bulletin, 103 (May), 411-23.

Anderson, Theodore W. and Yasuo Amemiya (1985), The Asymptotic Normal Distribution of Estimators in Factor Analysis under General Conditions, Technical Report 12, Econometric Workshop, Stanford University.

Anderson, Theodore W. and Yasuo Amemiya (1986), "Asymptotic Distribution in Factor Analysis and Linear Structural Relations," Technical Report 18, Econometric Workshop, Stanford University.

Batra, Rajeev and Michael L. Ray (1986), "Affective Responses Mediating Acceptance of Advertising," Journal of Consumer Research," 13 (September), 249.

Bentler, Peter M. (1983), "Some Contributions to Efficient Statistics for Structural Models: Specification and Estimation of Moment Structures," Psychometrika, 48 (December), 493-517.

________ (1989), EQS Structural Equations Program Manual, Los Angeles: BMDP Software.

________ (1990), "Comparative Fit Indexes in Structural Models," Psychological Bulletin, 107 (March), 238-46.

________ and T. Dijkstra (1985), "Efficient Estimation via Linearization in Structural Models," in Multivariate Analysis VI, Paruchuri R. Kirshnaiah ed., Amsterdam: North-Holland.

Bollen, Kenneth A. (1989), Structural Equations with Latent Variables, New York: Wiley.

________ and J. Scott Long (1993), Testing Structural Equation Models, Newbury Park, CA: SAGE.

Bollen, Kenneth A.(1995), AStructural Equation Models that are Nonlinear in Latent Variables: A Least Squares Estimator,≅ Sociological Methodology, 25, 223-251.

Boomsma, Anne (1983), On the Robustness of LISREL (Maximum Likelihood Estimation) Against Small Sample Size and Nonnormality, Amsterdam: Sociometric Research Foundation.

Browne, Michael W. (1982), "Covariance Structures," in Topics in Applied Multivariate Analysis, Douglas M. Hawkins ed., Cambridge: Cambridge University Press.

________ (1987), "Robustness of Statistical Inference in Factor Analysis and Related Models," Biometrika, 74 (June), 375-84.

Burt, Ronald S. (1976), "Interpretational Confounding of Unobserved Variables in Structural Equation Models," Sociological Methods and Research, 5 (August), 3-52.

Busemeyer, Jerome R. and Lawrence E. Jones (1983), "Analysis of Multiplicative Combination Rules When the Causal Variables are Measured With Error," Psychological Bulletin, 93 (May), 549-62.

Cohen, Jacob and Patricia Cohen (1983), Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences, Hillsdale, NJ: Lawrence Erlbaum.

Dwyer, F. Robert, Paul H. Schurr, and Sejo Oh. 1987, "Developing Buyer-Seller Relationships," Journal of Marketing, 51 (April): 11-

27.

Engel, James F., Rodger D. Blackwell and David T. Kollat (1978), Consumer Behavior, 3rd. ed., Hinsdale, IL: Dryden.

Feucht, Thomas E. (1989), "Estimating Multiplicative Regression Terms in the Presence of Measurement Error," Sociological Methods and Research, 17 (February), 257-82.

Gerbing, David W. and James C. Anderson (1988), "An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment," Journal of Marketing Research, 25 (May), 186-92.

Harlow, Lawrence L. (1985), Behavior of Some Elliptical Theory Estimators with Nonnormal Data in a Covariance Structures Framework: A Monte Carlo Study, Unpublished Dissertation, University of California, Los Angeles.

Hayduk, Leslie A. (1987), Structural Equation Modeling with LISREL: Essential and Advances, Baltimore, MD: Johns Hopkins Press.

Heide, Jan B. and George John (1992), "Do Norms Matter in Marketing Relationships?" Journal of Marketing, 56 (April), 32-44.

Howard, John A. (1977), Consumer Behavior: Application of Theory, New York: McGraw Hill.

________ (1989), Consumer Behavior in Marketing Strategy, Englewood Cliffs, NJ: Prentice Hall.

________ and Jagdish N. Sheth (1969), The Theory of Buyer Behavior, New York: Wiley.

Hu, Li-tze, Peter M. Bentler and Yutaka Kano (1992), "Can Test Statistics in Covariance Structure Analysis Be Trusted?" Psychological Bulletin, 112 (September), 351-362.

Jaccard, James, Robert Turrisi and Choi K. Wan (1990), Interaction Effects in Multiple Regression, Newbury Park, CA: SAGE Publications.

Jaccard, James, Robert and C. K. Wan (1995), AMeasurement Error in the Analysis of Interaction Effects Between Continuous Predictors Using Multiple Regression: Multiple Indicator and Structural Equation Approaches,≅ Psychological Bulletin, 117 (2), 348-357.

Jöreskog, Karl G. (1993), "Testing Structural Equation Models," in Testing Structural Equation Models, Kenneth A. Bollen and J. Scott Long eds., Newbury Park, CA: SAGE.

________ and Dag Sörbom (1989), LISREL 7: A Guide to the Program and Applications, 2nd ed., Chicago: SPSS, Incorporated.

________ (1993), PRELIS 2 User's Reference Guide , Chicago: Scientific Software International, Inc.

Jöreskog, Karl G. and Fan Yang (1996), ANonlinear Structural Equation Models: The Kenny and Judd Model with Interaction Effects,≅ Advances in Structural Equation Modeling Techniques, G.A. Marcoulides, R.E. Schumacker, eds., Hillsdale, NJ: LEA.

Kenny, David and Charles M. Judd (1984), "Estimating the Nonlinear and Interactive Effects of Latent Variables," Psychological Bulletin, 96 (July), 201-10.

Kohli, Ajay K. (1989), "Effects of Supervisory Behavior: The Role of Individual Differences Among Salespeople," Journal of Marketing, 53 (October), 40-50.

Laroche, Michel and John A. Howard (1980), "Non-linear Relations in a Complex Model of Buyer Behavior," Journal of Consumer Research, 6 (March), 377-88.

Ping, R. (1996a), "Latent Variable Interaction and Quadratic Effect Estimation: A Two-Step Technique Using Structural Equation Analysis, Psychological Bulletin, 119 (January), 166-175.

Ping, R. (1996b), "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients," *Multivariate Behavioral Research*, 31 (1), 95-120.

Neter, John, William Wasserman and Michael H. Kunter (1985), Applied Linear Statistical Models, Homewood, IL: Irwin.

Sharma, Subhash, Richard M. Durand, and Oded Gur-Arie (1981), "Identification and Analysis of Moderator Variables," Journal of Marketing Research, 28 (August), 291-300.

Sharma, Subhash, Srinivas Durvasula and William R. Dillon (1989), "Some Results on the Behavior of Alternative Covariance Structure Estimation Procedures in the Presence of Nonnormal Data," Journal of Marketing Research, 24 (May), 214-21.

Sherif, Muzafer and Carl I. Hovland (1961), Social Judgement:Assimilation and Contrast Effects in Communication and Attitude Change, New Haven: Yale University Press.

Stern, Louis W. and Torger Reve (1980), "Distribution Channels as Political Economies: A Framework for Comparative Analysis," Journal of Marketing, 44 (Summer), 52-64.

Tanaka, Jeffrey S. (1984), Some Results on the Estimation of Covariance Structure Models, Unpublished Dissertation, University of California, Los Angeles.

Teas, R. Kenneth (1981), "An Empirical Test of Modelsof Salespersons' Job Expectancy and Instrumentality Perceptions," Journal of Marketing Research, 18 (May), 209-26.

Walker, Orville C., Gilbert A. Churchill and Neil M. Ford (1977), "Motivation and Performance in Industrial Selling: Present Knowledge and Needed Research," Journal of Marketing Research, 14 (May), 156-68.

Weitz, Barton A. (1981), "Effectiveness in Sales Interactions: A Contingency Framework," Journal of Marketing, 45 (Winter), 85-103

Table 1

BASELINE SIMULATION POPULATION CHARACTERISTICS

Parameter	Population Variance	Population Coefficient	Parameter	Population Variance	Population Coefficient
X	2.15		ζ_Y	0.16	
Z	1.60		λ_{x1}		1.00
T	1.00		λ_{x2}		0.60
Corr(X,Z)		0.60	λ_{z1}		1.00
Corr(X,T)		0.60	λ_{z2}		0.60
Corr(T,Z)		0.60	λ_{t1}		1.00
ε_{x1}	0.36	λ_{t2}		0.60	
ε_{x2}	0.81	$\beta_{Y,X}$		-0.15	
ε_{z1}	0.49	$\beta_{Y,Z}$		0.35	
ε_{z2}	0.64	$\beta_{Y,T}$		0.25	
ε_{t1}	0.15	$\beta_{Y,XZ}$		0.70	
ε_{t2}	0.55	$\beta_{Y,XX}$		-0.50	

Table 2
BASELINE SIMULATION ESTIMATION SUMMARY

Variable	Estimation Technique[a]	E(β)[b]	Bias[c] Amount	%	MSE[d]	Var(β)[e]
X	Pop. Value	-0.150				
	2 Step	-0.142	0.008	-5.3	0.016	0.02
	LISREL 8	-0.140	0.010	-6.7	0.017	0.02
	K&Judd	-0.145	0.005	-3.3	0.014	0.01
	Regression	-0.083	0.067	-44.7	0.010	0.01
Z	Pop. Value	0.350				
	2 Step	0.334	-0.016	-4.6	0.014	0.01
	LISREL 8	0.339	-0.011	-3.1	0.015	0.01
	K&Judd	0.338	-0.012	-3.4	0.015	0.01
	Regression	0.301	-0.049	-14.0	0.009	0.01
T	Pop. Value	0.250				
	2 Step	0.241	-0.009	-3.6	0.019	0.02
	LISREL 8	0.247	-0.003	-1.2	0.020	0.02
	K&Judd	0.252	0.002	0.8	0.020	0.02
	Regression	0.250	0.000	0.0	0.008	0.02
XZ	Pop. Value	0.700				
	2 Step	0.718	0.018	2.6	0.012	0.01
	LISREL 8	0.715	0.015	2.1	0.011	0.01
	K&Judd	0.709	0.009	1.3	0.013	0.01
	Regression	0.540	-0.160	-22.9	0.026	0.00
XX	Pop. Value	-0.500				
	2 Step	-0.513	-0.013	2.6	0.009	0.01
	LISREL 8	-0.508	-0.008	1.6	0.009	0.01
	K&Judd	-0.516	-0.016	3.2	0.008	0.01
	Regression	-0.370	0.130	-26.0	0.019	0.00
All Coeff.	2 Step		0.013		0.014	0.01
	LISREL 8		0.010		0.014	0.01
	K&Judd		0.010		0.014	0.01
	Regression		0.099		0.014	0.01
Non-Lin. Coeff.	2 Step		0.016		0.011	0.01
	LISREL		0.012		0.010	0.01
	K&Judd		0.013		0.011	0.01
	Regression		0.146		0.023	0.00
			(RMS)[f]		(Avg)[g]	(Avg)[g]

[a] 2 Step = LISREL 8 estimates using the proposed two step approach. LISREL 8 = LISREL 8 estimates using the proposed single indicator specification and LISREL 8 constraint equations. K&Judd = LISREL 8 estimates using a Kenny and Judd (1984) specification and constraint equations.
[b] Average of coefficient estimates across all data sets.
[c] E(β) minus the population value.
[d] Mean Square Error = ∋ (population value - estimated value)^2/n .
[e] Variance of β = ∋(E(β) - estimated value)^2/n .
[f] RMS of column entries.
[g] Average of column entries.

Table 3
VARYING CONDITIONS SIMULATIONS POPULATION CHARACTERISTICS

Parameter[a]	Population Variance	Population Coefficient
All Data Sets:		
X	2.15	
Z	1.60	
T	1.00	
Corr(X,Z)		0.60
Corr(X,T)		0.60
Corr(T,Z)		0.60
ζ_Y	0.16	
$\beta_{Y,X}$	-0.15	
$\beta_{Y,Z}$	0.35	
$\beta_{Y,T}$	0.25	
High Reliability Samples:		
λ_{x1}		1.00
$\lambda_{x2}-\lambda_{x4}$		0.90
λ_{z1}		1.00
$\lambda_{z2}-\lambda_{z4}$		0.90
λ_{t1}		1.00
$\lambda_{t2}-\lambda_{t4}$		0.90
$\varepsilon_{x1}-\varepsilon_{x4}$	0.82	
$\varepsilon_{z1}-\varepsilon_{z4}$	0.61	
$\varepsilon_{t1}-\varepsilon_{t4}$	0.38	
Low Reliability Samples:		
λ_{x1}		1.00
$\lambda_{x2}-\lambda_{x4}$		0.60
λ_{z1}		1.00
$\lambda_{z2}-\lambda_{z4}$		0.60
λ_{t1}		1.00
$\lambda_{t2}-\lambda_{t4}$		0.60
$\varepsilon_{x1}-\varepsilon_{x4}$	2.81	
$\varepsilon_{z1}-\varepsilon_{z4}$	2.09	
$\varepsilon_{t1}-\varepsilon_{t4}$	1.31	
Small Nonlinear Coefficients:		
$\beta_{Y,XZ}$	-0.25	
$\beta_{Y,XX}$	0.15	
Large Nonlinear Coefficients:		
$\beta_{Y,XZ}$	0.70	
$\beta_{Y,XX}$	-0.50	

[a] $Y = \beta_{Y,X}X + \beta_{Y,Z}Z + \beta_{Y,T}T + \beta_{Y,XZ}XZ + \beta_{Y,XX}XX + \zeta_Y$
$x_i = \lambda_{xi}X + \varepsilon_{xi}$ (i=1,4)
$z_j = \lambda_{zj}Z + \varepsilon_{zj}$ (j=1,4)
$t_k = \lambda_{tk}T + \varepsilon_{tk}$ (k=1,4) .

Table 4

COEFFICIENT ESTIMATION SUMMARY FOR VARYING CONDITIONS SIMULATIONS

		Coefficient Estimate				
Simulation	Variable/		Bias[c]			
Condition[a]	Coefficient	E(β)[b]	Amount	%	MSE[d]	Var(β)[e]
SβLρ100	X -0.15	-0.160	-0.010	6.67%	0.057	0.05
LβLρ100	X -0.15	-0.161	-0.011	7.33%	0.063	0.06
SβLρ100	Z 0.35	0.330	-0.020	-5.71%	0.051	0.05
LβLρ100	Z 0.35	0.335	-0.015	-4.29%	0.055	0.05
SβLρ100	T 0.25	0.268	0.018	7.20%	0.061	0.06
LβLρ100	T 0.25	0.235	-0.015	-6.00%	0.056	0.05
LβLρ100	XZ 0.70	0.684	-0.016	-2.29%	0.054	0.05
SβLρ100	XZ -0.25	-0.237	0.013	-5.20%	0.059	0.05
LβLρ100	XX -0.50	-0.481	0.019	-3.80%	0.051	0.05
SβLρ100	XX 0.15	0.134	-0.013	-8.67%	0.060	0.06
SβHρ100	X -0.15	-0.148	0.002	-1.33%	0.003	0.00
LβHρ100	X -0.15	-0.154	-0.004	2.67%	0.004	0.00
SβHρ100	Z 0.35	0.349	-0.001	-0.29%	0.006	0.00
LβHρ100	Z 0.35	0.353	0.003	0.86%	0.002	0.00
SβHρ100	T 0.25	0.248	-0.002	-0.80%	0.003	0.00
LβHρ100	T 0.25	0.252	0.002	0.80%	0.002	0.00
LβHρ100	XZ 0.70	0.697	-0.003	-0.43%	0.005	0.00
SβHρ100	XZ -0.25	-0.254	-0.004	1.60%	0.004	0.00
LβHρ100	XX -0.50	-0.497	0.003	-0.60%	0.002	0.00
SβHρ100	XX 0.15	0.152	0.002	1.33%	0.003	0.00
SβLρ300	X -0.15	-0.145	0.005	-3.33%	0.013	0.01
LβLρ300	X -0.15	-0.153	-0.003	2.00%	0.012	0.01
SβLρ300	Z 0.35	0.360	0.010	2.86%	0.019	0.01
LβLρ300	Z 0.35	0.343	-0.007	-2.00%	0.015	0.01
SβLρ300	T 0.25	0.245	-0.005	-2.00%	0.009	0.00
LβLρ300	T 0.25	0.246	-0.004	-1.60%	0.011	0.01
LβLρ300	XZ 0.70	0.703	0.003	0.43%	0.012	0.01
SβLρ300	XZ -0.25	-0.248	0.002	-0.80%	0.015	0.01
LβLρ300	XX -0.50	-0.505	-0.005	1.00%	0.012	0.01
SβLρ300	XX 0.15	0.148	-0.002	-1.33%	0.012	0.01
SβHρ300	X -0.15	-0.149	0.001	-0.67%	0.002	0.00
LβHρ300	X -0.15	-0.150	0.000	0.00%	0.001	0.00
SβHρ300	Z 0.35	0.350	0.000	0.00%	0.001	0.00
LβHρ300	Z 0.35	0.349	-0.001	-0.29%	0.001	0.00
SβHρ300	T 0.25	0.249	-0.001	-0.40%	0.001	0.00
LβHρ300	T 0.25	0.251	0.001	0.40%	0.002	0.00
LβHρ300	XZ 0.70	0.700	0.000	0.00%	0.002	0.00
SβHρ300	XZ -0.25	-0.249	0.001	-0.40%	0.001	0.00
LβHρ300	XX -0.50	-0.500	0.000	0.00%	0.002	0.00
SβHρ300	XX 0.15	0.149	-0.001	-0.67%	0.001	0.00

[a] SβLρ100 = Small nonlinear β's, Low reliability (ρ), 100 case sample, LβHρ300 = Large nonlinear β's, High reliability (ρ), 300 case sample, etc.

[b] Average of coefficient estimates across all data sets.

[c] E(β) minus the population value.

[d] Mean Square Error = э (population value - estimated value)^2/n .

[e] Variance of β = э(E(β) - estimated value)^2/n .

Table 5
STANDARD ERROR SUMMARY FOR VARYING CONDITIONS SIMULATIONS

Simulation Condition[a]	Variable/ Coefficient		% False Negatives[b] ML	ML/ROBUST	ADF
SβLρ100	X	-0.15	83%	74%	[c]
LβLρ100	X	-0.15	81%	73%	
SβLρ100	Z	0.35	70%	71%	
LβLρ100	Z	0.35	72%	72%	
SβLρ100	T	0.25	74%	77%	
LβLρ100	T	0.25	75%	79%	
LβLρ100	XZ	0.70	17%	16%	
SβLρ100	XZ	-0.25	41%	31%	
LβLρ100	XX	-0.50	26%	28%	
SβLρ100	XX	0.15	66%	59%	
SβHρ100	X	-0.15	14%	12%	[c]
LβHρ100	X	-0.15	15%	13%	
SβHρ100	Z	0.35	11%	10%	
LβHρ100	Z	0.35	10%	10%	
SβHρ100	T	0.25	10%	10%	
LβHρ100	T	0.25	10%	11%	
LβHρ100	XZ	0.70	00%	00%	
SβHρ100	XZ	-0.25	00%	00%	
LβHρ100	XX	-0.50	00%	00%	
SβHρ100	XX	0.15	00%	00%	
SβLρ300	X	-0.15	31%	29%	30%
LβLρ300	X	-0.15	32%	27%	29%
SβLρ300	Z	0.35	28%	20%	23%
LβLρ300	Z	0.35	26%	22%	22%
SβLρ300	T	0.25	35%	33%	32%
LβLρ300	T	0.25	36%	31%	30%
LβLρ300	XZ	0.70	00%	00%	00%
SβLρ300	XZ	-0.25	27%	28%	25%
LβLρ300	XX	-0.50	00%	00%	00%
SβLρ300	XX	0.15	31%	27%	26%
SβHρ300	X	-0.15	00%	00%	00%
LβHρ300	X	-0.15	00%	00%	00%
SβHρ300	Z	0.35	00%	00%	00%
LβHρ300	Z	0.35	00%	00%	00%
SβHρ300	T	0.25	00%	00%	00%
LβHρ300	T	0.25	00%	00%	00%
LβHρ300	XZ	0.70	00%	00%	00%
SβHρ300	XZ	-0.25	00%	00%	00%
LβHρ300	XX	-0.50	00%	00%	00%
SβHρ300	XX	0.15	00%	00%	00%

---

[a] SβLρ100 = Small nonlinear β's, Low reliability (ρ), 100 case sample, LβHρ300 = Large nonlinear β's, High reliability (ρ), 300 case sample.
[b] Percent NS coefficient estimates based on t∃2.
[c] Not available in EQS when the sample size is less than the number of unique elements in the sample covariance matrix.

Table 6

MODEL FIT SUMMARY FOR VARYING CONDITIONS SIMULATIONS

	Percent Rejected Models Based on CFI Statistic[b]											
	Linear-terms-only Model						Figure 3 Model					
	CFI<.9			CFI<.95			CFI<.9			CFI<.95		
Simulation Condition[a]	ML	ML/ Robust	ADF	ML	ML/ Robust	ADF	ML	ML/ Robust	ADF	ML	ML/ Robust	ADF
SβLρ100	3	9	[c]	11	21	[c]	1	13	[c]	15	17	[c]
SβHρ100	1	3	[c]	2	4	[c]	1	2	[c]	4	1	[c]
SβLρ300	2	3	59	3	1	90	2	2	57	2	4	94
SβHρ300	3	1	30	3	2	78	2	2	35	3	4	82
LβLρ100	1	11	[c]	8	20	[c]	3	12	[c]	17	22	[c]
LβHρ100	2	2	[c]	3	2	[c]	1	1	[c]	1	3	[c]
LβLρ300	2	3	55	4	2	85	1	2	61	2	2	91
LβHρ300	2	1	34	2	3	74	2	2	26	4	3	79

[a] SβLρ100 = Small nonlinear β's, Low reliability (ρ), 100 case sample, LβHρ300 = Large nonlinear β's, High reliability (ρ), 300 case sample.

[b] Percent of replications in which the model did not fit the data, based on the CFI (Bentler 1990) statistic (lack of fit= the CFI value was < .9 or < .95).

[c] Not available in EQS when the sample size is less than the number of unique elements in the sample covariance matrix.

Figure 1
A PRODUCT INDICATOR INTERACTION AND QUADRATIC MODEL

(Click here to view Figure 1)

Figure 2
A SINGLE PRODUCT INDICATOR INTERACTION AND QUADRATIC MODEL

(Click here to view Figure 2)

Figure 3
SYNTHETIC DATA STRUCTURAL MODEL

(Click here to view Figure 3)

ENDNOTES

1. There are now several more, and most of them are refinements of the Kenny and Judd (1984) approach (which will be described next). Bollen 1995 proposed using a two-stage least squares estimator. Jöreskog and Yang (1996) provided additional details for the Kenny and Judd approach. Jaccard and Wan (1995) suggested using a subset of the Kenny and Judd indicators. Ping (1996a) suggested a two step refinement of the Kenny and Judd approach, and Ping (1996b) suggested an errors-in-variables regression approach.

2.The number of replications was a compromise between practicality and a very large number of replications. The simulations could not be automated, and each replication required the equivalent of three computer runs punctuated with manual coding activities.

3. The X, Z and constructs were generated to be unidimensional, and the indicators were created to meet the Kenny and Judd normality conditions.

4. EXCEL templates are available from the author (see the web site) to accomplish the equation (4) through (7) calculations.

5. Product indicators are not multivariate normal (Bollen 1989, Kenny and Judd 1984), and ML and generalized least squares estimates are not formally appropriate in this application because they assume multivariate normality (Bollen 1989). However, ML and generalized least squares estimates appear to be robust to departures from normality (Anderson and Amemiya 1985, 1986; Bollen 1989; Boomsma 1983; Browne 1987; Harlow 1985; Sharma, Durvasula and Dillon 1989; Tanaka 1984).

6. These results provide an empirical demonstration of the efficacy of the Kenny and Judd technique. As far as I know (in 1993) the only other evidence is provided by two data sets reported in Kenny and Judd's original article (Jaccard and Wan 1995 also demonstrated the efficacy of the Kenny and Judd technique).

7. The unbiased regression estimates for T appear to be the result of a chance combination of the population intercorrelations among X, Z and T and lack of sample intercorrelations with XX and XZ.

8. (In 1993) As far as I know this study provides first direct comparison of regression and structural equation estimates for models with nonlinear variables (Jaccard and Wan 1995 also compared regression and Kenny and Judd estimates). While regression estimates are known to be inefficient for variables measured with error, based on these results they may nevertheless be less variable than structural equation estimates for linear variables under these conditions.

9. Procedures for obtaining unidimensionality are suggested in Anderson and Gerbing 1982 p. 454, Gerbing and Anderson 1988, and Jöreskog 1993 p. 313 (see the discussion of the need for unidimensionality in Anderson and Gerbing 1988, Burt 1976, and Jöreskog 1993).

LATENT VARIABLE INTERACTION AND QUADRATIC EFFECT ESTIMATION: A TWO-STEP TECHNIQUE USING STRUCTURAL EQUATION ANALYSIS

(An earlier, but revised, version of Ping 1996, *Psychological Bulletin*, 119 (January), 166-175)
(Updated July, 2001)

ABSTRACT

This paper proposes an alternative estimation technique for latent variable interactions and quadratics that is useful with EQS and AMOS. First, measurement parameters for indicator loadings and errors of linear latent variables are estimated in a measurement model that excludes the interaction and quadratic variables. Next, these estimates are used to calculate values for the indicator loadings and error variances of the interaction and quadratic latent variables. Finally, these calculated values are specified as constants in the structural model containing the interaction and quadratic variables.

Interaction and quadratic effects are routinely reported in ANOVA to aid in the interpretation of significant main effects. However, interaction and quadratic effects are less frequently reported for survey data. Thus, researchers have called for the inclusion of interaction and quadratic variables in survey data analysis (Aiken & West, 1991; Cohen & Cohen, 1975, 1983; Jaccard, Turrisi & Wan, 1990). However, for unobserved or latent variables, there has been no adequate method of estimating interaction and quadratic effects until recently. Kenny and Judd (1984) proposed that, under certain conditions, interaction and quadratic latent variables could be adequately specified using products of indicators. They demonstrated their proposed technique using COSAN (McDonald, 1978) (now available in the SAS procedure CALIS), because at the time COSAN was the only structural equation software that accommodated the nonlinear constraints required to estimate these variables.

Hayduk (1987) subsequently implemented the Kenny and Judd technique using LISREL 7. However, the Hayduk approach required the specification of many additional latent variables to account for the loadings and error variances of the nonlinear indicators. The result is that the specification of latent variable interactions and quadratics is a tedious and error prone process in COSAN, EQS and AMOS.[1]

Recently, LISREL 8 provided a nonlinear constraint capability that can be used to implement the Kenny and Judd technique. However, EQS and AMOS have yet to provide an equivalent method of implementing the Kenny and Judd approach.

This paper proposes an alternative to the Hayduk technique that can be used with all structural equation analysis software. Because it creates no additional variables or equations, the proposed technique may be useful to EQS, AMOS and LISREL 7 users, which do not directly model these latent variables.

The proposed technique is implemented in two-steps. For indicators in mean deviation form,[2] loadings and error variances for the indicators of linear latent variables are estimated in a first-step measurement model. Then the nonlinear indicators of interaction and quadratic latent variables are created as products of the indicators of linear latent variables, as Kenny and Judd (1984) suggested. Next the loadings and error variances for these product indicators are calculated using the first step measurement model estimates plus equations derived from Kenny and Judd (1984) results. Finally the relations among the linear, interaction, and quadratic latent variables are estimated, using a second-step structural model in which these calculated loadings and error variances are specified as constants.

The balance of the paper describes this technique.

QUADRATIC AND INTERACTION EFFECT ESTIMATION

For latent variables X and Z, with indicators x_1, x_2, z_1 and z_2, Kenny and Judd (1984) proposed the interaction latent variable XZ could be specified with the product indicators x_1z_1, x_1z_2, x_2z_1, and x_2z_2. They also showed that the variance of product indicators such as x_1z_1 depends on measurement parameters associated with X and Z. Assuming that each of the latent variables X and Z is normally distributed and independent of the errors (ε_{x1}, ε_{x2}, ε_{z1}, and ε_{z2}) (X and Z may be correlated), that the errors are mutually independent, and that the indicators and the errors are normally distributed and in mean deviation form (i.e. have means of zero), the variance of the product indicator x_1z_1 is given by

$$Var(x_1z_1) = Var[(\lambda_{x1}X + \varepsilon_{x1})(\lambda_{z1}Z + \varepsilon_{z1})]$$

$$= \lambda_{x1}^2\lambda_{z1}^2 Var(XZ) + \lambda_{x1}^2 Var(X)Var(\varepsilon_{z1}) + \lambda_{z1}^2 Var(Z)Var(\varepsilon_{x1}) + Var(\varepsilon_{x1})Var(\varepsilon_{z1}) \quad (1)$$

$$= \lambda_{x1}^2\lambda_{z1}^2[Var(X)Var(Z) + Cov^2(X,Z)] + \lambda_{x1}^2 Var(X)Var(\varepsilon_{z1}) + \lambda_{z1}^2 Var(Z)Var(\varepsilon_{x1}) + Var(\varepsilon_{x1})Var(\varepsilon_{z1}) , \quad (1a)$$

for x_1 and z_1 with expected values of zero. In equations 1 and 1a λ_{x1} and λ_{z1} are the loadings of x_1 and z_1 on X and Z; ε_{x1} and ε_{z1} are the error terms for x_1 and z_1; Var(X), Var(Z), Var(x_1z_1), Var(ε_{x1}), and Var(ε_{z1}) are the variances of X, Z, x_1z_1, ε_{x1}, and ε_{z1}, respectively; and Cov(X,Z) is the covariance of X and Z.

In the quadratic case (where X = Z), the variance of the product indicator x_1x_1 is given by

$$Var(x_1x_1) = Var[(\lambda_{x1}X + \varepsilon_{x1})(\lambda_{x1}X + \varepsilon_{x1})]$$

$$= \lambda_{x1}^2\lambda_{x1}^2 Var(X^2) + 4\lambda_{x1}^2 Var(X)Var(\varepsilon_{x1}) + Var(\varepsilon_{x1}) , \quad (2)$$

$$= 2\lambda_{x1}^2\lambda_{x1}^2 Var^2(X) + 4\lambda_{x1}^2 Var(X)Var(\varepsilon_{x1}) + 2Var^2(\varepsilon_{x1}) . \quad (2a)$$

Kenny and Judd then specified equations 1a and 2a using COSAN by creating additional variables for the terms in these equations. For example, equation 1a required five additional variables, one each for $\lambda_x\lambda_z$, $Var(X)Var(Z)+Cov^2(X,Z)$, $Var(X)Var(\varepsilon_z)$, $Var(Z)Var(\varepsilon_x)$, and $Var(\varepsilon_x)Var(\varepsilon_z)$. These additional variables were then specified (constrained) to equal their respective equation 1a terms for COSAN estimation. Some creativity is required, however, to estimate equations 1 and 2 with EQS, AMOS and LISREL 7,[3] because these software products are not able to specify the nonlinear (product) terms in equation 1 or 2.

Hayduk's contribution was to provide a LISREL implementation of the Kenny and Judd technique. Hayduk's approach was to create additional latent variables to specify, for example, the right-hand side of equation 1a. It is difficult to do justice to Hayduk's approach in a few sentences, and the interested reader is directed to Hayduk (1987) Chapter 7 for details. In summary, to specify the first term of equation 1a, Hayduk created a chain of additional latent variables that affected the indicator x_1z_1. Using three additional chains of latent variables, the remaining three terms in equation 1a can be specified.

For a latent variable with many indicators, or for a model with several interaction or quadratic latent variables, the Hayduk approach of adding variables is arduous. For example, the single interaction model shown in Figure 1 requires an additional thirty latent variables to specify the loadings and error variances of the indicators for XZ.

As a result, researchers may find the specification of the additional variables or constraint equations[4] required by the Hayduk technique difficult. The number of additional variables required or generated by these techniques may also lead to estimation difficulties produced by the large matrices required to specify these additional variables.

The next section proposes a technique that requires the specification of no additional variables or constraint equations.

A PROPOSED ESTIMATION TECHNIQUE

In estimating structural equation models, Anderson and Gerbing (1988) proposed that the measurement specification of the model should be assessed separately from its structural specification, to ensure the unidimensionality of each of the latent variables in the model. This they argued avoids interpretational confounding (Burt, 1976), the interaction of the measurement and structural models. Interpretational confounding produces marked changes in the estimates of the measurement parameters when alternative structural models are estimated. They also noted that when a latent variable is unidimensional, the measurement parameter estimates for that latent variable should change trivially, if at all, between the measurement and structural model estimations (p. 418).

As a result, if X and Z are each unidimensional, that is their indicators have only one underlying construct each (Aker & Bagozzi, 1979; Anderson & Gerbing, 1988; Burt, 1973; Hattie, 1985; Jöreskog, 1970, 1971; McDonald, 1981), estimates of the parameters appearing in equations 1 and 2 are available in a measurement model that contains X and Z but excludes XX and XZ. To explain this result, the measurement parameters of a unidimensional latent variable are by definition unaffected by the presence or absence of other latent variables in a structural model. Consequently, other latent variables can be added or deleted from a measurement or structural model containing a unidimensional latent variable with no effect on the measurement parameter estimates for that latent variable. Thus if X and Z are unidimensional, the parameter estimates for equation 1 or 2 could be obtained from a measurement model that excludes XX and XZ. Similarly, the addition of XX and/or XZ to a structural model does not affect the measurement parameter estimates of X or Z in this structural model if X and Z are unidimensional.

As a result, the equation 1 and 2 loadings and error variances for product indicators such as x_1z_1 and x_1x_1 could be calculated using parameter estimates from a measurement model that excludes XX and XZ. Because these measurement parameter estimates should change trivially, if at all, between the measurement and structural model estimations (Anderson and Gerbing, 1988), these calculated loadings and error variances could then be used as fixed values (constants) in a structural equation model containing the interaction and quadratic latent variables XX and XZ.

In particular, for indicators in mean deviation form and under the Kenny and Judd normality assumptions stated in conjunction with equation 1, equations 1 and 2 can be simplified to

$$\mathrm{Var}(xz) = a^2\mathrm{Var}(XZ) + \mathrm{Var}(b)\ . \qquad (3)$$

In equation 3 Var(xz) is the variance of the indicator xz, Var(XZ) is the variance of the latent variable XZ, and $a = \lambda_x\lambda_z$. Var(b), the error variance for xz, is given by $\mathrm{Var}(b) = K\lambda_x^2\mathrm{Var}(X)\mathrm{Var}(\varepsilon_z) + K\lambda_z^2\mathrm{Var}(Z)\mathrm{Var}(\varepsilon_x) + K\mathrm{Var}(\varepsilon_x)\mathrm{Var}(\varepsilon_z)$, (K=2 if x=z, K=1 otherwise). Then if X and Z are each unidimensional, values for the loading "a" and the error variance for xz, Var(b), can be calculated using measurement model estimates for λ_x, λ_z, Var(X), Var(Z), $\mathrm{Var}(\varepsilon_x)$, and $\mathrm{Var}(\varepsilon_z)$. The loading and error variance of xz can subsequently be specified using these calculated values as fixed (constant) terms in a structural model involving XX and/or XZ, instead of variables to be estimated as the Kenny and Judd (1984) technique requires.

Consequently, the Figure 1 structural model could be estimated by setting the loadings and error variances for the product indicators equal to constants that are calculated using equation 3 and parameter estimates from a linear-latent-variable-only measurement model involving only X, Z and Y.

To illustrate this technique, the results of two tests of the technique's recovery of known parameters are presented.

EXAMPLES

ARTIFICIAL DATA SETS

Method The proposed technique was used to recover known parameters in two artificial data sets. Using a normal random number generator, two sets of 500 cases were created. One set of 500 cases contained values based on the Table 1 population characteristics for x_1, x_2, and Y in the Figure 2 quadratic model. The other set of 500 cases contained values based on the Table 1 population characteristics for x_1, x_2, z_1, z_2, and Y in the Figure 3 interaction model. These data sets were generated to meet the Kenny and Judd normality and mean deviation assumptions stated in conjunction with equation 1.

The covariance matrices for these two data sets are shown in Table 2. The Figure 2 structural model was specified by first estimating the parameters in a linear-latent-variable-only measurement model that excluded XX. Next the equation 3 loadings ("a's") and error variances (Var(b)'s), for the product indicators in Figure 2 were calculated using parameter estimates from this linear-latent-variable-only measurement model. Finally the Figure 2 structural model was estimated with the loadings and error variances of the nonlinear latent variables fixed at their respective "a" and Var(b) values.

The linear-latent-variable-only measurement model associated with the Figure 2 model was estimated using LISREL 7 and maximum likelihood. This produced the Table 3 estimates for the λ's, Var(ε)'s, and Var(X) to be used in calculating the equation 3 values for the product indicators of XX. Next the equation 3 values for $a_{x1,x1}$, $a_{x1,x2}$, $a_{x2,x2}$, $\mathrm{Var}(b_{x1,x1})$, $\mathrm{Var}(b_{x1,x2})$, and $\mathrm{Var}(b_{x2,x2})$ were computed (see Figure 2 for the equations, and Table 3 for the values and example calculations). Then the structural model shown in Figure 2 was specified by fixing the loading and error variance for each product indicator to the appropriate "a" and Var(b) values computed in the previous step. The results of the Figure 2 structural model estimation using LISREL 7 and maximum likelihood are shown in Table 4.

We repeated this process for the interaction model shown in Figure 3, and obtained the results shown in Tables 3 and 4.

To obtain a basis for comparing the efficacy of the proposed technique, Kenny and Judd, and Hayduk estimates were also generated. These estimates used the Figure 2 and 3 models. The Kenny and Judd estimates were produced using COSAN and generalized least squares, and the Hayduk estimates utilized LISREL 7 and maximum likelihood. The results are shown in Table 4.

Results The three estimation techniques produced essentially equivalent parameter estimates. The estimates were within a few points of the population values and each other. The squared average deviations from the population values (MSE's in Table 4) produced by each technique were also within a few points of each other. For the quadratic model, the overall MSE values for the three techniques (MSE-all parameters in the Quadratic Term Model portion of Table 4) were nearly identical. The MSE for the quadratic effect coefficients produced by the proposed technique (MSE-γ's) was slightly smaller than it was for the Hayduk and Kenny and Judd techniques. In the interaction model portion of the table, the all-parameter MSE's were also within a few points of each other. However, the all-parameter MSE's were slightly larger than they were for the quadratic model, the effect coefficient MSE's were smaller, and the Kenny and Judd technique produced the smallest effect coefficient MSE. Combining the parameter estimates for the two models (see the "Overall:" section of Table 4), the proposed technique produced MSE's that were the same or slightly smaller than the Hayduk and Kenny and Judd techniques.

To illustrate the use of the proposed technique, a field survey data analysis involving nonlinear latent variables is presented.

A FIELD SURVEY

Method. As part of a larger study of a social exchange view of long term buyer-seller relationships involving business firms, data were gathered from key informants in retailing firms concerning their loyalty to their primary economic exchange partner, their primary wholesaler; their satisfaction with that economic exchange partner, and the attractiveness of the best alternative wholesaler. Relationship satisfaction (SAT) and alternative attractiveness (ALT) were hypothesized to affect loyalty (LOY) (see Ping, 1993; Rusbult, Zembrodt & Gunn, 1982).

Since this is an illustration of the use of the proposed estimation technique, the study will simply be sketched. SAT, ALT and LOY were measured using multiple item Likert measures. The survey responses were used to create indicators of the independent variables (i.e., SAT and ALT) that were in mean deviation form. The responses were then used to assess the unidimensionality of SAT, ALT and LOY. They were also used to gauge the normality of the linear indicators using the skewness and kurtosis tests in LISREL 7's PRELIS.

Values for the product indicators were created for each survey response by forming all unique products of the values of the appropriate indicators of the linear latent variables, then appending these products to the response (see the comments regarding the formation of these indicators at the foot of Table 6). Next the linear-latent-variable-only measurement model for the Figure 4 model (i.e., with SAT, ALT and LOY only) was estimated. This was accomplished using the Table 5 variance-covariance matrix, maximum likelihood, and LISREL 7. The resulting measurement parameter estimates for the equation 3 "a's" and Var(b)'s are shown in Table 6.

The structural equation was then estimated. This was accomplished by calculating the Figure 4 product indicator loadings and error variances ("a's" and Var(b)'s), using the Table 6 measurement model estimates and equation 3 (see Table 6).[5] Then the loadings and error variances for the product indicators were fixed at these calculated values in the structural model, and the structural equation estimates shown in Table 7 were then produced using LISREL 7 and maximum likelihood. Table 7 also shows the maximum likelihood estimates using the Kenny and Judd technique for comparison.

Discussion. The estimates produced by the Kenny and Judd technique and the proposed technique were again similar. While some were higher and some were lower, the calculated "a's" and Var(b)'s produced by the proposed technique were within a few points of the Kenny and Judd estimates for the loadings and error variances of the product indicators. Similarly, the structural effect coefficients (γ=s) for the two techniques were comparable.

DISCUSSION

As the results in Tables 6 and 7 show, the measurement parameter estimates for the unidimensional SAT and ALT variables changed trivially between the linear-latent-variable-only measurement model and the Figure 4 structural model that contained the linear and nonlinear latent variables. Procedures for obtaining unidimensionality are suggested in Anderson and Gerbing (1982), Gerbing and Anderson (1988), and Jöreskog (1993). While there is no agreement on the detailed steps, the process of obtaining unidimensionality must balance concern for the content validity of a measure with its consistency. In the field survey example the estimation of single construct measurement models (Jöreskog, 1993) with a target comparative fit index (Bentler, 1990) of .99 produced the desired trivial difference in measurement parameters between the measurement and structural models.[6]

Had more than a trivial change been observed in the measurement parameters for the linear latent variables between the measurement and structural models (i.e., differences in the second decimal place), measurement parameter estimates for the linear latent variables from the previous structural model to recompute the "a's" and Var(b)'s, and thereby "converge" to the desired trivial change between structural model estimates of the measurement parameters for the linear latent variables[7].

The assumption that the error terms for linear indicators are independent can be relaxed. In equation 3 the expression for Var(b) would be changed by an additional covariance term (available in the measurement model) as follows,

$$\mathrm{Var}(b) = K\lambda_x^2\mathrm{Var}(X)\mathrm{Var}(\varepsilon_z) + K\lambda_z^2\mathrm{Var}(Z)\mathrm{Var}(\varepsilon_x) + K\mathrm{Var}(\varepsilon_x)\mathrm{Var}(\varepsilon_z) + (2K)\lambda_x\lambda_z\mathrm{Cov}(X,Z)\mathrm{Cov}(\varepsilon_x,\varepsilon_z)\ ,$$

(K=2 if x=z, K=1 otherwise).

LIMITATIONS

Just as in the Hayduk and Kenny and Judd techniques, the assumption of normality in the linear indicators cannot be relaxed. The derivation of the "a" and Var(b) terms is based on this assumption. Bollen (1989, Ch. 9) discusses appropriate normality tests involving indicator skewness and kurtosis. EQS and LISREL's PRELIS have implemented several of these tests. However, for typical sample sizes used in structural equation analysis, even small deviations from normality are likely to be statistically significant (Bentler, 1989). In addition, there is little guidance for determining when statistical nonnormality becomes practical nonnormality (Bentler, 1989). As a result, while the survey items were judged to be Anot nonnormal,≅ several items were statistically nonnormal using standard skewness and kurtosis tests (although the coefficients were not unreasonably large). In addition, the Mardia (1970) coefficient of multivariate nonnormality was significant (although not excessively so).

The robustness of the proposed, Hayduk and Kenny and Judd techniques to departures from normality is not known, and unreasonable departures from multivariate normality should be remedied. Bollen (1989) suggests transformation of the data (p. 425) (see Neter, Wasserman and Kunter, 1988 for alternatives to the $\log_e$ transformation), and Bentler (1989) discusses the deletion of cases that contribute to nonnormality (p. 228).

In the proposed, Hayduk, and Kenny and Judd techniques the product indicators are not normally distributed. This means that the customary maximum likelihood (ML) and generalized least squares (GLS) estimators are formally inappropriate for these techniques because they assume multivariate normality.

This presents several apparent difficulties in using these techniques: structural model parameter estimates, and the fit and significance statistics may be incorrect. However, based on available evidence (e.g., Anderson & Amemiya, 1985, 1986; Boomsma, 1983; Browne, 1987; Harlow, 1985; Sharma, Durvasula & Dillon, 1989; Tanaka, 1984) ML and GLS parameter estimates are robust against departures from normality (Jöreskog & Sörbom, 1989; Bollen, 1989). The results of the present study support this: the Figure 2 and 3 models were not multivariate normal (because the product indicators are not normally distributed), yet the proposed, Hayduk, and Kenny and Judd techniques reproduced the population parameters quite well using maximum likelihood and generalized least squares estimates[8].

For model fit and significance statistics, however, these estimation techniques should be used with caution (Bentler, 1989; Bollen, 1989; Hu, Bentler & Kano, 1992; Jöreskog & Sörbom, 1989).[9] Additional estimators that are less dependent on distributional assumptions should be used with these techniques to determine model fit and significance. EQS and LISREL provide asymptotic distribution free estimation[10] (Browne, 1982, 1984). EQS also provides linearized distribution free estimation (Bentler, 1983) and Robust statistics (Satorra & Bentler, 1988). For large models, fit indices (see Bollen & Long, 1993) may be appropriate (Kenny & Judd, 1984; Hayduk, 1987). This is obviously an area where additional work is needed.

Finally, mean deviation form for the indicators cannot be relaxed. The derivation of the "a" and Var(b) terms were based on this assumption. Further, mean deviation form is recommended to improve the interpretability of the linear effect coefficients in regression (see Aiken & West, 1991, and Jaccard, Turrisi & Wan, 1990).

CONCLUSION

The article has proposed an alternative to the Hayduk (1987) and Kenny and Judd (1984) techniques for estimating structural equation models with interaction or quadratic latent variables. The proposed technique is limited to indicators that are in mean deviation form and multivariate normal. In addition, the linear latent variables are assumed to be unidimensional, so that measurement model parameter estimates can be used in the structural model as constants. An iterative procedure was suggested to correct for slight differences in the measurement parameter estimates of linear latent variables between the measurement and structural models. The efficacy of the proposed technique was suggested by recovering known parameters in artificial data sets, and producing estimates for field survey data that are similar to Kenny and Judd estimates.

REFERENCES

Aiken, L. S., & West, S. G. (1991). Multiple Regression: Testing and Interpreting Interactions. Newbury Park, CA: SAGE Publications.

Aker, D. A., & Bagozzi, R. P. (1979). Unobservable variables in structural equation models with an application in industrial selling. Journal of Marketing Research, 16, 147-158.

Anderson, J. C., & Gerbing, D. W. (1982). Some methods for respecifying measurement models to obtain uni-dimensional construct measurement. Journal of Marketing Research, 19, 186-92.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. Psychological Bulletin 103, 411-23.

Anderson, T.W., & Amemiya, Y. (1985). The asymptotic normal distribution of estimators in factor analysis under general conditions (Tech. Rep. No. 12). Stanford, CA: Stanford University, Econometric Workshop, .

Bentler, P. M. (1983). Some contributions to efficient statistics for structural models: Specification and estimation of moment structures. Psychometrika 48, 493-517.

Bentler, P. M. (1989). EQS Structural Equations Program Manual. Los Angeles: BMDP Statistical Software.

Bollen, K. A. (1989). Structural Equations with Latent Variables. New York: Wiley.

Bollen, K.A., & Long, J. S. (1993). Testing Structural Equation Models. Newbury Park, CA: SAGE Publications.

Bollen, Kenneth A.(1995), AStructural Equation Models that are Nonlinear in Latent Variables: A Least Squares Estimator,≅ Sociological Methodology, 25, 223-251.

Boomsma, A. (1983). On the Robustness of LISREL (Maximum Likelihood Estimation) Against Small Sample Size and Nonnormality. Unpublished Doctoral Dissertation, University of Groningen, Groningen, The Netherlands.

Browne, M. W. (1982). Covariance structures. In D.M. Hawkins (Ed.), Topics in Applied Multivariate Analysis (pp. 72-141). Cambridge: Cambridge University Press.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology 37, 62-83.

Browne, M. W. (1987). Robustness of statistical inference in factor analysis and related models. Biometrika 74, 375-384.

Burt, R. S. (1973). Confirmatory factor-analysis structures and the theory construction process. Sociological Methods and Research 2, 131-187.

Burt, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. Sociological Methods and Research 5, 3-52.

Cohen, J., & Cohen, P. (1975). Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J., & Cohen, P. (1983). Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum.

Harlow, L. L. (1985). Behavior of Some Elliptical Theory Estimators with Nonnormal Data in a Covariance Structures Framework: A Monte Carlo Study. Unpublished Doctoral Dissertation, University of California, Los Angeles.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. Applied Psychological Measurement 9, 139-164.

Hayduk, L. A. (1987). Structural Equation Modeling with LISREL: Essentials and Advances. Baltimore, MD: Johns Hopkins Press.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? Psychological Bulletin 112, 351-362.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). Interaction Effects in Multiple Regression. Newbury Park, CA: SAGE Publications.

Jaccard, James, Robert and C. K. Wan (1995), AMeasurement Error in the Analysis of Interaction Effects Between Continuous Predictors Using Multiple Regression: Multiple Indicator and Structural Equation Approaches,≅ Psychological Bulletin, 117 (2), 348-357.

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. Biometrika 57, 239-251.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. Psychometrika 57, 409-426.

Jöreskog, K. G. (1993). Testing structural equation models. In K.A. Bollen and J.S. Long (Eds.), (pp. 294-316). Newbury Park, CA: SAGE.

Jöreskog, K. G., & Sörbom, D. (1989). LISREL 7 A Guide to the Program and Applications (2nd ed.). Chicago: SPSS, Inc.

Jöreskog, Karl G. and Fan Yang (1996), ANonlinear Structural Equation Models: The Kenny and Judd Model with Interaction Effects,≅ Advances in Structural Equation Modeling Techniques, G.A. Marcoulides, R.E. Schumacker, eds., Hillsdale, NJ: LEA.

Kenny, D., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. Psychological Bulletin 96, 201-210.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. Biometrika 57, 519-530.

McDonald, R. P. (1978). A simple comprehensive model for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology 31, 59-72.

McDonald, R. P. (1981). The dimensionality of tests and items. British Journal of Mathematical and Statistical Psychology 34, 100-117.

Neter, J., Wasserman, W. & Kunter, M. H. (1985). Applied Linear Statistical Models. Homewood, IL: Irwin.

Ping, R. A. (1993). The effects of satisfaction and structural constraints on retailer exiting, voice, loyalty, opportunism, and neglect. Journal of Retailing 69, 320-352.

Ping, R.A. (1995). A Parsimonious estimating technique for interaction and quadratic latent variables. Journal of Marketing Research, 32 (August), 336-347.

Ping, R.A. (1996), "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients," *Multivariate Behavioral Research*, 31 (1), 95-120.

Rusbult, C. E., Zembrodt, I. M., & Gunn, L. K. (1982). Exit, voice, loyalty, and neglect: responses to dissatisfaction in romantic involvement. Journal of Personal and Social Psychology 43, 1230-1242.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-squared statistics in covariance structure analysis. Proceedings of the American Statistical Association, 308-313.

Sharma, S., Durvasula, S., & Dillon, W. R. (1989). Some results on the behavior of alternative covariance structure estimation procedures in the presence of nonnormal data. Journal of Marketing Research 26, 214-221.

Tanaka, J. S. (1984). Some Results on the Estimation of Covariance Structure Models. Unpublished Doctoral Dissertation, University of California, Los Angeles.

Table 1. Artificial Data Set Population Characteristics

Quadratic Term Model:

Parameter[a]	Population Variance	Population Value
X	1.00	
ε_{x1}	0.15	
ε_{x2}	0.55	
E_Y	0.20	
λ_{x1}		1.00
λ_{x2}		0.60
$\gamma_{Y,X}$		.25
$\gamma_{Y,XX}$		-.50

Interaction Term Model:

Parameter[b]	Population Variance	Population Value
X	2.15	
Z	1.60	
$\psi_{X,Z}$	0.20	
ε_{x1}	0.36	
ε_{x2}	0.81	
ε_{z1}	0.49	
ε_{z2}	0.64	
E_Y	0.16	
λ_{x1}		1.00
λ_{x2}		0.60
λ_{z1}		1.00
λ_{z2}		0.70
$\gamma_{Y,X}$		-0.15
$\gamma_{Y,XZ}$		0.70
$\gamma_{Y,Z}$	0.35	

[a] $Y = \gamma_{Y,X}X + \gamma_{Y,XX}X^2 + E_Y$
$x_1 = \lambda_{x1}X + \varepsilon_{x1}$
$x_2 = \lambda_{x2}X + \varepsilon_{x2}$

[b] $Y = \gamma_{Y,X}X + \gamma_{Y,Z}Z + \gamma_{Y,XZ}XZ + E_Y$
$x_1 = \lambda_{x1}X + \varepsilon_{x1}$
$x_2 = \lambda_{x2}X + \varepsilon_{x2}$
$z_1 = \lambda_{z1}Z + \varepsilon_{z1}$
$z_2 = \lambda_{z2}Z + \varepsilon_{z2}$

Table 2. Artificial Data Set Sample Variance-Covariance Matrix

Quadratic Term Model:

	x_1	x_2	x_1x_1	x_1x_2	x_2x_2	y
x_1	1.272					
x_2	.605	.826				
x_1x_1	.202	.105	3.614			
x_1x_2	.208	.131	1.687	1.553		
x_2x_2	.162	.141	.860	1.175	1.731	
y	.186	.112	-1.317	-.729	-.443	.898

Interaction Term Model:

	x_1	x_2	z_1	z_2	x_1z_1	x_1z_2	x_2z_1	z_2z_2	y
x_1	2.376								
x_2	1.234	1.492							
z_1	.319	.139	2.129						
z_2	.206	.068	1.174	1.497					
x_1z_1	.106	.004	.179	.199	5.130				
x_1z_2	-.244	-.121	.174	.187	2.830	3.927			
x_2z_1	-.003	.074	.166	.095	2.516	1.326	2.885		
x_2z_2	-.125	-.161	.085	.096	1.333	1.908	1.572	2.153	
y	-.232	-.188	.595	.453	2.688	2.009	1.396	1.040	2.324

Table 3. Artificial Data Set Sample Measurement Model Parameter Estimates

Quadratic Term Model:

Parameter	Vari-ance	Value
X	1.009	
ε_{x1}	0.262	
ε_{x2}	0.463	
λ_{x1}		1.000
λ_{x2}		0.599

Equation 3 Values[a]

$a_{1,1}$ = 1.000 $Var(b_{1,1})$ = 1.198
$a_{1,2}$ = .599 $Var(b_{1,2})$ = .684
$a_{2,2}$ = .359 $Var(b_{2,2})$ = 1.101

Interaction Term Model:

Parameter	Vari-ance	Value
X	2.223	
Z	1.604	
$\psi_{X,Z}$	0.237	
ε_{x1}	0.152	
ε_{x2}	0.807	
ε_{z1}	0.524	
ε_{z2}	0.637	
λ_{x1}		1.000
λ_{x2}		0.554
λ_{z1}		1.000
λ_{z2}		0.731

Equation 3 Values[b]

$a_{1,1}$ = 1.000 $Var(b_{1,1})$ = 1.492
$a_{1,2}$ = .731 $Var(b_{1,2})$ = 1.646
$a_{2,1}$ = .554 $Var(b_{2,1})$ = 2.077
$a_{2,2}$ = .406 $Var(b_{2,2})$ = 1.644

[a] e.g., $a_{2,2} = \lambda_{x2}^2 = .599^2 = .359$,

$Var(b_{2,2}) = 4\lambda_{x2}^2 Var(X)Var(\varepsilon_{z2}) + 2Var(\varepsilon_{x2})^2$

$= 4(.599)^2(1.009)(.463)+2(.463)^2 \nabla 1.101$

[b] e.g., $a_{2,2} = \lambda_{x2}\lambda_{z2} = .554*.731 \nabla .406$,

$Var(b_{2,2}) = \lambda_{x2}^2 Var(X)Var(\varepsilon_{z2}) + \lambda_{z2}^2 Var(Z)Var(\varepsilon_{x2}) + Var(\varepsilon_{x2})Var(\varepsilon_{z2})$

$= .554^2(2.223)(.637)+.731^2(1.604)(.807)+.807(.637) \nabla 1.644$

Table 4. Structural Model Parameter Estimates

Quadratic Term Model:

Parameter	Population Variance	Population Value	Kenny & Judd Approach Variance	Kenny & Judd Approach Value	Hayduk Approach Variance	Hayduk Approach Value	Proposed Approach Variance	Proposed Approach Value
X	1.00		1.050		1.080		1.009	
ε_{x1}	0.15		0.243		0.243		0.264	
ε_{x2}	0.55		0.516		0.524		0.463	
E_Y	0.20		0.102		0.098		0.235	
λ_{x1}		1.00		1.000		1.000		1.000
λ_{x2}		0.60		0.585		0.592		0.600
$\gamma_{Y,X}$		0.25		0.275		0.274		0.290
$\gamma_{Y,XX}$		-.50		-0.573		-0.570		-0.494
MSE[a]-all parameters				.003		.004		.003
MSE[a]-γ's				.003		.003		.001

Interaction Term Model:

Parameter	Population Variance	Population Value	Kenny & Judd Approach Variance	Kenny & Judd Approach Value	Hayduk Approach Variance	Hayduk Approach Value	Proposed Approach Variance	Proposed Approach Value
X	2.15		2.277		2.368		2.223	
Z	1.60		1.643		1.605		1.604	
$\psi_{X,Z}$	`0.20		0.336		0.161		0.237	
ε_{x1}	0.36		0.122		0.144		0.327	
ε_{x2}	0.81		0.710		0.769		0.747	
ε_{z1}	0.49		0.450		0.462		0.538	
ε_{z2}	0.64		0.678		0.692		0.638	
E_Y	0.16		0.224		0.186		0.353	
λ_{x1}		1.00		1.000		1.000		1.000
λ_{x2}		0.60		0.543		0.537		0.599
λ_{z1}		1.00		1.000		1.000		1.000
λ_{z2}		0.70		0.722		0.720		0.737
$\gamma_{Y,X}$		-0.15		-0.145		-0.140		-0.132
$\gamma_{Y,XZ}$		0.70		0.695		0.780		0.666
$\gamma_{Y,Z}$	0.35		0.317		0.320		0.318	
MSE[a]-all parameters				.008		.007		.004
MSE[a]-γ's			.000		.002		.001	

Overall:

Parameter	Population Variance	Population Value	Kenny & Judd Approach Variance	Kenny & Judd Approach Value	Hayduk Approach Variance	Hayduk Approach Value	Proposed Approach Variance	Proposed Approach Value
MSE[a]-all parameters				.006		.007		.003
MSE[a]-γ's				.001		.003		.001

[a] Mean squared deviations from the population parameters.

Table 5.
Field Data Set Variance-Covariance Matrix

	s_1	s_2	s_3	s_4	a_5	a_1	a_2	a_3	a_4
s_1	.49								
s_2	.36	.53							
s_3	.41	.46	.62						
s_4	.37	.40	.44	.52					
s_5	.37	.43	.48	.43	.56				
a_1	-.28	-.34	-.35	-.29	-.31	1.00			
a_2	-.25	-.30	-.32	-.24	-.31	.71	.94		
a_3	-.26	-.33	-.38	-.32	-.34	.79	.76	.92	
a_4	-.21	-.29	-.32	-.22	-.29	.59	.61	.66	.76
l_1	-.06	-.03	.04	-.05	-.05	.03	.08	.05	.05
l_2	.00	.02	.03	.04	.02	.04	.07	.02	.03
l_3	.05	.11	.14	.11	.09	-.07	-.07	-.12	-.09
l_4	-.13	-.05	-.14	-.08	-.14	-.02	.06	.00	.05

	l_1	l_2	l_3	l_4
l_1	.70			
l_2	.54	.68		
l_3	.35	.42	.88	
l_4	.39	.41	.30	.96

Table 6. Field Data Set Measurement Model Parameter Estimates

Parameter	Variance	Parameter	Value	Parameter	Value
SAT	0.519	ε_{a3}	0.077	λ_{s5}	0.939
ALT	0.850	ε_{a4}	0.244	λ_{a1}	0.924
$\psi_{X,Z}$	-0.372	ε_{l1}	0.234	λ_{a2}	0.904
ε_{s1}	0.167	ε_{l2}	0.069	λ_{a3}	1.000
ε_{s2}	0.132	ε_{l3}	0.589	λ_{a4}	0.783
ε_{s3}	0.108	ε_{l4}	0.680	λ_{l1}	0.876
ε_{s4}	0.119	λ_{s1}	0.793	λ_{l2}	1.000
ε_{s5}	0.102	λ_{s2}	0.883	λ_{l3}	0.688
ε_{a1}	0.275	λ_{s3}	1.000	λ_{l4}	0.682
ε_{a2}	0.248	λ_{s4}	0.880		

Equation 3 Values (λ's = "a's", ε's = b's)[a]

$\lambda_{s1,s1} = .628$	$Var(\varepsilon_{s1,s1}) = .273$	$\lambda_{a3,a4} = .783$	$Var(\varepsilon_{a3,a4}) = .266$
$\lambda_{s1,s2} = .700$	$Var(\varepsilon_{s1,s2}) = .132$	$\lambda_{a4,a4} = .613$	$Var(\varepsilon_{a4,a4}) = .627$
$\lambda_{s1,s3} = .793$	$Var(\varepsilon_{s1,s3}) = .139$	$\lambda_{s1,a1} = .732$	$Var(\varepsilon_{s1,a1}) = .256$
$\lambda_{s1,s4} = .697$	$Var(\varepsilon_{s1,s4}) = .125$	$\lambda_{s1,a2} = .716$	$Var(\varepsilon_{s1,a2}) = .238$
$\lambda_{s1,s5} = .744$	$Var(\varepsilon_{s1,s5}) = .126$	$\lambda_{s1,a3} = .793$	$Var(\varepsilon_{s1,a3}) = .179$
$\lambda_{s2,s2} = .779$	$Var(\varepsilon_{s2,s2}) = .248$	$\lambda_{s1,a4} = .620$	$Var(\varepsilon_{s1,a4}) = .207$
$\lambda_{s2,s3} = .883$	$Var(\varepsilon_{s2,s3}) = .126$	$\lambda_{s2,a1} = .815$	$Var(\varepsilon_{s2,a1}) = .243$
$\lambda_{s2,s4} = .777$	$Var(\varepsilon_{s2,s4}) = .116$	$\lambda_{s2,a2} = .798$	$Var(\varepsilon_{s2,a2}) = .224$
$\lambda_{s2,s5} = .829$	$Var(\varepsilon_{s2,s5}) = .115$	$\lambda_{s2,a3} = .883$	$Var(\varepsilon_{s2,a3}) = .153$
$\lambda_{s3,s3} = 1.00$	$Var(\varepsilon_{s3,s3}) = .247$	$\lambda_{s2,a4} = .691$	$Var(\varepsilon_{s2,a4}) = .199$
$\lambda_{s3,s4} = .880$	$Var(\varepsilon_{s3,s4}) = .118$	$\lambda_{s3,a1} = .924$	$Var(\varepsilon_{s3,a1}) = .250$
$\lambda_{s3,s5} = .939$	$Var(\varepsilon_{s3,s5}) = .113$	$\lambda_{s3,a2} = .904$	$Var(\varepsilon_{s3,a2}) = .230$
$\lambda_{s4,s4} = .774$	$Var(\varepsilon_{s4,s4}) = .219$	$\lambda_{s3,a3} = 1.00$	$Var(\varepsilon_{s3,a3}) = .140$
$\lambda_{s4,s5} = .826$	$Var(\varepsilon_{s4,s5}) = .107$	$\lambda_{s3,a4} = .783$	$Var(\varepsilon_{s3,a4}) = .209$
$\lambda_{s5,s5} = .881$	$Var(\varepsilon_{s5,s5}) = .207$	$\lambda_{s4,a1} = .813$	$Var(\varepsilon_{s4,a1}) = .229$
$\lambda_{a1,a1} = .853$	$Var(\varepsilon_{a1,a1}) = .949$	$\lambda_{s4,a2} = .795$	$Var(\varepsilon_{s4,a2}) = .211$
$\lambda_{a1,a2} = .835$	$Var(\varepsilon_{a1,a2}) = .439$	$\lambda_{s4,a3} = .880$	$Var(\varepsilon_{s4,a3}) = .141$
$\lambda_{a1,a3} = .924$	$Var(\varepsilon_{a1,a3}) = .310$	$\lambda_{s4,a4} = .689$	$Var(\varepsilon_{s4,a4}) = .189$
$\lambda_{a1,a4} = .723$	$Var(\varepsilon_{a1,a4}) = .387$	$\lambda_{s5,a1} = .867$	$Var(\varepsilon_{s5,a1}) = .227$
$\lambda_{a2,a2} = .817$	$Var(\varepsilon_{a2,a2}) = .812$	$\lambda_{s5,a2} = .848$	$Var(\varepsilon_{s5,a2}) = .209$
$\lambda_{a2,a3} = .904$	$Var(\varepsilon_{a2,a3}) = .283$	$\lambda_{s5,a3} = .939$	$Var(\varepsilon_{s5,a3}) = .129$
$\lambda_{a2,a4} = .707$	$Var(\varepsilon_{a2,a4}) = .359$	$\lambda_{s5,a4} = .735$	$Var(\varepsilon_{s5,a4}) = .189$
$\lambda_{a3,a3} = 1.00$	$Var(\varepsilon_{a3,a3}) = .273$		

---

[a] Since SAT and ALT have 5 and 4 indicators respectively, SAT*SAT has 15 product indicators (=p*(p+1)/2, where p is the number of indicators), one for each unique product of the indicators of SAT, and that many sets of equation (3) "a's" and Var(b)'s (one set for each product indicator). Similarly ALT*ALT has 10 sets of "a's" and Var(b)'s, and SAT*ALT has 20 (=p*q, where p and q are the number of indicators of SAT and ALT respectively).

Table 7. Field Data Set Structural Parameter Estimates

	Variance			Variance			Variance/Value	
Parameter	Proposed	COSAN	Parameter	Proposed	COSAN	Parameter	Proposed (t-value)[a]	COSAN (t-value)[a]
SAT	0.519	0.554	$\lambda_{s4,s5}$	0.826	0.839	ε_{s3a2}	0.230	0.219
ALT	0.851	0.892	$\lambda_{s5,s5}$	0.881	0.897	ε_{s3a3}	0.140	0.152
λ_{s1}	0.793	0.741	$\lambda_{a1,a1}$	0.853	0.866	ε_{s3a4}	0.209	0.187
λ_{s2}	0.883	0.875	$\lambda_{a1,a2}$	0.835	0.821	ε_{s4a1}	0.229	0.210
λ_{s3}	1.000	1.000	$\lambda_{a1,a3}$	0.924	0.911	ε_{s4a2}	0.211	0.200
λ_{s4}	0.880	0.872	$\lambda_{a1,a4}$	0.723	0.734	ε_{s4a3}	0.141	0.149
λ_{s5}	0.940	0.931	$\lambda_{a2,a2}$	0.817	0.824	ε_{s4a4}	0.189	0.169
λ_{a1}	0.924	0.930	$\lambda_{a2,a3}$	0.904	0.906	ε_{s5a1}	0.227	0.238
λ_{a2}	0.905	0.890	$\lambda_{a2,a4}$	0.707	0.692	ε_{s5a2}	0.209	0.217
λ_{a3}	1.000	1.000	$\lambda_{a3,a3}$	1.00	1.00	ε_{s5a3}	0.129	0.109
λ_{a4}	0.783	0.774	$\lambda_{a3,a4}$	0.783	0.796	ε_{s5a4}	0.189	0.201
λ_{l1}	0.879	0.860	$\lambda_{a4,a4}$	0.613	0.600	ε_{s1s1}	0.273	0.251
λ_{l2}	1.000	1.000	$\lambda_{s1,a1}$	0.732	0.739	ε_{s1s2}	0.132	0.115
λ_{l3}	0.692	0.708	$\lambda_{s1,a2}$	0.716	0.731	ε_{s1s3}	0.139	0.147
λ_{l4}	0.686	0.671	$\lambda_{s1,a3}$	0.793	0.781	ε_{s1s4}	0.125	0.139
ε_{s1}	0.166	0.134	$\lambda_{s1,a4}$	0.620	0.610	ε_{s1s5}	0.126	0.141
ε_{s2}	0.133	0.101	$\lambda_{s2,a1}$	0.815	0.824	ε_{s2s2}	0.248	0.250
ε_{s3}	0.108	0.129	$\lambda_{s2,a2}$	0.798	0.806	ε_{s2s3}	0.126	0.114
ε_{s4}	0.119	0.101	$\lambda_{s2,a3}$	0.883	0.898	ε_{s2s4}	0.116	0.134
ε_{s5}	0.102	0.079	$\lambda_{s2,a4}$	0.691	0.678	ε_{s2s5}	0.115	0.111
ε_{a1}	0.275	0.194	$\lambda_{s3,a1}$	0.924	0.931	ε_{s3s3}	0.247	0.237
ε_{a2}	0.248	0.262	$\lambda_{s3,a2}$	0.904	0.900	ε_{s3s4}	0.118	0.136
ε_{a3}	0.077	0.124	$\lambda_{s3,a3}$	1.00	1.00	ε_{s3s5}	0.113	0.116
ε_{a4}	0.244	0.219	$\lambda_{s3,a4}$	0.783	0.772	ε_{s4s4}	0.219	0.202
ε_{l1}	0.232	0.247	$\lambda_{s4,a1}$	0.813	0.827	ε_{s4s5}	0.107	0.115
ε_{l2}	0.072	0.061	$\lambda_{s4,a2}$	0.795	0.808	ε_{s5s5}	0.207	0.217
ε_{l3}	0.587	0.569	$\lambda_{s4,a3}$	0.880	0.871	ε_{a1a1}	0.949	1.001
ε_{l4}	0.677	0.702	$\lambda_{s4,a4}$	0.689	0.694	ε_{a1a2}	0.439	0.479
$\lambda_{s1,s1}$	0.621	0.615	$\lambda_{s5,a1}$	0.867	0.853	ε_{a1a3}	0.310	0.321
$\lambda_{s1,s2}$	0.700	0.710	$\lambda_{s5,a2}$	0.848	0.840	ε_{a1a4}	0.387	0.374
$\lambda_{s1,s3}$	0.793	0.781	$\lambda_{s5,a3}$	0.939	0.930	ε_{a2a2}	0.812	0.813
$\lambda_{s1,s4}$	0.697	0.705	$\lambda_{s5,a4}$	0.735	0.748	ε_{a2a3}	0.283	0.267
$\lambda_{s1,s5}$	0.744	0.759	ε_{s1a1}	0.256	0.271	ε_{a2a4}	0.359	0.349
$\lambda_{s2,s2}$	0.779	0.769	ε_{s1a2}	0.238	0.227	ε_{a3a3}	0.273	0.237
$\lambda_{s2,s3}$	0.883	0.880	ε_{s1a3}	0.179	0.161	ε_{a3a4}	0.266	0.232
$\lambda_{s2,s4}$	0.777	0.798	ε_{s1a4}	0.207	0.216	ε_{a4a4}	0.627	0.639
$\lambda_{s2,s5}$	0.829	0.841	ε_{s2a1}	0.243	0.254	$\gamma_{LOY,SAT}$	0.031 (.3)	0.020 (.2)
$\lambda_{s3,s3}$	1.00	1.00	ε_{s2a2}	0.224	0.256	$\gamma_{LOY,ALT}$	0.097 (1.4)	0.099 (1.4)
$\lambda_{s3,s4}$	0.880	0.871	ε_{s2a3}	0.153	0.171	$\gamma_{LOY,SAT*ALT}$	0.384 (6.0)	0.379 (5.9)
$\lambda_{s3,s5}$	0.939	0.951	ε_{s2a4}	0.199	0.176	$\gamma_{LOY,ALT*ALT}$	0.127 (3.1)	0.119 (2.9)
$\lambda_{s4,s4}$	0.774	0.763	ε_{s3a1}	0.250	0.261	$\gamma_{LOY,SAT*SAT}$	0.153 (2.3)	0.161 (2.4)

[a] Approximate

Figure 1. A Covariance Structure Model.

(Click here to view Figure 1)

Figure 2. A Quadratic Model Using
The Proposed Approach.

(Click here to view Figure 2)

Figure 3. An Interaction Model Using
The Proposed Approach.

(Click here to view Figure 3)

Figure 4. A Field Survey Model Using The Proposed Approach.

(Click here to view Figure 4)

ENDNOTES

1. Since this article was written several additional approaches have been proposed, most of them refinements of the Kenny and Judd (1984) approach (described later). Bollen 1995 proposed using a two-stage least squares estimator. Jöreskog and Yang (1996) provided additional details for the Kenny and Judd approach. Jaccard and Wan (1995) suggested using a subset of the Kenny and Judd indicators. Ping (1995) suggested a single indicator approach to estimating latent variable interactions and quadratics, and Ping (1996) suggested an errors-in-variables regression approach.

2. An indicator in mean deviation form is the result of subtracting the mean of the indicator from the value of that indicator in each case. The resulting indicator has mean zero (see Aiken & West, 1991; Bollen, 1989 p. 13; Jaccard, Turrisi & Wan, 1990 p. 28; Kenny & Judd, 1984).

3. LISREL 8 provides constraint equations that can be used to implement the Kenny and Judd technique. Equation 1, for example, can be specified using two constraint equations, one for $\lambda_{x1}\lambda_{z1}$ and one for the balance of equation 1 after the $\lambda_{x1}^2\lambda_{z1}^2$Var(XZ) term.

4. The number of required constraint equations specifications in LISREL 8 is approximately equal to the number of additional variable specifications required by COSAN. Each interaction latent variable requires the specification of 2pq equations and each quadratic latent variable requires p(p+1) equations, where p and q are the number of indicators for the linear variables comprising the interaction or quadratic latent variable.

5. See the representative calculations in Table 3. An EXCEL spreadsheet is available on the web site to calculate the "a's" and Var(b)'s.

6. For each linear latent variable, a single construct measurement model (Jöreskog, 1993) was re-estimated until a target comparative fit index (Bentler, 1990) value of .99 was attained by serially deleting items that did not appear to degrade construct validity.

7. A reviewer suggested this procedure to deal with slight differences between the measurement parameter estimates from the linear-latent-variable-only measurement model and their estimates in the structural model. The effectiveness of this procedure for larger measurement parameter differences (i.e. in the first decimal place) is unknown.

8. Kenny and Judd (1984) and Hayduk (1987) reported similar results.

9. Results from recent investigations (see Jaccard & Wan, 1995; Ping, 1995) suggest that model fit and significance statistics from ML and possibly GLS estimators are robust to the addition of a few nonlinear indicators (e.g., x_1z_1) involving linear indicators (e.g. x_1 and z_1) that are normally distributed. However, the robustness of model fit and significance statistics from these estimators to the addition of many nonlinear indicators (i.e. over four) or nonlinear indicators composed of nonnormal linear indicators (typical of survey data) is unknown.

10. As Aiken and West (1991) warn, and other studies suggest (see Hu, Bentler & Kano, 1990; Jaccard & Wan 1995), -results from asymptotic distribution free estimation with less than very large sample sizes also seem to require cautious interpretation.

LATENT VARIABLE REGRESSION: A TECHNIQUE FOR ESTIMATING INTERACTION AND QUADRATIC COEFFICIENTS

(An earlier, but revised, version of Ping 1996, *Multivariate Behavioral Research*, 31 (1), 95-120)
(Updated July, 2001)

ABSTRACT

The article proposes a technique to estimate regression coefficients for interaction and quadratic latent variables that combines regression analysis with the measurement model portion of structural equation analysis (e.g., analysis involving EQS, LISREL, or AMOS). The measurement model provides parameter estimates that can be combined to correct the variance-covariance matrix used in regression, as Heise (1986) and others recommend. The proposed technique will provide coefficient estimates for regression models involving existing measures, or new measures for which a priori error estimates are not available.

For survey data, regression is the predominant data analysis technique in several social science literatures. It is widely used in other social science literatures, presumably because it is easily understood and available in popular statistics packages for microcomputers (e.g., SAS, SPSS, etc.).

Researchers in the social sciences have called for the inclusion of interaction and quadratic variables (e.g., xz and zz respectively in

$$y = b_0 + b_1x + b_2z + b_3xz \tag{1}$$

and

$$y = b'_0 + b'_1x + b'_2z + b_4zz) \tag{2}$$

in analyses of survey data with regression (Aiken & West, 1991; Cohen & Cohen, 1983; Jaccard, Turrisi & Wan, 1990). However, regression is known to produce coefficient estimates that are biased and inefficient for variables measured with error such as unobserved or latent variables (Bohrnstedt & Carter, 1971; Busemeyer & Jones, 1983; Cochran, 1968; Fuller, 1987; Gleser, Carroll & Gallo, 1987).

Recently, Heise (1986) proposed a regression approach to estimating interaction and quadratic coefficients for variables measured with error in survey data. The approach requires the researcher to have advance estimates of the errors in the measures used in the study. This limits the applicability of the technique to studies involving established measures with previously reported reliabilities.

The balance of the paper discusses a technique for estimating quadratic and interaction latent variables in survey data using regression, that avoids the requirement for a-priori estimates of reliability. The proposed technique uses estimates of measurement error provided by the measurement model step in structural equation analysis (see Anderson & Gerbing, 1988). These error estimates are used to correct the variance-covariance matrix used in regression. After a brief discussion of the techniques available for estimating interactions and quadratics in survey data using regression, the proposed technique is developed. The efficacy of the technique is then suggested by recovering known coefficients in synthetic data sets, and the proposed technique is applied to a field survey data set.

INTERACTION AND QUADRATIC ESTIMATION

Survey researchers who include latent variable interaction and quadratic terms in their models use two types of techniques: those that produce estimates of the coefficients for the interaction and quadratic terms in equations 1 and 2, and those that do not. Techniques that produce coefficient estimates for interactions and quadratics include regression and structural equation analysis (e.g., analysis involving AMOS, EQS or LISREL). Techniques that do not produce coefficient estimates for interactions and quadratics in survey data include ANOVA (see Maxwell & Delaney, 1993), subgroup analysis (see Jaccard, Turissi & Wan, 1990; Jöreskog, 1971), dummy variable regression (see Dillon & Goldstein, 1984), and the Chow test (Chow, 1960). These techniques that do not produce coefficient estimates for interactions and quadratics are also limited to testing for a single interaction or quadratic variable. The balance of the article will concentrate on techniques that produce coefficient estimates.

STRUCTURAL EQUATION TECHNIQUES

Kenny and Judd (1984), among others (e.g., Bollen 1995, Hayduk, 1987; Jaccrd & Wan 1995; Jöreskog & Yang 1996; Ping 1995, 1996; Wong & Long, 1987), have proposed an approach to specifying interaction and quadratic latent variables using structural equation analysis. In structural equation analysis the measured variables (indicators) are assumed to be linear functions of their unobserved (latent) variable. For an indicator x this relationship is specified as

$$Var(x) = \lambda_X^2 Var(\xi_X) + Var(\varepsilon_X) , \tag{3}$$

where Var(a) is the variance of a, λ_X is the (factor) loading of x on the latent variable ξ_X, ε_X is the error in measuring x, and ξ_X and ε_X are independent (Jöreskog, 1973; Keesling, 1972; Wiley, 1973). The Kenny and Judd approach involves specifying the indicators of a latent variable interaction, ξ_{XZ} for example, by using all possible products of the indicators of the latent variables ξ_X and ξ_Z. In particular for ξ_X and ξ_Z with indicators x_1, x_2, z_1, and z_2, respectively, ξ_{XZ} would have the indicators x_1z_1, x_1z_2, x_2z_1, and x_2z_2.[1] Under certain conditions the variance of these indicator products is given by

$$\begin{aligned} Var(x_iz_j) &= Var((\lambda_{Xi}\xi_X + \varepsilon_{Xi})(\lambda_{Zj}\xi_Z + \varepsilon_{Zj})) \\ &= \lambda_{Xi}\lambda_{Zj}Var(\xi_{XZ}) + \lambda_{Xi}Var(\xi_X)\varepsilon_{Zj} + \lambda_{Zj}Var(\xi_Z)\varepsilon_{Xi} + Var(\varepsilon_{Xi})Var(\varepsilon_{Zj}) . \end{aligned} \tag{4}$$

Specification of these indicator products is tedious however. The Kenny and Judd approach requires the specification of four dummy (non-linear) variables, one for $\lambda_{Xi}\lambda_{Zj}$, and one for each of the last three terms of equation 4. Hence a total of sixteen dummy variables would be required for the four product indicators of ξ_{XZ}. For models with several interactions or quadratics, or several indicators per variable, these dummy variables can overwhelm the model. For example a model with two linear latent variables (e.g., ξ_X and ξ_Z) having six indicators each, one interaction, and two quadratic variables requires the specification of three hundred and seventy-two additional dummy variables.[2] Aiken and West (1991) noted that this approach has been difficult for researchers to implement.

Regression Techniques

Perhaps for these reasons regression continues to be a popular alternative to structural equation analysis for estimating interactions and quadratic effects among latent variables. Two regression approaches are available. Researchers can ignore measurement error, sum the items to form a single measure of each concept, and form arithmetic products of these summed measures to create interaction and quadratic variables. For example, the interaction variable corresponding to the summed variables $X = x_1 + x_2$ and $Z = z_1 + z_2$ would have the form $XZ = (x_1 + x_2)(z_1 + z_2)$. Quadratic variables would be similarly constructed. These variables can then be added to the regression model. However, this approach has been criticized for its biased and inefficient coefficient estimates for both the summed variables and the interactions and quadratics (Bohrnstedt & Carter, 1971; Busemeyer & Jones, 1983; Cochran, 1968; Fuller, 1987; Gleser, Carroll & Gallo, 1987).

Warren, White and Fuller (1974) and Heise (1986) proposed an alternative approach for regression and variables measured with error, that involved correcting the regression moment matrix using appropriate error variance terms. Consider the variable x given by $x = X_T + e_x$, where x is the observed score, X_T is the true score, and e_x is error. The variance of the true score X_T could be estimated using the variance of the observed score x and the variance of the error e_x, and the regression variance-covariance matrix could be corrected for this error. Specifically,

$$Var(x) = Var(X_T) + Var(e_x) \tag{5}$$

where X_T and e_x are assumed to be uncorrelated. Hence

$$\begin{aligned} Var(e_x) &= Var(x) - Var(X_T) \\ &= Var(x) - \rho_{XX}Var(X) \\ &= Var(x)(1 - \rho_{XX}) \end{aligned}$$

where

$$\rho_{XX} = Var(X_T)/Var(x) = \text{the (a-priori) reliability of x.}$$

For an interaction variable xz,

$$\begin{aligned} Var(X_TZ_T) &= Var[(x-e_x)(z-e_z)] \\ &= Var(xz - xe_z - ze_x + e_xe_z) \\ &= Var(xz) - Var(z)Var(e_x) - Var(x)Var(e_z) + Var(e_x)Var(e_z) \\ &= Var(xz) - Var(z)Var(x)(1 - \rho_{XX}) - Var(x)Var(z)(1 - \rho_{ZZ}) \\ &\quad + Var(x)(1 - \rho_{XX})Var(z)(1 - \rho_{ZZ}) \end{aligned}$$

(see Bohrnstedt & Goldberger, 1969) could replace Var(xz) in the regression variance-covariance matrix, assuming x and z have a multivariate normal distribution with zero mean.

While useful, this approach is limited to situations where the a-priori errors or reliabilities of x and z are known. Nevertheless, a similar correction approach could be taken using a structural equation analysis. This approach is developed next.

A PROPOSED ESTIMATION TECHNIQUE

Structural equation modeling packages such as AMOS, EQS and LISREL can provide a straightforward estimate of λ_X, $Var(\xi_X)$, and $Var(\varepsilon_X)$ in equation 3 using the so-called measurement model (see Anderson & Gerbing, 1982; 1988) (see also Byrne, 1989). This measurement model is intended to gauge the adequacy of the assignment of indicators to latent variables, and in the process it produces estimates of the parameters in equation 3 (i.e., $Var(\varepsilon_X)$, λ_X and $Var(\xi_X)$).

For variables formed as sums of indicators such as $X = x_1 + x_2$, equation 3 becomes

$$\begin{aligned} Var(X) &= Var(x_1 + x_2) \\ &= Var[(\lambda_{X1}\xi_X + \varepsilon_{X1}) + (\lambda_{X2}\xi_X + \varepsilon_{X2})] \\ &= Var((\lambda_{X1} + \lambda_{X2})\xi_X) + Var(\varepsilon_{X1}) + Var(\varepsilon_{X2}) \\ &= Var(\Lambda_X\xi_X) + \theta_X, \\ &= \Lambda_X^2 Var(\xi_X) + \theta_X, \end{aligned} \tag{6}$$

where $\Lambda_X = \lambda_{X1} + \lambda_{X2}$, $\theta_X = Var(\varepsilon_{X1}) + Var(\varepsilon_{X2})$, x_1 and x_2 are independent of ε_{X1} and ε_{X2}, ε_{X1} and ε_{X2} are independent of each other, and x_1 and x_2 are multivariate normal with zero means. Since estimates of $\Lambda_X = \lambda_{X1} + \lambda_{X2}$, $\theta_X = Var(\varepsilon_{X1}) + Var(\varepsilon_{X2})$, and $Var(\xi_X)$ are available in a measurement model for ξ_X, they can be used to correct Var(X), and provide a consistent estimate of $Var(\xi_X)$. Rearranging equation 6, Var(X) in a regression variance-covariance matrix could be replaced by an estimate of $Var(\xi_X)$

$$Var(\xi_X) = (Var(X) - \theta_X)/\Lambda_X^2 \ . \tag{7}$$

CORRECTION EQUATIONS

The balance of the regression variance-covariance matrix could be corrected in a similar manner, using combinations of the uncorrected variance-covariance matrix entries and measurement model estimates. For example, consider the following regression model,

$$Y = b_0 + b_1X + b_2Z + b_3V + b_4W + b_5XX + b_6ZZ + b_7XZ + b_8VW + b_9VZ \ , \tag{8}$$

where X, Z, V, W and Y are sums of indicators and of the form

$$Q = q_1 + q_2 \ ,$$

XX, ZZ, XZ, VW, and VZ are of the form

$$PQ = (p_1+p_2)(q_1+q_2) \ ,$$

and p_1, p_2, q_1 and q_2 are indicators meeting the equation 6 conditions.

The correction for the diagonal term Var(X) in the variance-covariance matrix for equation 8 is given by equation 7 and the corrections for Var(Z), Var(V), Var(W), and Var(Y) are similar to those shown in equation 7. Under the assumptions for equation 6, the corrections for the other terms in the variance-covariance matrix for equation 8 can be determined. For example, the corrections for the equation 8 variance-covariance matrix diagonal terms composed of interactions such as Var(XZ), Var(VW) and Var(VZ) are given by

$$Var(\xi_X\xi_Z) = (Var(XZ) - Var(\xi_X)\Lambda_X^2\theta_Z - Var(\xi_Z)\Lambda_Z^2\theta_X - \theta_X\theta_Z)/\Lambda_X^2\Lambda_Z^2 \tag{9}$$

where $\Lambda_X = \lambda_{x1} + \lambda_{x2}$, $\Lambda_Z = \lambda_{z1} + \lambda_{z2}$, $\theta_X = Var(\varepsilon_{x1}) + Var(\varepsilon_{x2})$, and $\theta_Z = Var(\varepsilon_{z1}) + Var(\varepsilon_{z2})$ (see Appendix A).

The correction for quadratics such as Var(XX) and Var(ZZ) is similar

$$Var(\xi_X\xi_X) = (Var(XX) - 4Var(\xi_X)\Lambda_X^2\theta_X - 2\theta_X^2)/\Lambda_X^4 \ . \tag{10}$$

Off diagonal terms composed of linear variables such as Cov(X,Z), are given by

$$Cov(\xi_X,\xi_Z) = Cov(X,Z)/\Lambda_X\Lambda_Z \ . \tag{11}$$

Other combinations of linear terms such as Cov(X,Y), Cov(X,V), Cov(X,W), Cov(Z,Y), Cov(Z,V), Cov(Z,W) and Cov(V,W) are similar.

Mixed off-diagonal terms composed of linear and interaction or quadratic variables are also corrected. For example, terms such as Cov(V,XZ) are corrected as follows

$$Cov(\xi_V,\xi_X\xi_Z) = Cov(V,XZ)/\Lambda_V\Lambda_X\Lambda_Z \ . \tag{12}$$

The other combinations of linear and interaction or quadratic terms such as Cov(X,XX), Cov(Z,XZ), etc. are similar.

For the correction of off diagonal combinations of interactions and quadratics there are several cases: a covariance term composed of two quadratics, two interactions, or an interaction and a quadratic. The covariance of a quadratic and an interaction with a common linear term such as Cov(XX,XZ) is corrected with

$$\mathrm{Cov}(\xi_X\xi_X,\xi_X\xi_Z) = (\mathrm{Cov}(XX,XZ) - 2\mathrm{Cov}(\xi_X,\xi_Z)\Lambda_X\Lambda_Z\theta_X)/\Lambda_X^3\Lambda_Z\ . \tag{13}$$

Other combinations of interactions or quadratics are corrected similarly. For example, a covariance with two interactions with a common linear term such as Cov(VW,VZ) is corrected with

$$\mathrm{Cov}(\xi_V\xi_W,\xi_V\xi_Z) = (\mathrm{Cov}(VW,VZ) - \mathrm{Cov}(\xi_W,\xi_Z)\Lambda_W\Lambda_Z\theta_V)/\Lambda_V^2\Lambda_W\Lambda_Z\ . \tag{14}$$

A covariance with a combination of interactions or quadratics with no common terms such as Cov(XZ,VW) or Cov(XX,ZZ) is corrected with

$$\mathrm{Cov}(\xi_X\xi_Z,\xi_V\xi_W) = \mathrm{Cov}(XZ,VW)/\Lambda_X\Lambda_Z\Lambda_V\Lambda_W\ . \tag{15}$$

Equations 7 and 9-15 generalize to an arbitrary number of indicators for X, Z, V, W and Y (see Appendix A).

SYNTHETIC DATA EXAMPLES

To gauge the efficacy of this technique it was used to recover known coefficients in synthetic data sets. Using a normal random number generator, data sets composed of 100 replications of samples of 50, 100, and 150 cases were created. Each replication was generated using the Table 1 population characteristics for x_1, x_2, z_1, z_2, t_1, t_2 and y in the equation

$$Y = \beta_{Y,X}X + \beta_{Y,Z}Z + \beta_{Y,T}T + \beta_{Y,XZ}XZ + \beta_{Y,XX}XX + \zeta_Y \tag{16}$$

(see Appendix B for details).To gauge the effects of varying the simulation conditions the process was repeated for two additional levels of latent variable reliability (see Table 1).

The equation 16 model was estimated for each replication by creating the variables X (= $[x_1+x_2]/2$), Z (= $[z_1+z_2]/2$), T (= $[t_1+t_2]/2$), XX (= X*X), XZ (= X*Z), and Y (= y, a single indicator) in each case.[3] Then the linear-terms-only measurement model associated with equation 16 (i.e., involving only X, Z, T and Y-- see Figure 1) was estimated using EQS and maximum likelihood estimates. Specifically the λ's, θ_ε's and the variances and covariances of the latent variables ξ_X, ξ_Z, ξ_T, and ξ_Y were estimated. This produced estimates of the λ's, θ_ε's and Var(ξ)'s for use in equations 7 and 9-13. After using equation 7 and 9-13 to correct the equation 16 variance-covariance matrix, the coefficients in equation 16 were estimated using this corrected matrix and ordinary least squares regression. The results are shown in Table 2.

To obtain a basis for comparison uncorrected regression estimates were also generated for each replication. These estimates used the uncorrected equation 16 variance-covariance matrix and ordinary least squares. The results are also shown in Table 2, and will be discussed later.

To illustrate the use of the proposed technique a field survey data analysis involving interaction and quadratic latent variables is presented.

A FIELD SURVEY EXAMPLE

As part of a study of reactions to changes in overall inter-group satisfaction with an exchange relationship (e.g., a firm selling to another firm) data were gathered using multiple Likert items measuring overall satisfaction (SAT) of the subject group with the partner group, the attractiveness of the best alternative group (ALT), and the opportunism (OPP) (self interest seeking with guile, Williamson, 1975) (which can plausibly be viewed as a form of instrumental aggression) committed by the subject group on the partner group (see Ping, 1993).

Since the purpose is to illustrate the use of the proposed estimation technique the study will simply be summarized. SAT was measured using a seven-item scale, ALT used six items, and OPP was measured with eight items. The anticipated relationships among the study concepts were

$$OPP = b_1SAT + b_2ALT + \zeta. \tag{17}$$

Opportunism was expected to be negatively associated with satisfaction and positively associated with the attractiveness of the best alternative.

Because alternative attractiveness was a new measure developed for this study, an a-priori estimate of its reliability was not available, and the approaches suggested by Heise (1986) or Feucht (1989) were not feasible. In addition, a structural equation analysis using the Kenny and Judd (1984) approach produced an unacceptably low model-to-data fit, that was improved only by deleting items in the measures. Because these item deletions appeared to compromise the content validity of the established measures, the proposed technique was used.[4]

Two hundred eighty dyads were analyzed, and the resulting cases were used to produce the uncorrected variance-covariance matrix shown in Table 3. The uncorrected regression results shown in Table 4 were the result of testing the indicators for non normality, averaging the indicators of each concept, zero centering each indicator for the linear independent variables (i.e., s_1, s_2, ... , s_7, a_1, ... a_6), and entering the interaction and quadratic variables into the regression jointly (see Lubinski & Humphreys, 1990). Zero centering the indicator s_1, for example, is accomplished by subtracting the sample mean of s_1 from the value of s_1 in each case. The result is a mean of zero for s_1 which meets the equation 6 requirement for an indicator mean of zero.

The equation 17 regression results shown in Table 4 suggested that opportunism was weakly associated with satisfaction, and that alternatives had the larger association. Because these results were difficult to explain, interaction and quadratic terms were added to equation 17:

$$OPP = b_1SAT + b_2ALT + b_3SAT^2 + b_4SATALT + b_5ALT^2 + \zeta. \tag{17a}$$

The equation 17a uncorrected regression results shown in Table 4 suggested that opportunism was associated with both antecedents, but that the opportunism association with satisfaction may be contingent on the level of alternatives.

To obtain unbiased estimates of these associations, the equation 17 measurement model was estimated using LISREL 7 (see Figure 2). The resulting estimates for the indicator λ's and θ's, and the variances and covariances of SAT, ALT and OPP are shown in Table 5. These estimates and equations 7 and 9-13 were used to correct the Table 3 variance-covariance matrix, and produced the corrected matrix shown in Table 6. The corrected regression results shown in Table 7 suggested that the association between opportunism and satisfaction was contingent on the level of alternatives. In particular when alternatives were few (i.e., ALT was less than zero, its average), the negative association between satisfaction and opportunism was stronger (the coefficient of SAT was given by -.158+.213ALT), than when there were many alternatives (i.e., when ALT was above average or positive). These results will be discussed next.

DISCUSSION

When compared to the uncorrected equation 17a results, the equation 17 regression produced a simple but misleading view of the relationships between opportunism and its antecedents. Adding the uncorrected interaction and quadratic terms (equation 17a) clarified these relationships somewhat, but the coefficient estimates for both the linear and nonlinear variables were biased. The corrected estimates of the equation 17a coefficients, however, suggested that the relationship between opportunism and satisfaction was contingent on the level of alternatives.

In this example, removing the regression coefficient bias did not produce dramatically different estimates. However, the corrected Table 7 estimates could have been larger, smaller, or of different signs than the uncorrected Table 4 estimates. Bohrnstedt and Carter (1971) demonstrated that the extent and direction of regression coefficient bias, when there are multiple independent variables measured with error, depends not only on the reliabilities of the independent variables, but also on the population correlations among the independent variables. As a result, the uncorrected Table 4 coefficients could have born little resemblance to the population coefficients and their estimates given by the corrected Table 7 coefficients.

The proposed technique appeared to produce less biased coefficient estimates than uncorrected regression in the synthetic data sets (see Table 2). The average coefficient estimates (Sample Coefficient Average in Table 2) were within a few points of the population values for all three sample sizes and reliabilities, and as a result the biases, the differences between the sample average values and the population values, were small. However, the variances of the coefficient estimates and the average squared deviations of the estimates around the population values (MSE in Table 2) were larger than for uncorrected regression. Hence the proposed estimation technique appeared to reduce coefficient estimate bias at the expense of increased coefficient estimate variability.

Several assumptions made in the derivations of the proposed corrections were required and one was not. The assumption that the indicator error terms were mutually independent, and the assumption that the indicators have zero means, were not absolutely necessary-- they merely simplified the correction equations for this exposition. Relaxing the assumption of mutually independent error terms adds error covariance terms to equations 11-15, and error variance terms to equations 7, 9 and 10.

Relaxing the assumption of zero indicator means also changes the form of the corrections. Without this assumption the covariance of each combination of a linear terms and an interaction or quadratic term (e.g., Cov(SAT,SATSAT)) must also be corrected for the non zero means of the measures. However, transforming independent variables so that they have zero means (i.e., zero centering) is recommended when investigating interaction and quadratic terms in the presence of their linear component terms (Cronbach, 1987; Jaccard, Turrisi & Wan, 1990; Aiken & West, 1991).

The assumptions of indicator normality however cannot be relaxed. This assumption is required of latent variables in structural equation analysis, and enables the use of the measurement model estimates. However, measurement model estimates from Maximum Likelihood and Generalized Least Squares appear to be robust to departures form normality

(Anderson & Amemiya, 198; 1986; Boomsma, 1983; Browne, 1987; Harlow, 1985; Sharma, Durvasula & Dillon, 1989; Tanaka, 1984).

(At the time the paper was written) Perhaps the most serious limitation of the proposed technique is the lack of a formally appropriate significance testing statistic. Since the distribution of the standard errors estimated with the proposed technique is unknown, the p-values associated with these coefficient estimates should not be trusted. Ping (2001) proposed scaling the uncorrected coefficient standard errors using the ratio of the uncorrected standard error of the estimate to the corrected standard error of the estimate (this paper is on the web site).

Finally, modification of the variance-covariance matrix is not a feature of any popular regression software package, and the estimation of a measurement model adds procedural complexity. A spreadsheet package (LOTUS) was used in the simulations and the field survey example. The simulation variance-covariance matrix was created as an output file using SPSS and imported to the spreadsheet. Next the measurement model estimates were manually keyed into the spreadsheet and used to compute the corrected matrix.[5] This matrix was output to a file (using the spreadsheet's ability to print to a file in ASCII) that was then read by the final regression program. Measurement model estimation is described in some detail in Bentler (1989, p. 26), Byrne (1989), and Jöreskog and Sörbom (1989, p. 96). An EXCEL version of this spreadsheet is available on the web site.

SUMMARY

A technique for estimating regression effects involving interaction and quadratic latent variables has been proposed that is conceptually simple and appears to produce unbiased estimates that have a larger variance than uncorrected regression estimates. The technique appears to be suitable for studies in which all interactions and quadratics are probed, structural equation estimation is undesirable, or new measures are involved. The proposed technique involves several steps: I) zero center the indicators for the linear independent variables, create summed variables from the indicators (e.g., $X = x_1+x_2+...+x_n$), then create the interaction and quadratic terms (e.g., $XZ = X*Z$), ii) create the regression variance-covariance matrix, iii) estimate the measurement model parameters associated with the indicators of the summed variables (e.g., x_1, x_2, ..., x_n), iv) correct the variance-covariance matrix using equations 7 and 9-13 and the measurement model parameter estimates, and v) estimate the regression effects using the corrected regression variance-covariance matrix resulting from step iv (the coefficient standard errors should be computed using Ping 2001).

REFERENCES

Aiken, L.S. & West, S.G. (1991). Multiple Regression: Testing and Interpreting Interactions. Newbury Park, CA: SAGE Publications.

Anderson, J.C. & Gerbing, D.W. (1982). Some methods for respecifying measurement models to obtain unidimensional construct measurement. Journal of Marketing Research, 16, 186-92.

Anderson, J.C. & Gerbing, D.W. (1988). Structural equation modeling in practice: a review and recommended two-step approach. Psychological Bulletin, 103, 411-23.

Anderson, T.W. & Amemiya, Y. (1985). The asymptotic normal distribution of estimators in factor analysis under general conditions (Tech. Rep. No. 12). Stanford, CA: Stanford University, Econometric Workshop.

Bentler, P.M. (1989). EQS Structural Equations Program Manual. Los Angeles: BMDP Statistical Software.

Bentler, P.M. (1990). Comparative fit indices in structural models. Psychological Bulletin, 107, 238-246.

Bohrnstedt, G.W. & Carter, T.M. (1971). Robustness in regression analysis. In H.L. Costner (Ed.), Sociological Methodology (pp. 118-146). San Francisco: Jossey-Bass.

Bohrnstedt, G. & Goldberger, A. (1969). On the exact covariance of products of random variables. Journal of the American Statistical Association, 64, 1439-1442.

Bollen, Kenneth A.(1995), AStructural equation models that are nonlinear in latent variables: a least squares estimator,≃ Sociological Methodology, 25, 223-251.

Boomsma A. (1983). On the robustness of Lisrel (maximum likelihood estimation) against small sample size and nonnormality. Unpublished Doctoral Dissertation, University of Groningen, Groningen, The Netherlands.

Browne, M.W. (1987). Robustness of statistical inference in factor analysis and related models. Biometrika, 74, 375-84.

Busemeyer, J.R. & Jones, L.E. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. Psychological Bulletin, 93, 549-62.

Byrne, B.M. (1989). A Primer of LISREL. New York: Springer Verlag.

Cochran, W.G. (1968). Errors in measurement in statistics. Technometrics, 10, 637-666.

Chow, G.C. (1960). Tests of equality between sets of coefficients in two linear regressions. Econometrika, 28, 591-605.

Cohen, J. & Cohen, P. (1983). Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum.

Cronbach, L. (1987). Statistical tests for moderator variables: flaws in analysis recently proposed. Psychological Bulletin, 102, 414-417.

Dillon, W.R. & Goldstein, M. (1984). Multivariate Analysis Methods and Applications. New York: Wiley.

Efron, B. (1981). Nonparametric estimates of standard error: the jacknife, the bootstrap, and other resampling methods. Biometrika, 68, 589-599.

Feucht, T.E (1989). Estimating multiplicative regression terms in the presence of measurement error. Sociological Methods and Research, 17, 257-282.

Fuller, W.A. (1987). Measurement Error Models. New York: Wiley.

Gleser, L.J., Carroll, R.J. & Gallo, P.P. (1987). The Limiting distribution of least squares in an errors-in-variables linear regression model. The Annals of Statistics, 15, 220-233.

Harlow, L.L. (1985). Behavior of Some Elliptical Theory Estimators with Nonnormal Data in a Covariance Structures Framework: A Monte Carlo Study. Unpublished Doctoral Dissertation, University of California, Los Angeles.

Hayduk, L.A. (1987). Structural Equation Modeling with LISREL: Essential and Advances. Baltimore, MD: Johns Hopkins Press.

Heise, D.R. (1986). Estimating nonlinear models correcting for measurement error. Sociological Methods and Research, 14, 447-472.

Jaccard, J., & Wan, C.K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: multiple indicator and structural equation approaches. Psychological Bulletin, 117, 348-357.

Jaccard, J., Turrisi, R. & Wan, C.K. (1990). Interaction Effects in Multiple Regression. Newbury Park, CA: SAGE Publications.

Jaccard, James, Robert and C. K. Wan (1995), AMeasurement Error in the Analysis of Interaction Effects Between Continuous Predictors Using Multiple Regression: Multiple Indicator and Structural Equation Approaches,≃ Psychological Bulletin, 117 (2), 348-357.

Jöreskog, K.G. (1971). Simultaneous factor analysis in several populations. Psychometrika, 57, 409-26.

Jöreskog, K.G. (1973). A general method for estimating a linear structural equation system. In A.S. Goldberger and O.D. Duncan (Eds.), Structural Equation Models in the Social Sciences (pp. 85-112). New York: Academic Press.

Jöreskog, K.G., Sörbom, D. (1989). LISREL 7 A Guide to the Program and Applications, 2nd ed., Chicago: SPSS Inc.

Jöreskog, Karl G. and Fan Yang (1996), ANonlinear Structural Equation Models: The Kenny and Judd Model with Interaction Effects,≅ Advances in Structural Equation Modeling Techniques, G.A. Marcoulides, R.E. Schumacker, eds., Hillsdale, NJ: LEA.

Keesling, J.W. (1972). Maximum Likelihood Approaches to Causal Analysis. Unpublished Doctoral Dissertation, University of Chicago.

Kenny, D. & Judd, C.M. (1984). Estimating the nonlinear and interactive effects of latent variables. Psychological Bulletin, 96, 201-210.

Lubinski, D. & Humphreys, L.G. (1990). Assessing spurious Amoderator effects≅: illustrated substantively with the hypothesized (Asynergistic≅) relation between spatial and mathematical ability. Psychological Bulletin, 107, 385-393.

Maxwell, S.E. & Delaney, H.D. (1993). Bivariate median splits and spurious statistical significance. Psychological Bulletin, 113, 181-190.

Netemeyer, R.G., Johnson, M.W. & Burton, S. (1990). Analysis of role conflict and role ambiguity in a structural equations framework. Journal of Applied Psychology, 75, 148-157.

Ping, R.A. (1993). The effects of satisfaction and structural constraints on retailer exiting, voice, loyalty, opportunism, and neglect. Journal of Retailing, 69, 320-352.

______ (1995). A Parsimonious estimating technique for interaction and quadratic latent variables. Journal of Marketing Research.

______ (1996). Interaction and quadratic effect estimation: a two step technique using structural equation analysis. Psychological Bulletin.

______ (2001), "A Suggested Standard Error for Interaction Coefficients in Latent Variable Regression," *2001 Academy of Marketing Science Conference Proceedings*, Miami: Academy of Marketing Science.

Sharma, S., Durvasula, S. & Dillon, W.R. (1989). Some results on the behavior of alternative covariance structure estimation procedures in the presence of nonnormal data. Journal of Marketing Research, 26, 214-221.

Tanaka, J.S. (1984). Some results on the estimation of covariance structure models. Unpublished Doctoral Dissertation, University of California, Los Angeles.

Warren, R.D., White, J.K. & Fuller, W.A. (1974). Errors in variables analysis of managerial role performance. Journal of the American Statistical Association, 69, 886-893.

Wiley, D.E. (1973). The Identification problem for structural equation models with unmeasured variables. In A.S. Goldberger and O.D. Duncan (Eds.), Structural Equation Models in the Social Sciences (pp. 69-83). New York: Academic Press.

Williamson, O.E. (1975). Markets and Hierarchies: Analysis and Antitrust Implications. New York: Free Press.

Wong, S.K. & Long, J.S. (1987). Reparameterizing Nonlinear Constraints in Models with Latent Variables. (Technical Report). Pullman, WA: Washington State University.

APPENDIX A-- Correction Details for an Arbitrary Variance-Covariance Matrix

The following presents the proposed corrections for the elements of an arbitrary variance-covariance matrix in more detail.

The correction for the variance of the linear term $X = x_1 + x_2$, where $x_1 = \lambda_{X1}\xi_X + \varepsilon_{X1}$, $x_2 = \lambda_{X2}\xi_X + \varepsilon_{X2}$, x_1 and x_2 are independent of ε_{X1} and ε_{X2}, ε_{X1} and ε_{X2} are independent of each other, and x_1 and x_2 are multivariate normal with zero means, is as follows:

$$\begin{aligned} Var(X) &= Var(x_1 + x_2) \\ &= Var[(\lambda_{X1}\xi_X + \varepsilon_{X1}) + (\lambda_{X2}\xi_X + \varepsilon_{X2})] \\ &= Var((\lambda_{X1} + \lambda_{X2})\xi_X) + Var(\varepsilon_{X1}) + Var(\varepsilon_{X2}) \\ &= Var(\Lambda_X\xi_X) + \theta_X, \\ &= \Lambda_X^2 Var(\xi_X) + \theta_X, \end{aligned}$$

where Var(a) is the variance of a, $\Lambda_X = \lambda_{X1} + \lambda_{X2}$, $\theta_X = Var(\varepsilon_{X1}) + Var(\varepsilon_{X2})$. As a result, a correction for Var(X) is given by

$$Var(\xi_X) = (Var(X) - \theta_X)/\Lambda_X^2.$$

The correction for Cov(X,Z) is

$$\begin{aligned} Cov(X,Z) &= Cov(x_1 + x_2, z_1 + z_2) \\ &= Cov(x_1,z_1) + Cov(x_1,z_2) + Cov(x_2,z_1) + Cov(x_2,z_2) \\ &= Cov(\lambda_{X1}\xi_X+\varepsilon_{X1},\lambda_{Z1}\xi_Z+\varepsilon_{Z1}) + Cov(\lambda_{X1}\xi_X+\varepsilon_{X1},\lambda_{Z2}\xi_Z+\varepsilon_{Z2}) \\ &\quad + Cov(\lambda_{X2}\xi_X+\varepsilon_{X2},\lambda_{Z1}\xi_Z+\varepsilon_{Z1}) + Cov(\lambda_{X2}\xi_X+\varepsilon_{X2},\lambda_{Z2}\xi_Z+\varepsilon_{Z2}) \\ &= Cov(\xi_X,\xi_Z)(\lambda_{X1}\lambda_{Z1} + \lambda_{X1}\lambda_{Z2} + \lambda_{X2}\lambda_{Z1} + \lambda_{X2}\lambda_{Z2}) \\ &= Cov(\xi_X,\xi_Z)(\lambda_{X1} + \lambda_{X2})(\lambda_{Z1} + \lambda_{Z2}) \\ &= Cov(\xi_X,\xi_Z)\Lambda_X\Lambda_Z, \end{aligned}$$

and a correction for Cov(X,Z) is given by

$$Cov(\xi_X,\xi_Z) = Cov(X,Z)/\Lambda_X\Lambda_Z,$$

where $\Lambda_X = \lambda_{X1} + \lambda_{X2} + ... + \lambda_{zm}$, and $\Lambda_Z = \lambda_{Z1} + \lambda_{Z2} + ... + \lambda_{Zn}$.

Off-diagonal terms composed of linear and non linear variables such as Cov(V,XZ) are corrected as follows:

$$Cov(V,WX) = Cov(\Lambda_V\xi_V + E_V,[\Lambda_W\xi_W + E_W][\Lambda_X\xi_X + E_X]),$$

where $E_V = \varepsilon_{V1} + \varepsilon_{V2} + ... + \varepsilon_{Vp}$, $E_W = \varepsilon_{W1} + \varepsilon_{W2} + ... + \varepsilon_{Wq}$, and $E_X = \varepsilon_{X1} + \varepsilon_{X2} + ... + \varepsilon_{Xm}$. Hence

$$Cov(V,WX) = Cov(\xi_V,\xi_W\xi_X)\Lambda_V\Lambda_W\Lambda_X,$$

and

$$Cov(\xi_V,\xi_W\xi_X) = Cov(V,WX)/\Lambda_V\Lambda_W\Lambda_X,$$

where $\Lambda_V = \lambda_{V1} + \lambda_{V2} + ... + \lambda_{Vp}$, $\Lambda_W = \lambda_{W1} + \lambda_{W2} + ... + \lambda_{Wq}$, and $\Lambda_X = \lambda_{X1} + \lambda_{X2} + ... + \lambda_{Xm}$.

The covariance of an interaction is given by

$$Cov(VW,XZ) = Cov(V,X)Cov(W,Z) + Cov(V,Z)Cov(W,X), \tag{i}$$

(see Kenny and Judd 1984), and

$$\begin{aligned} Cov(VW,XZ) &= Cov(\xi_V,\xi_X)\Lambda_V\Lambda_X Cov(\xi_W,\xi_Z)\Lambda_W\Lambda_Z \\ &\quad + Cov(\xi_V,\xi_Z)\Lambda_V\Lambda_Z Cov(\xi_W,\xi_X)\Lambda_W\Lambda_X \\ &= Cov(\xi_V\xi_W,\xi_X\xi_Z)\Lambda_V\Lambda_W\Lambda_X\Lambda_Z . \end{aligned}$$

A correction for Cov(VW,XZ) is therefore given by

$$Cov(\xi_V\xi_W,\xi_X\xi_Z) = Cov(VW,XZ)/\Lambda_V\Lambda_W\Lambda_X\Lambda_Z, \tag{ii}$$

where $\Lambda_V = \lambda_{V1} + \lambda_{V2} + ... + \lambda_{Vp}$, $\Lambda_W = \lambda_{W1} + \lambda_{W2} + ... + \lambda_{Wq}$, $\Lambda_X = \lambda_{X1} + \lambda_{X2} + ... + \lambda_{Xm}$, and $\Lambda_Z = \lambda_{Z1} + \lambda_{Z2} + ... + \lambda_{Zn}$.

By equality ii the correction for the covariance of two quadratics such as Cov(XX,ZZ) is

$$Cov(\xi_X\xi_X,\xi_Z\xi_Z) = Cov(XX,ZZ)/\Lambda_X^2\Lambda_Z^2,$$

where $\Lambda_X = \lambda_{X1} + \lambda_{X2} + ... + \lambda_{Xm}$, and $\Lambda_Z = \lambda_{Z1} + \lambda_{Z2} + ... + \lambda_{Zn}$.

For the variance of an interaction

$$\begin{aligned} Var(XZ) &= Cov(XZ,XZ) \\ &= Var(X)Var(Z) + Cov(X,Z)^2, \end{aligned}$$

using equality i. Hence

$$\begin{aligned} Var(XZ) &= [\Lambda_X^2 Var(\xi_X) + \theta_X][\Lambda_Z^2 Var(\xi_Z) + \theta_Z] + [Cov(\xi_X,\xi_Z)\Lambda_X\Lambda_Z]^2 \\ &= Cov(\xi_X\xi_Z,\xi_X\xi_Z)\Lambda_X\Lambda_W\Lambda_X\Lambda_Z + Var(\xi_X)\Lambda_X^2\theta_Z \\ &\quad + Var(\xi_Z)\Lambda_Z^2\theta_X + \theta_X\theta_Z, \end{aligned}$$

and

$$Var(\xi_X\xi_Z) = (Var(XZ) - Var(\xi_X)\Lambda_X{}^2\theta_Z - Var(\xi_Z)\Lambda_Z{}^2\theta_X - \theta_X\theta_Z)/\Lambda_X{}^2\Lambda_Z{}^2 ,$$

where $\Lambda_X = \lambda_{X1} + \lambda_{X2} +...+ \lambda_{Xm}$, $\Lambda_Z = \lambda_{Z1} + \lambda_{Z2} +...+ \lambda_{Zn}$, $\theta_X = Var(\varepsilon_{X1}) + Var(\varepsilon_{X2}) +...+ Var(\varepsilon_{Xm})$, $\theta_Z = Var(\varepsilon_{Z1}) + Var(\varepsilon_{Z2}) +...+ Var(\varepsilon_{Zn})$.

The correction for a quadratic such as Var(XX) is similar:

$$\begin{aligned} Var(XX) &= 2Var(X)^2 \\ &= 2[\Lambda_X{}^2Var(\xi_X) + \theta_X]^2 \\ &= Var(\xi_X\xi_X)\Lambda_X{}^4 + 4Var(\xi_X)\Lambda_X{}^2\theta_X + \theta_X{}^2 , \end{aligned}$$

and

$$Var(\xi_X\xi_X) = (Var(XX) - 4Var(\xi_X)\Lambda_X{}^2\theta_X - 2\theta_X{}^2)/\Lambda_X{}^4 ,$$

where $\Lambda_X = \lambda_{X1} + \lambda_{X2} +...+ \lambda_{Xm}$ and $\theta_X = Var(\varepsilon_{X1}) + Var(\varepsilon_{X2}) +...+ Var(\varepsilon_{Xm})$.

For the covariance of a quadratic and an interaction that has common linear terms such as Cov(XX,XZ),

$$\begin{aligned} Cov(XX,XZ) &= 2Var(X)Cov(X,Z) \\ &= 2[\Lambda_X{}^2Var(\xi_X) + \theta_X]Cov(\xi_X,\xi_Z)\Lambda_X\Lambda_Z \\ &= Cov(\xi_X\xi_X,\xi_X\xi_Z)\Lambda_X{}^2\Lambda_X\Lambda_Z + 2Cov(\xi_X,\xi_Z)\Lambda_X\Lambda_Z\theta_X , \end{aligned}$$

and

$$Cov(\xi_X\xi_X,\xi_X\xi_Z) = (Cov(XX,XZ) - 2Cov(\xi_X,\xi_Z)\Lambda_X\Lambda_Z\theta_X)/\Lambda_X{}^3\Lambda_Z ,$$

where $\Lambda_X = \lambda_{X1} + \lambda_{X2} +...+ \lambda_{Xm}$, $\Lambda_Z = \lambda_{Z1} + \lambda_{Z2} +...+ \lambda_{Zn}$, and $\theta_X = Var(\varepsilon_{X1}) + Var(\varepsilon_{X2}) +...+ Var(\varepsilon_{Xm})$.

For a combination of interactions with common linear terms such as Cov(VW,VZ)

$$\begin{aligned} Cov(VW,VZ) &= Var(V)Cov(W,Z) + Cov(V,Z)Cov(W,V) \\ &= [\Lambda_V{}^2Var(\xi_V) + \theta_V]Cov(\xi_W,\xi_Z)\Lambda_W\Lambda_Z \\ &\qquad + Cov(\xi_V,\xi_Z)\Lambda_V\Lambda_ZCov(\xi_W,\xi_Z)\Lambda_W\Lambda_V \\ &= Cov(\xi_V\xi_W,\xi_V\xi_Z)\Lambda_V{}^2\Lambda_W\Lambda_Z + Cov(\xi_W,\xi_Z)\Lambda_W\Lambda_Z\theta_V , \end{aligned}$$

and

$$Cov(\xi_V\xi_W,\xi_V\xi_Z) = (Cov(VW,VZ) - Cov(\xi_W,\xi_Z)\Lambda_W\Lambda_Z\theta_V)/\Lambda_V{}^2\Lambda_W\Lambda_Z ,$$

where $\Lambda_V = \lambda_{V1} + \lambda_{V2} +...+ \lambda_{Vp}$, $\Lambda_W = \lambda_{W1} + \lambda_{W2} +...+ \lambda_{Wq}$, $\Lambda_Z = \lambda_{Z1} + \lambda_{Z2} +...+ \lambda_{Zn}$, and $\theta_V = Var(\varepsilon_{V1}) + Var(\varepsilon_{V2}) +...+ Var(\varepsilon_{Vp})$.

By induction a correction for Var(X) is given by $Var(\xi_X) = (Var(X) - \theta_X)/\Lambda_X{}^2$, where

$$\Lambda_X = \lambda_{X1} + \lambda_{X2} +...+ \lambda_{Xm},$$
$$\theta_X = Var(\varepsilon_{X1}) + Var(\varepsilon_{X2}) +...+ Var(\varepsilon_{Xm}),$$

and m is the number of indicators of ξ_X. The other corrections are similarly generalized. For example $Cov(\xi_V\xi_W,\xi_V\xi_Z) = (Cov(VW,VZ) - Cov(\xi_W,\xi_Z)\Lambda_W\Lambda_Z\theta_V)/\Lambda_V{}^2\Lambda_W\Lambda_Z$ where

$$V = v_1 + v_2 +...+ v_m,$$
$$W = w_1 + w_2 +...+ w_n,$$
$$Z = z_1 + z_2 +...+ z_p,$$
$$\Lambda_V = \lambda_{V1} + \lambda_{V2} +...+ \lambda_{Vm},$$
$$\Lambda_W = \lambda_{W1} + \lambda_{W2} +...+ \lambda_{Wn},$$
$$\Lambda_Z = \lambda_{Z1} + \lambda_{Z2} +...+ \lambda_{Zp},$$
$$\theta_V = Var(\varepsilon_{V1}) + Var(\varepsilon_{V2}) +...+ Var(\varepsilon_{Vm}),$$
$$\theta_W = Var(\varepsilon_{W1}) + Var(\varepsilon_{W2}) +...+ Var(\varepsilon_{Wn}),$$
$$\theta_Z = Var(\varepsilon_{Z1}) + Var(\varepsilon_{Z2}) +...+ Var(\varepsilon_{Zp}),$$

and m, n, and p are the number if indicators for V, W and Z respectively.

APPENDIX B-- Synthetic Data Set Creation Details

The data for equation 16 was generated as follows. Let **M** be an n by 1 vector of random normal variates with mean 0 and variance 1, where n is the number of cases. The n by 3 matrix **P** with columns that were the population values for the n by 1 vectors **X**, **Z**, and **T** were determined by **P** = **M**(1 1 1)**C**', where (1 1 1) is a 1 by 3 unit vector and **C** is a lower triangular matrix such that

$$\mathbf{CC}' = \begin{bmatrix} V_X & & \\ r(V_Z V_X)^2 & V_Z & \\ r(V_T V_X)^2 & r(V_T V_Z)^2 & V_T \end{bmatrix}$$

where V_* is the variance of $_*$, and r is the correlation between **X**, **Z** and **T**. The n by 4 matrices of observed values **x**, **z** and **t** for the population vectors **X**, **Z** and **T**, respectively, were given by **x** = (.6**P**(1 0 0) + **N**($0,\theta_{\varepsilon x}$))(1 1 1 1) , **z** = (.6**P**(0 1 0) + **N**($0,\theta_{\varepsilon z}$))(1 1 1 1) and **t** = (.6**P**(0 0 1) + **N**($0,\theta_{\varepsilon t}$))(1 1 1 1), where (1 0 0), (0 1 0), (0 0 1), and (1 1 1 1) are vectors of 1's, and the N($0,\theta_{\varepsilon *}$)'s are n by 1 vectors of random normal variates with mean 0 and variance $\theta_{\varepsilon *}$. The values for the n by 1 vector for dependent variable **Y** was determined by

$$\mathbf{Y} = b_X\mathbf{X} + b_Z\mathbf{Z} + b_T\mathbf{T} + b_{XX}\mathbf{XX} + b_{XZ}\mathbf{XZ} + \boldsymbol{\zeta}_{\mathbf{Y}} ,$$

where the b_*'s are the scalar effects of * on **Y**, and $\boldsymbol{\zeta}_{\mathbf{Y}}$ is an n by 1 vector of random normal variates with mean 0 and variance equal to .16 .

TABLE 1
Synthetic Data Sets Population Characteristics

		Population			
	Coefficient	Variance[b]			
Parameter[a]		All	Mixed	Low	High
X		2.15			
Z		1.60			
T		1.00			
Corr(X,Z)	0.60				
Corr(X,T)	0.60				
Corr(T,Z)	0.60				
ε_{x1}			0.36	1.15	0.19
ε_{x2}			0.81	1.15	0.19
ε_{z1}			0.49	0.85	0.14
ε_{z2}			0.64	0.85	0.14
ε_{t1}			0.15	0.53	0.09
ε_{t2}			0.55	0.53	0.09
ε_{y}		0			
ζ_{Y}		0.16			
λ_{x1}	1.00				
λ_{x2}	0.60				
λ_{z1}	1.00				
λ_{z2}	0.60				
λ_{t1}	1.00				
λ_{t2}	0.60				
λ_{y}	1.00				
$\beta_{Y,X}$	-0.15				
$\beta_{Y,Z}$	0.35				
$\beta_{Y,T}$	0.25				
$\beta_{Y,XX}$	-0.50				
$\beta_{Y,XZ}$	0.70				

[a] $Y = \beta_{Y,X}X + \beta_{Y,Z}Z + \beta_{Y,T}T + \beta_{Y,XX}XX + \beta_{Y,XZ}XZ + \zeta_Y$
$Var(x_1) = \lambda_{x1}^2 Var(X) + Var(\varepsilon_{x1})$
$Var(x_2) = \lambda_{x2}^2 Var(X) + Var(\varepsilon_{x2})$
$Var(z_1) = \lambda_{z1}^2 Var(Z) + Var(\varepsilon_{z1})$
$Var(z_2) = \lambda_{z2}^2 Var(Z) + Var(\varepsilon_{z2})$
$Var(t_1) = \lambda_{t1}^2 Var(T) + Var(\varepsilon_{t1})$
$Var(t_2) = \lambda_{t2}^2 Var(T) + Var(\varepsilon_{t2})$

[b] Mixed Variance: $\rho_X=.75$, $\rho_Z=.69$, $\rho_T=.69$.
Low Variance: ρ for X, Z and T = .6 .
High Variance: ρ for X. Z and T = .9 .

TABLE 2
Synthetic Data Set Regression Results

Coefficient	Population Value	Sample Size	Sample Coefficient Average	Bias	Variance	MSE
Uncorrected Regressions:						
$b_{Y,X}$	-0.15	n=50	0.081	0.231	0.000	0.067
		100	0.072	0.222	0.000	0.057
		150	0.068	0.218	0.000	0.049
$b_{Y,Z}$	0.35	50	0.199	-0.151	0.133	0.046
		100	0.201	-0.149	0.109	0.033
		150	0.202	-0.148	0.099	0.028
$b_{Y,T}$	0.25	50	0.307	0.057	0.005	0.018
		100	0.306	0.056	0.002	0.012
		150	0.304	0.054	0.002	0.010
$b_{Y,XX}$	-0.50	50	-0.340	0.155	0.000	0.033
		100	-0.350	0.150	0.000	0.024
		150	-0.350	0.147	0.000	0.019
$b_{Y,XZ}$	0.70	50	0.544	-0.156	0.138	0.047
		100	0.547	-0.153	0.115	0.035
		150	0.549	-0.151	0.101	0.028
Corrected Regressions:						
$b_{Y,X}$	-0.15	n=50	-0.174	-0.024	0.146	0.128
		100	-0.169	-0.019	0.119	0.106
		150	-0.168	-0.018	0.042	0.035
$b_{Y,Z}$	0.35	50	0.390	0.040	0.040	0.057
		100	0.379	0.029	0.042	0.055
		150	0.373	0.023	0.045	0.055
$b_{Y,T}$	0.25	50	0.218	-0.032	0.053	0.039
		100	0.228	-0.022	0.043	0.034
		150	0.236	-0.014	0.036	0.031
$b_{Y,XX}$	-0.50	50	-0.438	0.062	0.148	0.200
		100	-0.468	0.032	0.166	0.193
		150	-0.480	0.020	0.172	0.189
$b_{Y,XZ}$	0.70	50	0.785	0.085	0.147	0.220
		100	0.746	0.046	0.166	0.206
		150	0.736	0.036	0.158	0.188

TABLE 2 (Con't.)
Synthetic Data Set Results

Coefficient	Population Value	Reliability	Sample Coefficient Average	Bias	Variance	MSE
Uncorrected Regressions:						
$b_{Y,X}$	-0.15	ρ=High	-0.040	0.101	0.016	0.052
		Mixed	0.047	0.197	0.002	0.058
		Low	0.223	0.373	0.016	0.061
$b_{Y,Z}$	0.35	High	0.300	-0.050	0.049	0.029
		Mixed	0.104	-0.246	0.216	0.048
		Low	0.197	-0.153	0.103	0.028
$b_{Y,T}$	0.25	High	0.264	0.014	0.008	0.011
		Mixed	0.339	0.089	0.000	0.014
		Low	0.312	0.062	0.003	0.013
$b_{Y,XX}$	-0.50	High	-0.460	0.032	0.018	0.028
		Mixed	-0.350	0.149	0.000	0.024
		Low	-0.230	0.267	0.015	0.021
$b_{Y,XZ}$	0.70	High	0.670	-0.030	0.059	0.045
		Mixed	0.546	-0.154	0.120	0.037
		Low	0.431	-0.269	0.193	0.029
Corrected Regressions:						
$b_{Y,X}$	-0.15	ρ=High	-0.156	-0.006	0.013	0.012
		Mixed	-0.184	-0.034	0.062	0.046
		Low	-0.182	-0.032	0.237	0.207
$b_{Y,Z}$	0.35	High	0.351	0.001	0.014	0.014
		Mixed	0.399	0.049	0.020	0.036
		Low	0.397	0.047	0.084	0.114
$b_{Y,T}$	0.25	High	0.241	-0.009	0.011	0.009
		Mixed	0.222	-0.028	0.037	0.027
		Low	0.221	-0.029	0.080	0.064
$b_{Y,XX}$	-0.50	High	-0.493	0.007	0.071	0.075
		Mixed	-0.474	0.026	0.105	0.122
		Low	-0.419	0.081	0.289	0.383
$b_{Y,XZ}$	0.70	High	0.725	0.025	0.036	0.046
		Mixed	0.746	0.046	0.118	0.152
		Low	0.796	0.096	0.299	0.413

TABLE 3
Field Data Variance-Covariance Matrix

	SAT	ALT	OPP	SATSAT	SATALT	ALTALT
SAT	0.408					
ALT	-0.290	0.645				
OPP	-0.098	0.141	0.359			
SATSAT	-0.216	0.221	-0.001	0.546		
SATALT	0.221	-0.225	-0.007	-0.454	0.545	
ALTALT	-0.225	0.215	0.065	0.367	-0.524	0.928

TABLE 4
Uncorrected Field Data Regression Results

Dependent Variable	Independent Variable	Coefficient b	Beta	p-value	F-value and (p)	R^2 and (p) of difference
OPP	SAT	-.123	-.131	.09	11.92 (.00)	.09
	ALT	.163	.219	.00		
OPP	SAT	-.181	-.193	.01		
	ALT	.189	.254	.00		
	SATSAT	-.069	-.058	.47		
	SATALT	.195	.240	.10	7.14 (.00)	.14 (.006)
	ALTALT	.120	.193	.04		

TABLE 5
Field Data Measurement Model Results

Parameter	Variance	Parameter	Value
ε_{s1}	0.300	λ_{s1}	0.790
ε_{s2}	0.156	λ_{s2}	0.799
ε_{s3}	0.216	λ_{s3}	0.736
ε_{s4}	0.133	λ_{s4}	0.877
ε_{s5}	0.101	λ_{s5}	1.000
ε_{s6}	0.131	λ_{s6}	0.860
ε_{s7}	0.107	λ_{s7}	0.929
ε_{a1}	0.368	λ_{a1}	0.816
ε_{a2}	0.267	λ_{a2}	0.948
ε_{a3}	0.251	λ_{a3}	0.921
ε_{a4}	0.111	λ_{a4}	1.000
ε_{a5}	0.224	λ_{a5}	0.814
ε_{a6}	0.343	λ_{a6}	0.644
ε_{o1}	0.762	λ_{o1}	0.553
ε_{o2}	0.639	λ_{o2}	0.711
ε_{o3}	0.510	λ_{o3}	0.837
ε_{o4}	0.173	λ_{o4}	0.946
ε_{o5}	0.191	λ_{o5}	0.890
ε_{o6}	0.282	λ_{o6}	1.000
ε_{o7}	0.250	λ_{o7}	0.890
ε_{o8}	0.357	λ_{o8}	0.958

PHI

	SAT	ALT	OPP
SAT	.525		
ALT	-.381	.815	
OPP	-.119	.184	.425

TABLE 6
Field Data Corrected Variance-Covariance Matrix

	SAT	ALT	OPP	SATSAT	SATALT	ALTALT
SAT	0.525					
ALT	-0.381	0.815				
OPP	-0.119	0.184	0.425			
SATSAT	-0.345	0.351	-0.002	0.949		
SATALT	0.351	-0.357	-0.012	-0.819	0.953	
ALTALT	-0.357	0.341	0.104	0.681	-0.927	1.516

TABLE 7
Field Data Corrected Regression Results

Dependent Variable	Independent Variable	Coefficient b	Coefficient Beta	p-value (Approx.)	F-Value and (p) (Approx.)	R^2 and (p) of difference (Approx.)
OPP	SAT	-.158	-.176	.03		
	ALT	.210	.291	.00	12.94 (.00)	.10
	SATSAT	-.053	-.079	.54		
	SATALT	.213	.319	.05	8.47 (.00)	.16 (.001)
	ALTALT	.138	.261	.01		

Figure 1- Equation 16 Measurement Model

(Click here to view Figure 1)

Figure 2- Equation 17 Measurement Model

(Click here to view Figure 2)

ENDNOTES

1. Ping (1995) has proposed the use of a single indicator for ξ_{XZ}, $x{:}z = (x_1 + x_2)(z_1 + z_2)$.

2. A LISREL 8 implementation of the Kenny and Judd (1984) technique requires the specification of two dummy variables for equation 4, one for $\lambda_{xi}\lambda_{zj}$ and another for the balance of equation 4 (see Jaccard & Wan, 1995). LISREL 8 then creates the remaining Kenny and Judd dummy variables using partial derivatives. While the specification work with LISREL 8 is reduced, the number of dummy variables produced is the same, and the lengthy execution times, convergence problems and improper solutions potential produced by the resulting large information matrix is unchanged. Ping (1996) has proposed a two-step approach that uses fixed (not estimated) values for the loadings and errors of the indicators of ξ_{XZ} for example, and reduces the size of the resulting information matrix (see Jaccard and Wan, 1995 for an alternative approach that uses a subset of the indicators, and Ping 1995 for a single indicator approach).

3. The indicators x_1, x_2, z_1, z_2, t_1, and t_2 were generated to meet the equation 6 conditions. In particular, the indicators were generated to have zero means.

4. The measurement model for SAT, ALT, OPP, SATSAT, SATALT, and ALTALT was specified using the Kenny and Judd (1984) approach and the full measures. The indicators for the interaction and quadratic latent variables were specified using all unique products of the indicators for SAT and ALT. The Kenny and Judd equations for the loadings and the error terms for these indicator products were coded using COSAN (a subprocedure of the SAS procedure CALIS) (LISREL 8 could also be used), the measurement model was estimated, but the resulting model-to-data fit was unacceptable (e.g., the Comparative Fit Index (Bentler, 1990) was .6). Model fit was subsequently improved by deleting items from each measure (see Anderson & Gerbing, 1982; and Netemeyer, Johnson & Burton, 1990). However, the measures for SAT and OPP were well established, and item deletion appeared to substantially reduce their content validity. As a result the structural equation analysis approach was abandoned.

5. EQS will create an ASCII file containing the measurement model parameter estimates that also can be imported to the spreadsheet.

A SUGGESTED STANDARD ERROR FOR INTERACTION COEFFICIENTS IN LATENT VARIABLE REGRESSION

(An earlier, but revised, version of Ping 2001, *Proceedings of the Academy of Marketing Research 2001 Annual Conference*, Miami: Academy of Marketing Science.)
(Revised September, 2001)

ABSTRACT

Latent variable regression, a path coefficient estimation technique for interactions and quadratics in unobserved or latent variables, uses an error-attenuated covariance matrix (e.g., available from SAS, SPSS, etc.) which is then adjusted for measurement error, as input to ordinary least squares regression. However, it was proposed without a standard error for the path coefficients. This research suggests an approximate standard error for interaction coefficients in latent variable regression, and provides an example of its use.

In experiments with categorical independent variables (e.g., experiments analyzed with ANOVA), interactions, e.g. XZ in

1) $Y = b_0 + b_1X + b_2Z + b_3XZ + b_4XX + \zeta_Y$,

and quadratics (e.g., XX in Equation 1) are investigated to help interpret significant main effects (e.g., b_1 and b_2, the X-Y and Z-Y effects respectively in Equation 1). However, interactions and quadratics are seldom investigated in theoretical model validation studies using survey data, even when theory suggests their existence, possibly because they have been difficult to detect (see for example Podsakoff, Todor, Grover and Huber 1984; also see McClelland and Judd 1993). Until recently they may also have been difficult to specify with latent variables (Aiken and West 1991).

Nevertheless significant latent variable interactions have been reported (e.g., Baumgartner and Bagozzi 1995; Lusch and Brown 1996; Osterhuis (1997); Ping 1994, 1999; Singh 1998), and authors have called for the investigation of these variables in survey research (Aiken and West 1991; Blalock 1965; Cohen 1968; Cohen and Cohen 1975, 1983; Freidrich 1982; Howard 1989; Jaccard, Turrisi and Wan 1990; Kenny 1985). Their argument is that failing to consider interactions and quadratics in the population model increases the risk of false negative research findings (Type II error) because a interaction (or a quadratic) can mask a significant conditional effect, and misleading positive research findings because the effect could be conditional (see Ping 1996c for details).

There has been considerable progress recently in estimating interactions in survey data using regression (Aiken and West 1991; Jaccard, Turissi and Wan 1990; Ping 1996a) and using structural equation analysis (Bollen 1995; Hayduk 1987; Jaccard and Wan 1995; Jöreskog and Sörbom 1996; Kenny and Judd 1984; Ping, 1995, 1996b; Wong and Long 1987), and in interpreting these results (see Aiken and West 1991; Jaccard, Turissi and Wan 1990, Denters and Van Puijenbroek 1989).

In particular, Ping (1996a) suggested adjusting the covariance matrix used in ordinary least squares (OLS) regression to estimate interactions involving unobserved or latent variables with multiple indicators (see Heise 1986 for another covariance matrix adjustment approach to the errors-in-variables problem that involves OLS regression). Using simulated data, his results suggested the proposed technique, which he termed latent variable regression (LVR), performed adequately by producing unbiased and consistent regression coefficients. However, the suggested technique included no standard error for the LVR coefficients (i.e., b_1 through b_4, in Equation 1). This renders LVR useless for theoretical model testing, the major use of errors-in-variables analysis in the social sciences.

This research suggests an approximate standard error for the linear (i.e., b_1 and b_2 in Equation 1) and interaction (i.e., b_3 in Equation 1) regression coefficients produced by LVR. We will not argue for the use of LVR or the proposed standard error (however see p. 7). We will simply show that the proposed standard error performs adequately, and that point out that with it LVR could be used in theoretical model testing.

The paper begins with a brief review of LVR. Then it investigates the suggested standard error using simulated data sets. It concludes with a pedagogical example using LVR and the suggested standard error.

Latent Variable Regression

Ping (1996a) proposed using an attenuated covariance matrix (e.g., available from SAS, SPSS, etc.) that has been adjusted for measurement error as input to ordinary least squares regression in order to detect latent variable interactions and quadratics. Loadings and measurement errors from a structural equation analysis measurement model are used to provide the adjustments for this covariance matrix. For example, for unidimensional latent variables *X*, *Z*, *XZ*, and *Y* meeting the Kenny and Judd (1984) normality assumptions (i.e., indicators are multivariate normal with zero means and independent of their measurement errors, and measurement errors are independent of each other), a measurement model for *X*, *Z* and *Y* could be estimated, and the adjusted or disattenuated variance, Var(*X*), associated with X = $x_1+x_2+...+x_n$ could be estimated using

$$\mathrm{Var}(X) = [\mathrm{Var}(\mathrm{X})-\theta_X]/\Lambda_X^2,$$

where Var(X) is the attenuated variance of X (available from SAS, SPSS, etc.), θ_x is the sum of the measurement errors of x_i (= Var[ε_{x1}] +...+ Var[ε_{xn}]), and Λ_X is the sum of the loadings of x_i on *X* (= λ_{x1} +...+ λ_{xn}). The variances of *Z* and *Y* are similar.

The adjusted covariance of *X* and *Z* could be estimated from

$$\mathrm{Cov}(X,Z) = \mathrm{Cov}(\mathrm{X,Z})/\Lambda_X\Lambda_Z,$$

where Cov(X,Z) is the attenuated covariance of X and Z (=z_1+...+z_m). The covariances of *Y* with *X*, and *Z* are similar.

The adjusted covariance of *XZ* with *Y* could be estimated with

$$\mathrm{Cov}(XZ,Y) = \mathrm{Cov}(\mathrm{XZ,Y})/\Lambda_X\Lambda_Z\Lambda_Y.$$

The covariances of *XZ* with *X* and *Z* are similar.

Finally, the adjusted variance of *XZ* could be estimated using

$$\mathrm{Var}(XZ) = (\mathrm{Var(XZ)} - \Lambda_X^2\mathrm{Var}(X)\theta_Z - \Lambda_Z^2\mathrm{Var}(Z)\theta_X - \theta_X\theta_Z)/\Lambda_X^2\Lambda_Z^2.$$

A Suggested Latent Variable Regression Coefficient Standard Error

The coefficient standard errors (SE=s) (i.e., the SE=s of b_1, b_1, b_3, and b_4 in Equation 1) produced by ordinary least squares (OLS) regression using an error-adjusted covariance matrix are incorrect because they assume variables measured without error (e.g., Warren, White and Fuller 1974). A common approach to this problem (e.g., in instrumental variables and two-stage least squares-- see Hanushek and Jackson 1977) is to adjust the SE from unadjusted OLS regression by changes in its standard error, RMSE (= $[\Sigma[y_i - \hat{y}_i]^2]^2$, where y_i and $\hat{y}_i$. are observed and estimated y=s respectively). Thus an adjusted SE for LVR coefficients would involve the SE from unadjusted OLS regression, and a ratio of the standard error from unadjusted OLS regression and the standard error from LVR. Specifically, it is well known that SE=s produced by unadjusted OLS regression (i.e., an SE produced using covariances unadjusted for measurement error) are attenuated by measurement error and OLS regression RMSE is reduced by accounting for measurement error (e.g., Warren, White and Fuller 1974, see Myers 1986 for additional citations). Thus an adjustment for the SE from unadjusted OLS regression would be the ratio of the RMSE produced by unadjusted OLS regression to that produced by LVR, or

$$SE_{LVR} = SE_U{*}RMSE_U/RMSE_{LVR}\ ,$$

where SE_{LVR} is the suggested LVR standard error, SE_U is the SE produced by unadjusted OLS regression (i.e., the SE produced using covariances unadjusted for measurement error), $RMSE_U$ is the standard error produced by unadjusted OLS regression, and $RMSE_{LVR}$ is the standard error produced by LVR.

Table 2 shows the performance of this suggested standard error in detecting known population structural coefficients (i.e., b=s in $Y = b_0 + b_1X + b_2Z + b_3XZ + b_4W + \zeta_Y$) under several combinations of reliability, structural coefficient size, and sample size. These results include the performance of structural equation analysis (i.e., LISREL 8) for reference, and were produced using the Table 3 population parameters and the simulation procedure described in Appendix A.

Overall, the suggested standard error performed adequately as a coefficient standard error for LVR. Table 2 shows the bias (i.e., the ratio of the average of the standard errors to the root mean square difference of the coefficients from their population values) of the suggested standard errors and the LISREL 8 for comparison. While the downward bias of the suggested standard error was considerable in some cases, it was similar to the downward bias experienced by LISREL 8 and the LISREL 8 biases reported by Jaccard and Wan (1995), and thus it was judged to be acceptable. Further, the number of significant coefficients produced by the suggested standard error was similar to LISREL 8 over the simulation conditions, except for the low reliability-small sample condition, where the power of the test for the suggested standard error (i.e., the number of true positive interpretations) was slightly higher than LISREL 8.

An Example

In business-to-business relationships, relationship neglect (allowing the relationship to deteriorate) (*NEG*) has been argued to be associated with overall satisfaction with that relationship (*SAT*), the attractiveness of alternative relationships (*ALT*), investment in the relationship (*INV*), and the cost to switch relationships (*SCT*) (see Ping 1993). To illustrate the use of the adjusted coefficient standard error, we will add a satisfaction-alternative attractiveness interaction (*SxA*) to the variables argued to be associated with neglect, and estimate the structural equation

2) $NEG = b_0 + b_1 SAT + b_2 ALT + b_3 INV + b_4 SCT + b_5 SxA + \zeta_{NEG}$.

Because we this is a pedagogical example, we will omit the details of the study (see Ping 1993 for a test of this model without an interaction). The unadjusted covariance matrix for these variables, along with the results from a measurement model of *SAT*, *ALT*, *INV*, *SCT*, and *NEG* are shown in Table 4. The spreadsheet calculations used to adjust the attenuated covariance matrix, plus the adjusted covariance matrix are also shown in Table 4. The resulting coefficient estimates for Equation 2 are shown in Table 1.

To produce the Table 1 estimates for LVR, first we verified that *SAT*, *ALT*, *INV*, *SCT*, and *NEG* were unidimensional using single construct (structural equation) measurement models (see Jöreskog 1993). Then we mean centered the indicators of *SAT*, *ALT*, *INV*, *SCT*, and *NEG* (s, a, i, sc, and n respectively in Table 4) by subtracting the mean of an indicator from its case values using SPSS. Mean centering produces indicators and constructs with zero means, and is required for LVR (and interaction estimation with structural equation analysis without an intercept term in Equation 2-- see Jöreskog and Yang 1996). Next, the mean centered indicators for each construct were summed for each case to form regression variables SAT, ALT, INV, SCT, and NEG, and these new variables were added to each case, again using SPSS. The interaction SxA (= SAT times ALT) was also added to each case using SPSS.

Then the unadjusted covariance matrix of SAT, ALT, INV, SCT, NEG, and SxA was produced using SPSS, and a measurement model for *SAT*, *ALT*, *INV*, *SCT*, and *NEG* was estimated using LISREL 8 and maximum likelihood. Measurement model parameters for *SxA* are not necessary because the covariance matrix entries involving an interaction are adjusted using the unadjusted interaction and the >liner-terms-only= measurement model (e.g., involving *SAT*, *ALT*, *INV*, *SCT*, and *NEG*-- see Equation 2). Next, the measurement model loadings and measurement errors were used to adjust the covariance matrix of SAT, ALT, INV, SCT, NEG, and SxA to produce an adjusted covariance matrix for *SAT*, *ALT*, *INV*, *SCT*, *NEG*, and *SxA* using an Excel spreadsheet available on the web site.

This adjusted covariance matrix for *SAT*, *ALT*, *INV*, *SCT*, *NEG*, and *SxA* was input to SPSS to produce *i*) LVR coefficient estimates and *ii*) an LVR standard error ($RMSE_{LVR}$). Next the unadjusted covariance matrix was input to SPSS to estimate *iii*) unadjusted coefficient standard errors (SE_U) and *iv*) an unadjusted OLS regression standard error ($RMSE_U$).

Then adjusted coefficient standard errors were calculated by multiplying the unadjusted coefficient standard errors (SE_U-- result [*iii*]) by $RMSE_U$ (result [*vi*]) divided by $RMSE_{LVR}$ (result [*ii*]). Finally, the significance of the LVR coefficients was estimated by dividing the LVR coefficients (result [*i*]) by the adjusted coefficient standard errors just calculated.

The LISREL 8 estimates shown in Table 1 were produced by first attempting to use the Kenny and Judd (1984) approach of itemizing the SXA interaction with all unique products of the indicators of SAT and ALT. The resulting itemization of SXA produced 25 product-indicators and the LISREL 8 structural model for Equation 2 did not fit the data. Next, we attempted to use the Jaccard and Wan (1995) approach of itemizing SXA with a consistent subset of its 25 product-indicators. However, while several consistent subsets of the product-indicators were found, none were judged to be content valid because none Aspanned≅ the indicators of SAT and ALT (i.e., all the indicators of SAT and ALT were not represented in any of these subsets of product indicators). Finally, using the Ping (1995) approach of itemizing SXA with a single product indicator that is the sum of the SAT items times the sum of the ALT indicators, we were able to estimate the Equation 2 model.

Several comments may be of interest. The spreadsheet calculations involved multiplying the unadjusted covariances by functions of the number of items in the measures involved (e.g., G7, G8, etc. in Table 4). This was required because the unadjusted covariances in Table 4 were based on averaged, rather than summed, indicators. Multiplying each unadjusted covariance by functions of the number of items in its measures reverses the effect of averaging and produces usable variances and covariances.

Because of the number of steps involved, LVR is as tedious to use as other structural equation techniques for interactions. In addition, it cannot be used to jointly estimate multiple dependent variables as structural equation analysis can. Thus it appears to be no real threat to the popularity of LISREL, AMOS, EQS, etc.

However, since it does not involve specifying a structural equation model (e.g., LISREL) that involves an interaction (and experiencing estimation difficulties such as those described above), LVR might be useful to substantive researchers not interested in learning the intricacies of specifying and estimating interactions with structural equation analysis (e.g., LISREL). It may also be effective in estimating several interactions or probing for interactions to explain non-significant associations (a disordinal interaction can mask a significant association). The specification of several interactions noticeably increases the nonnormality of a structural equation model, and virtually guarantees unacceptable model-to-data fit; it can also increase structural model convergence and improper solution difficulties.

REFERENCES

Aiken, Leona S. and Stephen G. West (1991), *Multiple Regression: Testing and Interpreting Interactions*, Newbury Park, CA: SAGE Publications.

Baumgartner, Hans and Richard P. Bagozzi (1995), ASpecification, Estimation, and Testing of Moment Structure Models Based on Latent Variates Involving Interactions Among the Exogenous Constructs,≅ *Sociological Methods and Research*, 24 (2) November, 187-213.

Blalock, H. M., Jr. (1965), "Theory Building and the Concept of Interaction," *American Sociological Review*, 30, 374-81.

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.

Cohen, Jacob (1968), "Multiple Regression as a General Data-Analytic System," *Psychological Bulletin*, 70, 426-43.

_______ and Patricia Cohen (1975), *Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences*, Hillsdale, NJ: Lawrence Erlbaum.

Cohen, Jacob (1983), *Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences*, Hillsdale, NJ: Lawrence Erlbaum.

Denters, Bas and Rob A. G. Van Puijenbroek (1989), AConditional Regression Analysis,≅ *Quality and Quantity*, 23 (February), 83-108.

Freidrich, R. J. (1982), AIn Defense of Multiplicative Terms in Multiple Regression Equations,≅ *American Journal of Political Science*, 26, 797-833.

Hanushek, Eric A. and John E. Jackson (1977), *Statistical Methods for Social Scientists*, New York: Academic Press.

Hayduk, Leslie A. (1987), *Structural Equation Modeling with LISREL: Essential and Advances*, Baltimore, MD: Johns Hopkins Press.

Heise, David R. (1986), AEstimating Nonlinear Models,≅ *Sociological Methods and Research*, 14 (4) May, 447-472.

Howard, John A. (1989), *Consumer Behavior in Marketing Strategy*, Englewood Cliffs, NJ: Prentice Hall.

Jaccard, James, Robert Turrisi and Choi K. Wan (1990), *Interaction Effects in Multiple Regression*, Newbury Park, CA: SAGE Publications.

_______ and C. K. Wan (1995), AMeasurement Error in the Analysis of Interaction Effects Between Continuous Predictors Using Multiple Regression: Multiple Indicator and Structural Equation Approaches,≅ *Psychological Bulletin*, 117 (2), 348-357.

Jöreskog, Karl G. (1993), "Testing Structural Equation Models," In *Testing Structural Equation Models*, Kenneth A. Bollen and J. Scott Long eds., Newbury Park, CA: SAGE Publications.

_______ and Dag Sörbom (1996), *LISREL 8 User=s Reference Guide*, Chicago: Scientific Software International, Inc.

_______ and F. Yang (1996), ANonlinear Structural Equation Models: The Kenny and Judd Model with Interaction Effects,≅ *Advances in Structural Equation Modeling Techniques*, G.A. Marcoulides, R.E. Schumacker, eds., Hillsdale, NJ: LEA.

Kenny, David (1979), *Correlation and Causality*, New York: Wiley.

_______ and C. M. Judd (1984), "Estimating the Nonlinear and Interactive Effects of Latent Variables," *Psychological Bulletin*, 96, 201-10.

Lusch, Robert F. and James R. Brown (1996), AInterdependency, Contracting, and Relational Behavior in Marketing Channels,≅ *Journal of Marketing*, 60 (4) October, 19-38.

McClelland, G. H. and C. M. Judd (1993), AStatistical Difficulties of Detecting Interactions and Moderator Effects,≅ *Psychological Bulletin*, 114 (2), 376-390.

Myers, Raymond H. (1986), *Classical and Modern Regression with Applications*, Boston: Duxbury Press, 211.

Osterhuis, Thomas L (1997), APro-social Consumer Influences Strategies: When and How Do They Work?≅ *Journal of Marketing*, 61 (4) October, 16-29.

Ping, R. (1993), "The Effects of Satisfaction and Structural Constraints on Retailer Exiting, Voice, Loyalty, Opportunism, and Neglect," *Journal of Retailing*, 69 (Fall), 320-352.

_______ (1994), "Does Satisfaction Moderate the Association Between Alternative Attractiveness and Exit Intention in a Marketing Channel?≅ *Journal of Retailing*, 22 (4) Fall, 364-371.

_______ (1995), "A Parsimonious Estimating Technique for Interaction and Quadratic Latent Variables," *The Journal of Marketing Research*, 32 (August), 336-347.

_______ (1996a), "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients,≅ A

Multivariate Behavioral Research, 31 (1), 95-120.

_______ (1996b), "Latent Variable Interaction and Quadratic Effect Estimation: A Two-Step Technique Using Structural Equation Analysis, *The Psychological Bulletin*, 119 (January), 166-175.

_______ (1996c), "Estimating Latent Variable Interactions and Quadratics: The State of this Art," *The Journal of Management*, 22 (1), 163-183

_______ (1999), "Unexplored Antecedents of Exiting in a Marketing Channel,≅ *Journal of Retailing*, 75 (2) Summer, 218-242.

Podsakoff, Philip M., William D. Tudor, Richard A. Grover and Vandra L. Huber (1984), ASituational Moderators of Leader Reward and Punishment Behaviors: Fact or Fiction?≅ *Organizational Behavior and Human Performance*, 34, 21-63.

Singh, J. (1998), AStriking a Balance in Boundary-spanning Positions: an Investigation of Some Unconventional Influences of Role Stressors and Job Characteristics on Job Outcomes of Salespeople,≅ *Journal of Marketing*, 62 (3) July, 69-86.

Warren, Richard D., Joan K. White and Wayne A. Fuller (1974), AAn Errors-In-Variables Analysis of Managerial Role Performance,≅ *Journal of the American Statistical Association*, 69 (328) December, 886-893.

Wong, S. K. and J. S. Long (1987), "Reparameterizing Nonlinear Constraints in Models with Latent Variables," Technical Report, Washington State University.

Appendix A- Simulated Data Set Creation

To produce the Table 2 results, the model $Y = b_0 + b_1X + b_2Z + b_3W + b_4XZ + \zeta_Y$ was estimated using simulated data sets that met the Kenny and Judd (1984) normality assumptions (indicators are multivariate normal with mean zero and independent of their measurement errors, and measurement errors are independent of each other) with the Table 3 population parameters. These parameters are the original Kenny and Judd (1984) values for the variances of *X*, *Z*, and *W*, and polar but plausible values for model validation studies. E.g., the loadings and measurement errors produce reliabilities of .7 and .9, and the structural parameters (i.e., b=s and ζ=s) correspond to R^2 's of .10 and .50.

X, *Z*, *W*, and their indicators x_1, ... , x_4, z_1, ... , z_4, w_1, ..., w_4 were created using PRELIS, and a normal random number generator. Each data set contained 100 or 300 cases and was replicated 100 times. For each of these data sets PRELIS and the Table 3 population parameters were used to generate *Y* and its single indicator y_1, again using a normal random number generator. Next, the indicators x_1, ... , x_4, z_1, ... , z_4, w_1, ..., w_4 were summed to form X, Z and W, and these variables and Y (= y_1) were mean centered by subtracting the mean of each variable from each case value for that variable to produce zero means for all variables. XZ (= X times Z) was added to the data set, and the attenuated covariance matrix for X, Z, W, Y and XZ was generated using SPSS. Then the raw data for x_1, ... , x_4, z_1, ... , z_4, w_1, ..., w_4, and y_1 was used in a measurement model with LISREL 8 and maximum likelihood (ML) to produce the loadings and measurement errors for x_1, ... , x_4, z_1, ... , z_4, w_1, ..., w_4. Next these measurement parameters were used to adjust the attenuated covariance matrix, and the resulting adjusted covariance matrix was input to SPSS=s matrix regression procedure to produce latent variable regression (LVR) structural coefficients, and an LVR Root Mean Squared Error (RMSE). Then the attenuated covariance matrix was input to SPSS=s matrix regression procedure to produce unadjusted coefficient standard errors and an unadjusted RMSE. Finally the raw data was input to LISREL 8 to produce (ML) estimates for comparison purposes.

Table 1-- Coefficient Estimates for the Example

Latent Variable (OLS) Regression Results: Dependent Variable = NEG

Variable	Standardized Coefficient	Unstandardized Coefficient	SE[a]	t-value	(p-value)
SAT	-.317	-.279	.068	-4.05	(.000)
ALT	.214	.147	.051	2.84	(.005)
INV	-.217	-.168	.056	-2.98	(.003)
SCT	.066	.044	.045	0.97	(.333)
SxA	-.127	-.080	.050	-1.57	(.117)
Const.		.000	.035	0.00	(1.00)

LISREL 8 (GLS) Results: Endogenous Variable = NEG

Variable	Standardized Coefficient	Unstandardized Coefficient	SE	t-value
SAT	-.337	-.327	.088	-3.70
ALT	.207	.143	.054	2.63
INV	-.185	-.156	.071	-2.20
SCT	.042	.028	.058	0.49
SxA	-.108	-.072	.052	-1.37

χ^2/df/p = 309/216/0.28E-4, GFI = .88, AGFI = .84, CFI = .99, RMSEA = .04

[a] Suggested coefficient standard errors (i.e., coefficient standard errors from unadjusted regression multiplied by the ratio of unadjusted regression Root Mean Squared Errors [RMSE] to adjusted regression RMSE).

Table 2-- Bias in Coefficient Standard Error (SE) and Significant Coefficient Counts for the Suggested Standard Errors (Sugg. SE) and LISREL 8 (L8)

	Bias[a] in SE		Significant Coefficients		
	LVR[b] with Sugg. SE	L8	LVR[b] with Sugg. SE	L8	Difference
lsi100: b_1,b_2,b_4[c]	0.831	0.881	110	82	28
b_3	0.833	0.841	33	29	4
lsi300: b_1,b_2,b_4	0.917	0.946	191	179	12
b_3	0.818	0.823	55	54	1
lsn100: b_1,b_2,b_4	0.878	0.941	106	73	33
b_3	0.959	0.942	0	4	(4)
lsn300: b_1,b_2,b_4	0.958	0.998	185	169	16
b_3	0.913	0.924	25	21	4
lli100: b_1,b_2,b_4	0.884	0.866	265	273	(8)
b_3	0.837	0.825	96	99	(3)
lli300: b_1,b_2,b_4	0.936	0.944	291	289	2
b_3	0.810	0.814	100	100	0
lln100: b_1,b_2,b_4	0.919	0.932	295	289	6
b_3	0.932	0.944	8	4	4
lln300: b_1,b_2,b_4	0.978	0.980	299	298	1
b_3	0.918	0.908	27	28	(1)
hsi100: b_1,b_2,b_4	0.929	0.918	80	81	(1)
b_3	0.938	0.944	25	24	1
hsi300: b_1,b_2,b_4	0.974	0.984	183	180	3
b_3	0.891	0.885	53	56	(3)
hsn100: b_1,b_2,b_4	0.978	0.983	81	80	1
b_3	0.991	0.980	0	5	(5)
hsn300: b_1,b_2,b_4	1.000	1.000	183	182	1
b_3	0.938	0.923	23	27	(4)
hli100: b_1,b_2,b_4	0.926	0.918	296	297	(1)
b_3	0.930	0.931	100	100	0
hli300: b_1,b_2,b_4	0.980	0.984	300	300	0
b_3	0.880	0.883	100	100	0
hln100: b_1,b_2,b_4	0.987	0.978	296	297	(1)
b_3	0.982	0.972	0	4	(4)
hln300: b_1,b_2,b_4	1.000	1.000	300	300	0
b_3	0.929	0.905	19	27	(8)

[a] The average of the suggested coefficient standard errors divided by the root mean square difference of the coefficients from their population values. A bias less than 1 suggests the suggested standard error underestimates the actual coefficient variation (measured in this case the root mean square deviation from the population value).
[b] Latent Variable Regression.
[c] b_1, b_2, and b_4 are the coefficients of *X*, *Z*, and *W* in 100 data sets with *l*ower reliability, *s*mall coefficients, a population *i*nteraction, and a sample size of *100*. The next line shows this information for the *XZ* coefficients. Subsequent lines show combinations that include *h*igh reliability, *l*arge coefficients, with *n*o population *i*nteraction, and a sample size of *300*.

Table 3-- Population Parameters for Simulated Data Sets

Parameter[a]	Population Variance	Population Coefficient
All Data Sets:		
X	2.15	
Z	1.60	
W	1.00	
Corr(X,Z)		0.20
Corr(X,W)		0.20
Corr(Z,W)		0.20
λ_y		1.00
ε_y	0.00	
b_0		0.00
High Reliability Samples (ρ = .9):		
λ_{x1}		1.00
λ_{x2}-λ_{x4}		0.90
λ_{z1}	1.00	
λ_{z2}-λ_{z4}		0.90
λ_{w1}		1.00
λ_{w2}-λ_{w4}		0.90
ε_{x1}-ε_{x4}	0.82	
ε_{z1}-ε_{z4}	0.61	
ε_{w1}-ε_{w4}	0.38	
Low Reliability Samples (ρ = .7):		
λ_{x1}		1.00
λ_{x2}-λ_{x4}		0.70
λ_{z1}	1.00	
λ_{z2}-λ_{z4}		0.70
λ_{w1}		1.00
λ_{w2}-λ_{w4}		0.70
ε_{x1}-ε_{x4}	2.21	
ε_{z1}-ε_{z4}	1.65	
ε_{w1}-ε_{w4}	1.03	
Small Coefficients (R^2 = .10)		
ζ_Y	1.6	
$b_{Y,X}$		-0.15
$b_{Y,Z}$		0.17
$b_{Y,W}$		0.20
$b_{Y,XZ}$		0.12
Large Coefficients (R^2 = .50):		
ζ_Y	0.8	
$b_{Y,X}$		-0.35
$b_{Y,Z}$		0.37
$b_{Y,W}$		0.40
$b_{Y,XZ}$		0.30

[a] $Y = b_{Y,X}X + b_{Y,Z}Z + b_{Y,W}W + b_{Y,XZ}XZ + \zeta_Y$

$x_i = \lambda_{xi}X + \varepsilon_{xi}$

$z_i = \lambda_{zi}Z + \varepsilon_{zi}$

$w_i = \lambda_{wi}W + \varepsilon_{wi}$

Table 4-- Unadjusted Covariances and Measurement Model Results for the Example

		[Col A]	[Col B]	[Col C]	[Col D]	[Col E]	[Col F]	[Col G]
	Unadjusted Covariances:							
		SAT	ALT	INV	SCT	NEG	SxA	
Row 1]	SAT	0.4440214						
Row 2]	ALT	-0.2986497	0.7387649					
Row 3]	INV	0.1690241	-0.1894359	0.6089392				
Row 4]	SCT	0.1703342	-0.3365397	0.4214534	0.9570907			
Row 5]	NEG	-0.0053209	0.0034348	-0.0116940	0.0833171	0.5073915		
Row 6]	SxA	0.2559515	-0.2630964	-0.0110285	0.0544937	0.0775070	0.6710534	
	Loadings:							
		1	2	3	4	5	Sum	# of Indicators
Row 7]	s	0.8020420	0.8935347	1	0.8880973	0.93956542	4.52323953	5
Row 8]	a	0.9388913	0.9082904	1	0.7782776		3.62545943	4
Row 9]	i	0.8971429	0.9864732	1	0.9910887	0.75174001	4.62644498	5
Row 10]	sc	0.9298091	1.0108595	1	1.0286587		3.96932738	4
Row 11]	n	1.2562465	1.4469622	1	0.9756666		4.67887537	4
	Measurement Errors:							
		1	2	3	4	5	Sum	
Row 12]	s	0.16497635	0.1304105	0.09750127	0.11863208	0.10498881	0.61650901	
Row 13]	a	0.26669386	0.2477777	0.08196178	0.23768765		0.83412104	
Row 14]	i	0.44856484	0.1266073	0.09591831	0.11563109	0.44733779	1.23405942	
Row 15]	sc	0.29333068	0.2127870	0.17113061	0.21136906		0.88861735	
Row 16]	n	0.23881176	0.0646641	0.58677303	0.67803413		1.56828306	
	Adjusted Covariance Matrix:[a]							
		SAT	ALT	INV	SCT	NEG	SxA	
Row 17]	SAT	0.512423355 (= [G7^2*A1-F12]/F7^2)[b]						
Row 18]	ALT	*-0.364233127*	0.83583034 (= [G8^2*B2-F13]/F8^2)					
Row 19]	INV	*0.201925726*	*-0.22588210*	0.653589285 (= [G9^2*C3-F14]/F9^2)				
Row 20]	SCT	*0.189742817*	*-0.37417633*	*0.459002695*	0.915539276 (= [G10^2*D4-F15]/F10^2)			
Row 21]	NEG	*-0.005028333*	*0.00323979*	*-0.010804488*	*0.083317100*	0.299196894 (= [G11^2*E5-F16]/F11^2)		
Row 22]	SxA	0.345060666	-0.35402114	-0.014536385	0.066974076	0.080812176	0.93852398 (= [G7^2*G8^2*F6 -F7^2*A17*F13 -F8^2*B18*F12 -F12*F13]/F7^2*F8^2)	
		(=G7^2*G8*A6/[F7^2*F8])	(=G8^2*G7*B6/[F8^2*F7])	(=G7*G8*G9*C6/[F7*F8*F9])	(=G7*G8*G10*D6/[F7*F8*F10])	(=G7*G8*G11*E6/[F7*F8*F11])		

Cov(SAT,ALT) = G7*G8*A2/[F7*F8]. Italicized adjusted covariances are similar

---

[a] SPSS requires a square matrix as input to matrix regression.
[b] Spreadsheet formula for covariance matrix entry.

Hypothesized Associations and Unmodeled Latent Variable Interactions/Quadratics: An F-Test, Lubinski and Humphreys Sets, and Shortcuts Using Reliability Loadings

Robert A. Ping
Associate Professor of Marketing
Raj Soin College of Business Administration
Wright State University
Dayton, OH 45435
rping@wright.edu (937) 775-3047 (FAX) -3545

Topic Areas:
Interactions

Hypothesized Associations and
Unmodeled Latent Variable Interactions/Quadratics:
An F-Test, Lubinski and Humphreys Sets,
and Shortcuts Using Reliability Loadings

When hypothesized associations are non significant in model tests with survey data, researchers typically report this with little further analysis. However, an unmodeled population interaction or a quadratic may have produced this nonsignificant association. In this case, the association's significance varies with the levels of a moderating variable in the study, and its non-significance does not hold for all levels of this moderating variable. The association even may be significant in a future study.

Model-testers are discouraged from looking for interactions or quadratics in survey data after the hypothesized model has been estimated, as experimental researchers routinely do in ANOVA to better interpret associations, because it appears unscientific. In addition, structural equation analysis techniques for detecting interactions and quadratics are difficult to use.

This article discusses these matters, and proposes an accessible approach for post-hoc probing for latent variable interactions and quadratics to better interpret significant and nonsignificant associations.

Hypothesized Associations and
Unmodeled Latent Variable Interactions/Quadratics:
An F-Test, Lubinski and Humphreys Sets,
and Shortcuts Using Reliability Loadings

Testing structural equation models in survey data, when one or more variables is measured with error, has received considerable attention since the 19th century (see Fuller, 1980 for a summary). The resulting approaches to specifying measurement error could be grouped into errors-in-variables approaches that use regression (see Fuller, 1991 for a summary), and covariant structure analysis (structural equation) approaches that use LISREL, EQS, AMOS, etc. (e.g., Bentler & Weeks, 1980; Jöreskog, 1970; and McDonald, 1978). These specification approaches were extended to multiplicative interactions and quadratics (e.g., XZ and XX respectively in

$$Y = b_0 + b_1X + b_2Z + b_3XZ + b_4XX + \zeta \,, \qquad (1$$

where ζ is error or structural disturbance) by Kenny and Judd (1984) for structural equation analysis, and Heise (1986) for errors-in-variables.

However, approaches for specifying interactions and quadratics with measurement error have been slow to diffuse in the Social Sciences (Cortina, Chen & Dunlap, 2001). Errors-in-variables approaches are inaccessible to many substantive researchers, and structural equation analysis approaches when applied to interactions and quadratics have proven difficult for substantive researchers to use (Aiken & West, 1991). (The extant interaction/quadratic specification approaches include Kenny & Judd, 1984; Algina & Moulder, 2001; Bollen, 1995; Jaccard & Wan, 1995; Jöreskog & Yang, 1996; Jöreskog, 2000; Klein & Moosbrugger, 2000/Schermelleh-Engle, Kein & Moosbrugger, 1998/Klein & Muthén, 2002; Marsh, Wen & Hau, 2004; Mathieu, Tannenbaum & Salas, 1992; Moulder & Algina, 2002; Ping, 1995, 1996a and 1996c; and Wall & Amemiya, 2001 and 2003. With rare exceptions these approaches are unknown outside of the methods literature (Cortina, Chen & Dunlap, 2001) and they are tedious to use (Jöreskog & Yang, 1996).) Perhaps as a result, interactions and quadratics are comparatively rare in published model tests involving latent variables and survey data.

Nevertheless, failing to consider unmodeled population interactions or quadratics in a survey model can produce erroneous research findings and misleading recommendations. For example, with a significant interaction in Equation 1, the coefficient of Z factors to

$$Y = b_0 + b_1X + (b_2 + b_3X)Z + b_4XX + \zeta \, , \qquad \text{(1a}$$

rather than b_2 (see Aiken and West, 1991). Because the (factored) coefficient of Z in Equation 1a is now a variable that depends on the various levels of X in the survey, the magnitude, sign, and statistical significance of $b_2 + b_3X$ are variable, and thus very different from the coefficient of Z in an Equation 1 without XZ (i.e., b_2' in

$$Y = b_0' + b_1'X + b_2'Z + \zeta' \,). \qquad \text{(1b}$$

For example, b_2' could be nonsignificant, while $b_2 + b_3X$ could be significant over part(s) of the range of X in the survey. In this event, it is not the case that Z is unassociated with Y in the study. The hypothesized Z-Y association is simply conditional, and its significance depends on the various levels of X observed in the study.

This has important implications. Because b_2' is approximately the same as $b_2 + b_3X_{avg}$, where X_{avg} is the average of X in the study (see Aiken and West, 1991), if X_{avg} is low (small), $b_2 + b_3X_{avg}$ may be numerically small, and thus b_2' may be nonsignificant. In different words, Z may appear to be unrelated to Y in an Equation 1b model, when in the Equation 1 model, for larger values of X in the study $b_2 + b_3X$ may be significant as hypothesized. This also implies that with a significant interaction XZ in Equation 1, the Z-Y association in Equation 1b may be significant in the next study if X_{avg} is larger (higher) in that study.

Alternatively, with a significant interaction in Equation 1, b_2' could be significant, but $b_2 + b_3X$ could be nonsignificant over part of the range of X in a study. In this case, the customary interpretation based on the significance of b_2', that Z was associated with Y, is incorrect: there is a group of respondents in the study where changes in Z had no association with Y. Further, any "management implications" of the apparently significant Z-Y association in Equation 1b due to b_2' may be misleading. Again, there is a group of respondents for which "managing" (changes in) Z had no effect on Y.

Similarly, with a significant quadratic such as XX in Equation 1, Equation 1 can be refactored into $Y = b_0 + (b_1 + b_4X)X + b_2Z + b_3XZ + \zeta$, and the coefficient of X is given by $b_1 + b_4X$, rather than by b_1' in Equation 1b. In this case, the association between X and Y depends on the particular level of X at which this association is evaluated. (Interpreting quadratics and "the association between X and Y depending on the particular level of X at which this association is evaluated" is discussed later.) As a result, b_1' could be significant while $b_1 + b_4X$ could be nonsignificant, or vice versa, which creates the same interpretation and implications issues as a significant interaction.

Thus, for improved interpretation of study results, interactions and quadratics should be investigated post-hoc (i.e., after the hypothesized model has been estimated). Specifically, this may provide plausible explanations for hypothesized but nonsignificant associations, which avoids casting a shadow on the relevant theory because it appears not to apply, and it improves the interpretation of significant associations that may be conditional.

Unfortunately, survey researchers are discouraged from post-hoc probing for interactions (e.g., Aiken & West, 1991; Cohen & Cohen, 1983) on grounds this is unscientific because these variables were not hypothesized. However, the logic of model testing and its variables can easily be separated from the logic of discovery and its variables (interactions/quadratics) (e.g., Hunt, 1983) as long as any discovered interactions/quadratics and their relationships in the hypothesized model are properly presented as hypothetical and to be tested in a later study. Specifically, any interactions/quadratics discovered in post-hoc probing should be presented as potentially an artifact of the sample: Their existence in any population and thus in other samples/studies should be viewed as an empirical question to be answered in later studies.

However, given the estimation difficulties mentioned above in identifying significant interactions or quadratics among latent variables in survey data, substantive researchers who want to probe for these variables may decide to use approaches such as ordinary least squares (OLS) regression or analyzing subsets of data (e.g., median splits). Unfortunately regression estimates of interaction or quadratic structural coefficients for latent variables are well known to be biased (see the demonstrations in Aiken and West, 1991). Similarly, subset analysis is criticized in the psychometric literature for a variety of reasons, including its reduced ability to detect interactions or quadratics (see Maxwell and Delaney, 1993 and the citations therein).

There are other post-hoc estimation concerns as well, such as correlations among interactions and quadratics that can produce no significant interactions or quadratics when several interactions or quadratics are estimated jointly, and extant search techniques, such as forward selection and backward elimination, can be indeterminate in that they can produce different subsets of significant interactions and/or quadratics.

This article discusses these matters: It proposes an accessible approach for, and an example of, post-hoc probing for latent variable interactions and quadratics in model tests involving survey data.

A Suggested Approach

Because of the potential for detecting spurious interactions or quadratics (unhypothesized interactions or quadratics that do not exist in the population and are significant by chance), an F-test is desirable to determine if any unhypothesized interactions or quadratics are likely to be significant above the level of chance. To accomplish this, after the hypothesized structural model has been estimated, all possible interactions and quadratics should be added to the hypothesized model. To decrease the attendant amount of specification work, reliability loadings and measurement errors for these interactions and quadratics could be used (e.g., Mathieu, Tannenbaum and Salas, 1992; see Cortina, Chen and Dunlap, 2001, and Appendix A).

An F-Test

To reduce the likelihood of spurious (chance) interactions or quadratics, the increase in R^2 (e.g., the "Squared Multiple Correlations for Structural Equations" in LISREL) due to adding all implied interactions and quadratics to a model should be significant. A test statistic that assesses this increase is

$$F = [(R_2^2 - R_1^2)/(k_2 - k_1)] / [(1- R_2^2)/(N - k_2 - 1)] \qquad (2$$

where R_2^2 is the total explained variance (Squared Multiple Correlations for Structural Equations) in the structural model with the interactions and quadratics added, R_1^2 is the total explained variance in the structural model with no interactions and quadratics added, k_1 is the number of exogenous variables (predictors) in the structural model without the interactions and quadratics, k_2 is the number of exogenous variables in the structural model plus the number of interactions and quadratics added, and N is the number of cases (see for example Jaccard, Turrisi and Wan, 1990). This F statistic has $k_2 - k_1$ and $N - k_2 - 1$ degrees of freedom.

Calculating F

With a single endogenous or dependent variable (e.g., in the structural model $Y = b_1X + b_2Z + b_3W + \zeta$, Y is the endogenous or dependent variable and there are three exogenous variables or predictors of Y on the right-hand side of the equal sign) the F statistic is a straightforward calculation. Specifically, $k_1 = 3$ and R_1^2 is the explained variance in Y. The interactions are those involving the exogenous variables XZ, XW and ZW, and the quadratics are those involving these exogenous variables XX, ZZ, and WW. So, $k_2 = k_1 + 6 = 9$ and R_2^2 is the explained variance of Y in the structural equation with the interactions and quadratics added (i.e., $Y = c_1X + c_2Z + c_3W + c_4XZ + c_5XW + c_6ZW + c_7XX + c_8ZZ + c_9WW$).

Multiple Endogenous Variables With multiple dependent or endogenous variables the suggested F-test is performed multiple times, once for each endogenous variable. (An overall F-test is discussed later.) First, the linear equations implied by the structural model are written out, and the relevant interactions/quadratics are added. For example, in the structural model with the structural equations

$$Z = d_1X \qquad (3a$$
$$A = d_2X + d_3Z \qquad (3b$$
$$Y = d_4Z \qquad (3c$$
$$B = d_5Z + d_6C\ , \qquad (3d$$

the interactions implied by the model are those involving the variables on the right-hand side of each equation (i.e., XZ and ZC, and not XC--C and X did not occur together in any equation). The quadratics in the variables that comprise these interactions are the quadratics in the variables on the right-hand side of each equation (i.e., XX, ZZ, and CC). Thus, with the relevant interactions and quadratics added to their respective equations the origonal equations become

$$Z = d_1'X + d_7XX \quad (3e$$
$$A = d_2'X + d_3'Z + d_8XX + d_9XZ + d_{10}ZZ \quad (3f$$
$$Y = d_4'Z + d_{11}ZZ \quad (3g$$
$$B = d_5'Z + d_6'C + d_{12}ZZ + d_{13}ZC + d_{14}CC \quad (3h$$

(d_1 through d_6 will change with the addition of the interactions and quadratics). F is computed for Equation 3e with R_1^2 equal to the explained variance for Z in the structural model with Equations 3a, 3b, 3c and 3d all specified together, and $k_1 = 1$. R_2^2 for Equation 3e is Z's explained variance in the structural model with Equations 3e through 3h all specified together, and $k_2 = k_1 + 1$. Similarly, for the F of Equation 3f, R_1^2 for Equation 3b is A's explained variance in the structural model with Equations 3a through 3d specified together, and $k_1 = 2$. R_2^2 for Equation 3f is A's explained variance the structural model with Equations 3e through 3h specified together, and $k_2 = k_1 + 3$. F's for Equations 3g and 3h are computed as F for Equations 3e and Equation 3f, respectively, were.

Several comments may be of interest. If F is significant, it means there is likely to be one or more non-spurious interactions or quadratics in the population model (represented by the present sample). However, because interactions and quadratics are usually highly correlated, none of the interactions or quadratics in the structural equation with all the relevant interactions and quadratics added may be significant. We will discuss this matter later.

If F is not significant, it suggests it is unlikely there are any population interactions or quadratics in the population model. Because the reliability loadings and measurement errors are approximations, so are R_2^2 and F. If the p-value of F is in a neighborhood of 0.05, R_2^2 could be re-estimated using more exact loadings and errors for the interactions and quadratics (e.g., by standardizing X, Z and W so that their variances are equal to 1). However, experience suggests that more exact loadings and errors do not improve the F statistic materially, and it is usually less tedious to conclude there are no population interactions or quadratics in the model.

Examples

We will reanalyze model test data involving the latent variables T, U, V, W, and Y, and their indicators t_i (i = 1,5), u_j (j = 1,4), v_k (k = 1,4), w_p (p = 1,4), and y_q (q = 1,5), involving at least 200 usable survey responses. (Other study details have been omitted to skirt details that are of lesser importance to the example.) The example will also develop a suggested approach for finding significant interactions/quadratics.

The latent variables T through W were hypothesized to be associated with Y, the measures for the latent variables were judged to be unidimensional, valid and reliable, and the unidimensionally specified measurement model for T, U, V, W, and Y was judged to fit the data. However, estimating the structural model

$$Y = \beta_1 T + \beta_2 U + \beta_3 V + \beta_4 W + \zeta \qquad (4$$

(Model I) using LISREL 8 and Maximum Likelihood estimation suggested that T, V and W were not associated with Y (see Table 1). Nevertheless, these non-significant associations seemed counterintuitive. Thus, the possibility that quadratics in T, V or W (i.e., TT, VV, or WW), or interactions involving these variables (i.e., TU, TV, TW, UV, UW or VW), were "masking" the nonsignificant associations was investigated using the approach suggested above.

To do so, each indicator of the independent and dependent variables was mean-centered by subtracting the indicator's average from its value in each of the cases (see Aiken and West, 1991). Then, the indicators for the constituent variables of the interactions and quadratics (i.e., T, U, V and W) were summed and averaged in each of the cases, and a single indicator for each of these interactions and quadratics was formed as the product of the indicator averages of the relevant constituent variables and added to each case (see Ping, 1995). (Averaging reduces the magnitude of the variance of these product-variables, which avoids estimation difficulties produced by a large determinant of the input covariance matrix.) Specifically, the single indicators t:t (= $[(t_1+t_2+t_3+t_4+t_5)/5]^2$), t:u (= $[(t_1+t_2+t_3+t_4+t_5)/5][(u_1+u_2+u_3+u_4)/4]$), t:v (= $[(t_1+t_2+t_3+t_4+t_5)/5][(v_1+v_2+v_3+v_4)/4]$), t:w (= $[(t_1+t_2+t_3+t_4+t_5)/5][(w_1+w_2+w_3+w_4)/4]$), u:u, u:v, u:w, v:v, v:w, and w:w were added to each case.

Next, the Equation 4 structural model was re-specified to include all interactions and quadratics:

$$Y = \beta_1 T + \beta_2 U + \beta_3 V + \beta_4 W + \beta_{TT} TT + \beta_{UU} UU + \beta_{VV} VV + \beta_{WW} WW$$
$$+ \beta_{TU} TU + \beta_{TV} TV + \beta_{TW} TW + \beta_{UV} UV + \beta_{UW} UW + \beta_{VW} VW + \zeta \,. \qquad \text{(4a}$$

These interactions and quadratics were specified using the Equations A7, A7a, A8 and A8a reliability approximations for the product indicator loadings and measurement errors. The previous structural model (i.e., without the interactions and quadratics) values for the correlations among the latent variables, and coefficient alphas for T, U, V and W, were used to calculate the interaction/quadratic reliabilities. In addition, the variances of the interactions and quadratics were freed, and the interactions and quadratics were allowed to covary with each other and the other exogenous variables. (With multivariate normality, interactions and quadratics have zero correlations with their constituent variables (Kendall and Stewart, 1958, see Kenny and Judd, 1984). However, it is well known that interactions and quadratics are correlated with their constituent variables in (non-normal) real-world data, and not correlating these variables diminishes model-to-data fit and thus it may bias structural coefficient estimates. Similarly, with multivariate normality the variance of interactions and quadratics should equal their Kendall and Stewart (1958) values (see Kenny and Judd, 1984). However, constraining interaction/quadratic variances to their Kendall and Stewart values in real-world data can produce several difficulties, including lack of estimation convergence.)

The resulting structural model was estimated using LISREL 8 and Maximum Likelihood estimation (chi square/df/p-value/RMSEA/GFI/AGFI = 659/394/0.0/0.055/0.950/0.799). The squared multiple correlations from this model and those from the previous structural model (without the additional interactions or quadratics) were used in Equation 2 to calculate the proposed F statistic with $R_1^2 = 0.12$, $R_2^2 = 0.24$, $k_1 = 4$, $k_2 = k_1 + 10$, and $N = 200$. The resulting F statistic with 10 and 185 degrees of freedom was significant ($F = 2.95$, $p = 0.002$), which suggested there was likely to be at least one significant interaction or quadratic in the population Equation 4 model.

However, none of the interactions or quadratics in Equation 4 when they were all estimated together was significant (not reported). Experience suggests this is common in real-world data because interactions and quadratics are typically highly correlated. Thus, Equation 4 was probed for significant interactions and quadratics.

As implied earlier, depending on the search technique, different search results can obtain. However, Lubinski and Humphreys' (1990) suggestion that an interaction, XZ for example, should be estimated with its relevant quadratics, XX and ZZ, suggests an alternative search approach: Gauge each interaction with its relevant quadratics (Step 1); then estimate a final model containing only the significant interaction(s)/quadratic(s) from each of these tests (Step 2). This avoids mistaking an interaction for it's related quadratic (see Lubinski and Humphreys, 1990), or vice versa, and the number of interactions/quadratics to be jointly tested in Step 2 is reduced to the number of interactions or fewer, which should materially reduce masking.

Thus, the interaction TU was tested with its relevant quadratics TT and UU. (A significant UU could not be used to explain the nonsignificant T, V or W associations with Y. However, a significant UU would mean the U-Y association was conditional, and it should be interpreted accordingly.) To accomplish this the path coefficients (β's) for the paths between the dependent variable Y and each of the interactions and quadratics in Equation 4a were fixed at zero so their (LISREL) modification indices (MI's) could be examined. This model was estimated (Estimation A--chi square/df/p-value/RMSEA/GFI/AGFI = 695/409/0.0/0.056/0.844/0.798) and the interaction or quadratic β with the largest MI in the Lubinski and Humphreys interaction set TU, TT and UU was found (i.e., other MI's were ignored). Because β for TT-Y had the largest MI of the three, its path coefficient was freed and the model was re-estimated. In this re-estimation the TT-Y path was significant and the β's for TU and UU had MI's that suggested they were non-significant (i.e., below 3.8 which roughly corresponds to a t-value of 2 with 1 degree of freedom). Next, this process of examining the MI's in Estimation A for the Lubinski and Humphreys interaction set for TV was repeated, and VV had the largest MI. Because its MI was above 3.8 its path to Y was significant when it was freed. The process was repeated for each of the other the Lubinski and Humphreys interaction sets (i.e., TW, TT and WW; UV, UU and VV; UW, UU and WW; and VW, VV and WW) using the Estimation A MI's. Finally, the significant interaction(s)/quadratic(s) from the Step 1 Lubinski and Humphreys interaction sets were estimated in a structural model in which only they were specified (Step 2).

The results are shown in Table 2 (Part a), along with the results of estimating the Table 2 (Part a) model using the Ping (1995) estimation approach (with averaged loadings and measurement errors).

Next, the structural equations

$$U = \beta_5 T + \beta_6 W + \zeta \qquad (5a$$

$$W = \beta_7 V + \zeta \qquad (5b$$

$$T = \beta_8 W + \zeta \qquad (5c$$

were added to Equation 4 to create Model II. For Equation 4, estimating this model produced results similar to those in Table 1 (only the U-Y association was significant). The T-U and W-U associations were significant in Equation 5a, and the V-W association in Equation 5b was significant (not reported). As a result, we probed not only the nonsignificant hypothesized Y associations with T, V and W in Equation 4, and the W-T association in Equation 5c for unmodeled interactions or quadratics that might be used as explanations for nonsignificance, but also the significant associations for any unmodeled interactions or quadratics that would mean they were contingent instead of linear as hypothesized in Equations 4, 5a, 5b and 5c. The procedure was the same as before with one addition: Interactions were not relevant for Equations 5b and 5c, and the relevant quadratics were added to those equations to obtain

$$U = \beta_5'T + \beta_6'W + \beta_{TT}'TT + \beta_{TW}'TW + \beta_{WW}'WW + \zeta\ , \qquad (5d$$

$$W = \beta_7'V + \beta_{VV}'VV + \zeta\ , \qquad (5e$$

and

$$T = \beta_8'W + \beta_{WW}''WW + \zeta \qquad (5f$$

for the F-tests. In this case F for Equation 4a was again significant (F = 2.88, k_1 = 4, k_2 = k_1 + 10, N

= 200, $R_1^2 = 0.13$, and $R_2^2 = 0.25$). Similarly F for Equation 5d was (barely) significant ($F = 2.44$, $k_1 = 2$, $k_2 = k_1 + 3$, $N = 200$, $R_1^2 = 0.31$, and $R_2^2 = 0.34$), F for Equation 5e was nonsignificant ($F = 1.94$, $k_1 = 1$, $k_2 = k_1 + 1$, $N = 200$, $R_1^2 = 0.31$, and $R_2^2 = 0.32$), and F for Equation 5f was significant ($F = 9.36$, $k_1 = 1$, $k_2 = k_1 + 1$, $N = 200$, $R_1^2 = 0.01$, and $R_2^2 = 0.06$). (Overall confidence of the significant F's was at least 93% using a Bonferroni approach, which is discussed later.) Next, Lubinski and Humphreys (L&H) interaction sets for Equation 4a suggested TT and VV were again significant, and the L&H interaction set for Equation 5d suggested TT was significant. In Equation 5f L&H interactions sets did not apply, but the modification index for WW suggested it was significant. In the Step 2 structural model with TT and VV added to Equation 4, TT added to Equation 5a and WW added to Equation 5c, these quadratics were significant (not reported).

Discussion

The next step would normally be to interpret the significant post-hoc interactions and quadratics. However, because they were not hypothesized, theoretical justifications were developed for them to further reduce the likelihood of their being an artifact of the data. Stated differently, if a post-hoc interaction or quadratic cannot be theoretically justified, it should not be interpreted or used as an explanation for nonsignificance association because this difficulty with theoretical justification will likely reoccur in subsequent studies. In fact, if an interaction or quadratic cannot be theoretically justified, it probably should not be included in the Step 2 estimation.

Interpreting the theoretically plausible Table 2 Part b results (from probing Model I), U remained significant with the addition of TT and VV; T and V were moderated by TT and VV, respectively, (i.e., TT and VV were significant); and W was now significant. In the conditional structural coefficient for T, ($\beta_1 + \beta_{TT}T$) (Equation 4 with TT and VV added, $Y = \beta_1 T + \beta_2 U + \beta_3 V +$

$\beta_4W + \beta_{TT}TT + \beta_{VV}VV + \zeta = (\beta_1 + \beta_{TT}T)T + \beta_2U + \beta_3V + \beta_4W + \beta_{VV}VV + \zeta$), for a low level of existing T in the study (e.g., T = - 2.96 in Table 3--T was mean centered), small changes in that level of T were nonsignificantly associated with Y. This was true for increasing levels of T, until at a high level of existing T in the study (e.g., T = 1.8), small changes in that level of T were negatively and significantly associated with Y. The V-Y association was similar, except this moderated association became significant just above the average of V in the study (see Table 3). Specifically, for levels of V above the study average of V (V = 0), small changes in those levels of V were negatively and significantly associated with Y.

Turning to the second half of the example (Model II), the nonsignificant associations in Equation 4 for T and V were likely to be explained by the suppressing effects of TT and VV in the sample, and with these quadratics added, W was significant in the sample. Similarly, the nonsignificant W-T association in Equation 5f was likely explained by the suppressing effect of WW in the sample. Since these associations were determined to be theoretically plausible, they ought to be significant in subsequent studies; and with TT and VV added, the W-Y association ought to be significant. However, these results may have been an artifact of the sample and their existence in the study population and thus in other samples from that population is an empirical question to be answered in future studies.

Parenthetically, the significant U-Y association in Equation 4, the significant W-U association in Equation 5a, and the significant V-W association in Equation 5e were likely to be unconditional (unmoderated) in the study, and these associations would be candidates for the customary "provisional confirmation" interpretation (i.e., they were not disconfirmed). The W-Y association in Equation 4 might also receive this interpretation. However, since its significance was the result of TT and VV, and thus may be an artifact of the sample, a caveat such as, "...if confirmed in subsequent studies...," should probably be added.

There was one Step 1 combination of an interaction and its related quadratics in Equation 4 with modification indices that suggested the interaction and both quadratics were significant, and the modification indices were trivially different from each other. This supports Lubinski and Humphreys' (1990) observation that interactions and quadratics can be difficult to distinguish.

In this case the "tie" was broken using substantive theory. Specifically, the interaction seemed implausible. However, this also suggests that the "exact" nonlinear form of an association (i.e., interaction or quadratic) occasionally might be difficult to determine in real-world data without considering relevant theory.

In general, the "winner" in case of "ties" should probably be the interaction/quadratic best supported by theory. This seemingly "backward science" of observing a relationship then finding theory to explain it should not conflict with the logic of science (justification) (see Hunt, 1989): The observed relationship was within the logic of discovery. Finding theoretical support for an observed relationship simply suggests its observation should be likely in the future.

It is possible that when the significant interactions and quadratics resulting from each of the Step 1 Lubinski and Humphreys interaction sets are estimated together (Step 2), one or more of these interactions/quadratics will be nonsignificant. In that case, several approaches could be taken to further investigate the "true" set of (population) interactions and quadratics (e.g., "trimming" nonsignificant interactions/quadratics, backward selection, etc.). However, these approaches could be characterized as capitalizing on chance, and the suggestion of theoretically justifying the significant Step 1 interactions/quadratics, then estimating them in Step 2 (and not proceeding further) may be more defensible.

The Step 2 results with significant quadratics illustrates several points. Experience suggests, and authors believe, that quadratics are more likely than their presence in published substantive research suggests (e.g., Howard, 1989). In addition, while the modification indices for the TV interaction in Equation 4a was significant when considered in its Lubinski and Humphreys interaction set, it was judged to be plausible for one moderation (T moderates V) but implausible for the other (V moderates T). This underscores that a significant interaction, in this case TV, creates two moderations, the T-Y association with the factored coefficient $(\beta_T + \beta_{TV}V)T$, and the V-Y association with the factored coefficient $(\beta_V + \beta_{TV}T)V$, and theoretical support should probably apply to both of the moderations.

For emphasis, these post-hoc results should be presented as an artifact of the study's sample, and their existence in a population should be viewed as an empirical question to be answered in future studies. Specifically, they should be hypothesized in a future study, with theoretical support, and these hypotheses should be tested. However, there are several difficulties with this approach. In addition to the question of the contribution-versus-journal-space of a replication based solely on a previously observed interaction/quadratic (this matter is discussed later), experience suggests that providing theoretical support for an interaction/quadratic can be demanding. Interested readers are directed to Howard (1989); Jaccard, Turrisi and Wan (1990); and Kenny and Judd (1984) for accessible discussions and examples of quadratics in substantive data, and to Aiken and West (1991), and the citations therein, Ajzen and Fishbein (1980), Kenny and Judd (1984) and Ping (1994, 1999) for interactions.

The suggested F-test can become an overall test with multiple structural equations using a

Bonferroni approach to significance (e.g., Model II) (see Neter, Kunter, Nachtsheim and Wasserman, 1996; however, also see Perenger, 1998). Specifically, the confidence of multiple F-tests is greater than 1 minus the sum of the p-values of each test. Thus, in the example the confidence of the three significant F-tests was at least 93% (= 1 - [0.002 + 0.065 + 0.002]).

Although the Table 2 (Part a) reliability results were similar to the Ping (1995) (Part b) results for Model I, final estimation of a Step 2 model should probably use a technique that does not use approximate loadings and error variances, because the behavior of the reliability loadings and measurement errors under all circumstance is unknown.

The model-to-data fit in the example illustrates the utility of single indicator interaction/quadratic specification. In the example, Model I had five latent variables, and specifying an additional 10 interactions and quadratics produced a fit of RMSEA/GFI/AGFI = 0.055/0.950/0.799. While there is little agreement on fit indices (see Bollen and Long, 1993), a Root Mean Error of Approximation (RMSEA) (Steiger, 1990) of 0.05 suggests a close model-to-data fit (see Brown and Cudeck, 1993; Jöreskog, 1993).

As previously mentioned, the coefficient alpha reliability loadings and measurement error variances used in the example, were approximate. Improved estimations could use standardized variables (see Equation A2) and latent variable reliabilities (see Equation A1), or a different interaction/quadratic estimation approach. However, experience suggests the suggested (coefficient alpha) reliability approach may be trustworthy for latent variables with moderate to lower intercorrelations, high reliability, and moderate to larger sample sizes (e.g., Table 2 Part a versus Part b).

Using Maximum Likelihood (ML) estimation is not without its apparent difficulties. While it is now widely believed that ML estimates are robust to "reasonable" departures from normality (i.e., in survey data), their standard errors are believed to be biased (in unknown directions) (see Cortina, Chen and Dunlap, 2001 for more). Nevertheless, in the present case, it is probably safe to use the customary t-value cutoff of 2 in absolute value to judge post-hoc significance because the results are exploratory. However, if more precision is desired, EQS offers an ML (ROBUST) estimator that is less sensitive to data distributional assumptions.

Future Research

Experience suggests that coefficient alpha reliability loadings and measurement error variance estimates are trustworthy under the conditions mentioned earlier. However, it would be interesting to investigate these estimates under more demanding conditions. Specifically, the reliability approximations could be formally investigated using a "best conditions" (realistic) scenario: artificial data sets with many cases, high reliability, and low correlations between XZ, for example, and X and Z. If they appear sufficiently unbiased and efficient, a "worse case" simulation scenario of fewer cases, lower reliabilities and higher intercorrelations might be investigated. This is usually done first for normal data to see if there is any point in repeating the investigation with non-normal data.

In the example, significant interactions and quadratics were identified given a significant F-Test. However, experience suggests the suggested F-test may be conservative. Specifically, even with a nonsignificant F-test, significant interactions and quadratics can frequently be found in survey data. Thus, the F-test may be (conservatively) biased in structural equation analysis, and artificial data sets and an approach similar to that described above could be used to investigate these matters further.

While the logic of discovery does not necessarily require identifying "the" set of population interactions/quadratics, it would be interesting to recover known interactions/quadratics in artificial data sets (see above) using the suggested approach to gauge its efficacy. Specifically, the suggested approach could be formally investigated using several scenarios: normal and nonnormal artificial data sets with many or few cases, high or low reliability, and low and high correlations between XZ, for example, and X and Z.

Summary and Conclusion

The article provided arguments against neglecting the possibility of significant unmodeled interactions or quadratics in theoretical model tests using survey data, and thus for post-hoc probing for interactions/quadratics (after the hypothesized model has been estimated). For example, the article argued that significant unmodeled interactions or quadratics might explain hypothesized but nonsignificant model associations. Not exploring this possibility casts a shadow on the theory that generated the association, it appears not to explain the association. Similarly, in the present era of infrequent replication, it may be unnecessarily risky to interpret (or provide "management implications" for) a significant model association without checking first to see if the association could actually be contingent or moderated by another variable, and thus not significant in some study circumstances.

To guard against finding them by chance, the article suggested an F-Test of the additional explained variance from adding all possible interactions and quadratics. The article also suggested the use of reliability approximations for interaction/quadratic loadings and error variances as a laborsaving approach.

Because interactions and quadratics are correlated, they may suppress each other in a model in which all possible interactions and quadratics are all estimated together. Thus, the article suggested an approach to detecting significant interactions and quadratics using an implication from Lubinski and Humphrey (1990) that some of this suppression involves an interaction (XZ) and its related quadratics (XX and ZZ). The suggested procedure was to write the structural equations implied by the model. Then, all implied interactions and quadratics are added to each of these structural equations, and the change in an F statistic from adding these interactions/quadratics is examined. Next, if this F-test is significant, any significant interaction/quadratic in each Lubinski and Humphreys interaction set is found using modification indices. Next, theoretical justification for each significant interaction/quadratic that results is provided. Then, the interactions/quadratics that were successfully theoretically justified are added to the hypothesized model (the model with no post-hoc interactions/quadratics) and the resulting model is estimated. Finally, the significant interactions/quadratics are interpreted in detail to identify the resulting zones of significance and nonsignificance for the moderated association (e.g., Table 3).

The article argued that this post-hoc probing for interactions/quadratics (after the hypothesized model has been estimated) was within the logic of science as long as any significant interaction/quadratic that was found was also presented as an artifact of the study's sample, and its existence in the study population was viewed as an empirical question to be answered in subsequent studies. A Scenario Analysis could provide an easily conducted subsequent study to be reported along with the study where interactions and/or quadratics were post-hoc probed. Scenario analysis has been used elsewhere in the Social Sciences, and it is an experiment in which subjects, usually students, read written scenarios in which they are asked to imagine they are the subjects of an experiment in which variables are verbally manipulated. Then, these subjects are asked to complete a questionnaire containing the study measures, in this case the same questionnaire as the study with post-hoc probing (see Ping, 2004). The results of this research design when compared with other research designs such as cross sectional surveys (see for example Rusbult, Farrell, Rogers and Mainous, 1988), have been reported to be similar enough to suggest that scenario analysis may be useful in "validating" interaction/quadratic(s) discovered post-hoc in a previous study. When reported with the post-hoc probing study, it would provide a replication without the rigors of finding an outlet for a replication.

References

Aiken, Leona S. & Stephen G. West (1991). *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: SAGE Publications.

Ajzen Icek & Martin Fishbein (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NY: Prentice-Hall.

Algina, James & Bradley C. Moulder (2001). A note on estimating the Jöreskog-Yang model for latent variable interaction using Lisrel 8.3. *Structural Equation Modeling*, 8 (1) 40-52.

Anderson, James C. & David W. Gerbing (1982). Some methods for respecifying measurement models to obtain unidimensional construct measurement. *Journal of Marketing Research*, 19 (November), 453-60.

Anderson, James C. & David W. Gerbing (1984). The effect of sampling error on convergence, improper solutions, and goodness of fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49, 155-173.

Anderson, James C. & David W. Gerbing (1988). Structural equation modeling in practice: a review and recommended two-step approach. *Psychological Bulletin*, 103 (May), 411-23.

Bentler, Peter M. & D. G. Weeks (1980). Linear structural equations with latent variables. *Psychometrika*, 45, 289-308.

Bollen, Kenneth A. & J. Scott Long (1993). Introduction. In *Testing Structural Equation Models*, K. A. Bollen and J. S. Long, eds., Newbury Park, CA: SAGE.

Bollen, Kenneth A. (1995). Structural equation models that are nonlinear in latent variables: a least squares estimator. *Sociological Methodology*, 25, 223-251.

Browne, Michael W. & Robert Cudeck (1993). Alternative ways of assessing model fit. In *Testing*

Structural Equation Models, K. A. Bollen et al. eds. Newbury Park CA: SAGE.

Busemeyer, Jerome R. & Lawrence E. Jones (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin*, 93 (May), 549-62.

Cohen, Jacob & Patricia Cohen (1983). *Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Cortina, Jose M. Gilad Chen & William P. Dunlap (2001). Testing interaction effects in Lisrel: examination and illustration of available procedures. *Organizational Research Methods*, 4 (4), 324-360.

Fuller, Wayne A. (1980). Properties of some estimators for the errors-in-variables model. *The Annals of Statistics*, 8 (2-March), 407-422).

Fuller, Wayne A. (1991). Regression estimation in the presence of measurement error. In *Measurement Errors in Surveys*, Paul; P. Biemer et al. eds. NY: Wiley, 617-635.

Heise, David R. (1986). Estimating nonlinear models. *Sociological Methods and Research*, 14 (4-May), 447-472.

Howard, John A. (1989). *Consumer Behavior in Marketing Strategy.* Englewood Cliffs, NJ: Prentice Hall.

Hunt, Shelby D. (1983). *Marketing Theory: The Philosophy of Marketing Science.* Homewood, IL: Irwin.

Jaccard, James, Robert Turrisi & Choi K. Wan (1990). *Interaction Effects in Multiple Regression.* Newbury Park, CA: SAGE Publications.

Jaccard, James & C. K. Wan (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: multiple indicator and structural equation approaches. *Psychological Bulletin*, 117 (2), 348-357.

Jöreskog, Karl G. (1970). A general method for analysis of covariance structures. *Biometrika*, 57 (August), 239-251.

Jöreskog, Karl G. (1993). Testing structural equation models. In *Testing Structural Equation Models*, Kenneth A. Bollen and J. Scott Long eds. Newbury Park, CA: SAGE.

Jöreskog, Karl G. & Fan Yang (1996). Nonlinear structural equation models: the Kenny and Judd model with interaction effects. *Advances in Structural Equation Modeling Techniques*, G.A. Marcoulides, R.E. Schumacker, eds. Hillsdale, NJ: LEA.

Jöreskog, Karl G. (2000). Latent variable scores and their uses. (on-line paper), http://www.ssicentral.com/lisrel/ techdocs/lvscores.pdf.

Kendall, M. G. & A. Stewart (1958). *The advanced theory of statistics* (Vol. 1). London: Griffin.

Kenny, David & Charles M. Judd (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, 96 (July), 201-10.

Klein, A. G. & H. Moosbrugger (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, 65, 457-474.

Klein, A. G. & B. O. Muthén (2002). Quasi maximum likelihood estimation of structural equation models with multiple interactions and quadratic effects. Unpublished ms., Graduate School of Education, UCLA.

Lubinski, D. & Humphreys, L.G. (1990). Assessing spurious moderator effects: illustrated substantively with the hypothesized ("synergistic") relation between spatial and mathematical ability. *Psychological Bulletin*, 107, 385-393.

Marsh, Herbert W., Zhonglin Wen & Kit-Tai Hau (2004). Structural equation models of latent interactions: evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, 9 (3), 275-300.

Mathieu, J. E., S. I. Tannenbaum & E. Salas (1992). Influences of individual and situational characteristics on measuring of training effectiveness. *Academy of Management Journal,* 35, 828-847.

Maxwell, Scott E. & Harold D. Delaney (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, 113 (1), 181-190.

McDonald, R. P. (1978). A simple comprehensive model of the analysis of covariant structures. *British Journal of Mathematical and Statistical Psychology*, 31, 59-72.

McClelland, G. H. & C. M. Judd (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, 114 (2), 376-390.

Molder, Bradley C. & James Algina (2002). Comparison of methods for estimating and testing latent variable interactions. *Structural Equation Modeling*, 9 (1), 1-19.

Neter, John, Michael H. Kunter, Christopher J. Nachtsheim & William Wasserman (1996). *Applied linear statistical models*. Homewood, IL: Irwin.

Perenger, T. V. (1998), What's wrong with Bonferroni adjustments? *British Medical Journal*, 316 18 Apr), 1236-1238.

Ping, R. (1994). Does satisfaction moderate the association between alternative attractiveness and exit intention in a marketing channel?" *Journal of the Academy of Marketing Science*, 22 (1 Fall), 364-371.

Ping, R. (1995). A parsimonious estimating technique for interaction and quadratic latent variables. *The Journal of Marketing Research*, 32 (August), 336-347.

Ping, R. (1996a). Latent variable interaction and quadratic effect estimation: a two-step technique using structural equation analysis. *Psychological Bulletin*, 119 (January), 166-175.

Ping, R. (1996c). Latent variable regression: a technique for estimating interaction and quadratic coefficients. *Multivariate Behavioral Research*, 31 (1), 95-120.

Ping, R. (1999). Unexplored antecedents of exiting in a marketing channel. *Journal of Retailing*, 75 (2), 218-241.

Ping, R. (2004). On assuring valid measures for theoretical models using survey data. *Journal of Business Research*, 57 (2, February), 125-141.

Rusbult, C. E., D. Farrell, D. Rogers & A. G. Mainous (1988). Impact of exchange variables on exit, voice loyalty and neglect: an integrative model of responses to declining job satisfaction. *Academy of Management Journal*, 39 (September), 599-627.

Schermelleh-Engle, K., A. Kein & H. Moosbrugger (1998). Estimating nonlinear effects using a latent moderated structural equations approach. In *Interaction and Nonlinear Effects in Structural Equation Modeling*, R. E, Schumacker and G. A. Marcoulides, eds. Mahwah, NJ: Erlbaum.

Steiger, J. H. (1990). Structural model evaluation and modification: an interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.

Wall, M. M. & Y. Amemiya (2001). Generalized appended product indicator procedure for nonlinear structural equation analysis. *Journal of Educational and Behavioral Statistics*, 26, 1-29.

Wall, M. M. & Y. Amemiya (2003). A method of moments technique for fitting interaction effects in structural equation models. *British Journal of Mathematical & Statistical Psychology*, 56, 47-63.

Werts, C. E., R. L. Linn & K. G. Jöreskog (1974). Intraclass reliability estimates: testing structural assumptions. *Educational and Psychological Measurement*, 34, 25-33.

Table 1-- Abbreviated Results of Estimating the Equation 4 Structural Model[a]

$$Y = -.044T + .210U - .151V + .118W + \zeta\ (= .392)$$
$$(-.53)\quad (3.13)\quad (-1.98)\quad (1.93)\qquad\qquad (6.71)\quad \text{(t-values)}$$

---

[a] Using LISREL 8 with maximum likelihood estimation. T, U, V, W were allowed to intercorrelated but their measurement errors were not.

Table 2-- Abbreviated Results of Estimating the Equation 4 Structural Model with All (Relevant) Interactions and Quadratics Added, and the Path Coefficients (β=s) for TT and VV Freed (All other interaction and quadratic path coefficients were fixed at zero)

LISREL 8 with Equations A7, A7a, A8 and A8a reliability loadings and measurement errors:(Part a

$$Y = -.127T + .226U - .189V + .132W -.146TT -.139VV + \zeta\ (= .362)$$
$$(-1.37)\quad (3.40)\quad (-2.22)\quad (2.21)\quad (-2.20)\quad (-2.25)\qquad\qquad (6.62)\quad \text{(t-values)}$$

χ^2 = 522	GFI[a] = .87	CFI[b] = .97
df = 377	AGFI[a] = .82	RMSEA[c] = .04

Abbreviated Results of Estimating the Equation 4 Structural Model with All (Relevant) Interactions and Quadratics Added, and the Path Coefficients (β=s) for TT and VV Freed (All other interaction and quadratic path coefficients were fixed at zero)

LISREL 8 with Ping (1995) (Measurement Model) loadings and measurement errors for TT and VV: (Part b

$$Y = -.126T + .225U - .190V + .132W -.119TT -.136VV + \zeta\ (= .362)$$
$$(-1.36)\quad (3.40)\quad (-2.24)\quad (2.21)\quad (-2.19)\quad (-2.26)\qquad\qquad (6.62)\quad \text{(t-values)}$$

χ^2 = 522	GFI[a] = .87	CFI[b] = .97
df = 377	AGFI[a] = .82	RMSEA[c] = .04

---

[a] Shown for completeness only-- GFI and AGFI may be inadequate for fit assessment in larger models (see Anderson and Gerbing, 1984).
[b] .90 or higher indicates acceptable fit (see McClelland and Judd, 1993).
[c] .05 suggests close fit, .051-.08 suggests acceptable fit (Brown and Cudeck, 1993; Jöreskog, 1993).

Table 3-- Unstandardized T-Y and V-Y Associations Implied by the Table 2 Part b Results

T-Y Association				V-Y Association				
T Value[a]	T Coef-ficient [b]	SE of T Coef-ficient[c]	t-value	V Value[d]	V Coef-ficient[e]	SE of V Coef-ficient[f]	t-value	
-2.96	0.22	0.20	1.10	-1.8	0.05	0.15	0.36	
-2	0.11	0.16	0.66	-1	-0.05	0.12	-0.44	
-1	-0.00	0.13	-0.07	0	-0.19	0.10	-1.75	
0	-0.12	0.12	-1.03	1	-0.32	0.12	-2.63	
1	-0.24	0.13	-1.82	1.2	-0.35	0.13	-2.71	
1.8	-0.34	0.15	-2.16					
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(Column)

---

[a] T was mean centered and ranged from -2.96 (= low) to 1.8 in the study.
[b] The coefficient of T was (-.126 - .119T).
[c] The Standard Error (SE) of the T coefficient was

$$\sqrt{Var(b_T+b_{TxT}T)} = \sqrt{(Var(b_T)+T^2Var(b_{TxT})+2TCov(b_T,b_{TxT})} ,$$

where $Var(b_T)$, for example, is the square of the Standard Error (SE) of b_T (= -.126) in Part b of Table 2, $Cov(b_T,b_{TxT}) = r*SE_{bT}*SE_{bTxT}$, and r is the correlation of b_T and b_{TxT} (available in LISREL 8).
[d] V was mean centered and ranged from -1.8 (= low) to 1.2 in the study.
[e] The coefficient of V was (-.190 - .136V).
[f] The Standard Error (SE) of the V coefficient was

$$\sqrt{Var(b_V+b_{VxV}V)} = \sqrt{Var(b_V)+V^2Var(b_{VxV})+2VCov(b_V,b_{VxV})} ,$$

where $Var(b_V)$, for example, is the squares of the Standard Error (SE) of b_V (= - .190) in Part b of Table 2, $Cov(b_V,b_{VxV}) = r*SE_{bV}*SE_{bVxV}$, and r is the correlation of b_V and b_{VxV} (available in LISREL 8).

Appendix A--Reliability Loadings and Measurement Errors

Werts, Linn and Jöreskog (1974) suggested the latent variable reliability (ρ_X) of a measure of a unidimensional latent variable X (i.e., the measure has only one underlying latent variable) is

$$\rho_X = \Lambda_X{}^2 \text{Var}(X)/[\Lambda_X{}^2 \text{Var}(X) + \theta_X] , \qquad \text{(A1}$$

where Λ_X is the sum of the loadings of the indicators of X, Var(X) is the error disattenuated (i.e., measurement model) variance of X, and θ_X is the sum of the measurement error variances of the indicators of X. It is also well known that θ_X is

$$\theta_X = \mathit{Var}(X)(1 - \rho_X) , \qquad \text{(A1a}$$

where *Var*(X) is the error-attenuated variance of X (e.g., obtained from SAS, SPSS, etc.). Solving Equation A1 for the loading of X, Λ_X, and substituting Equation A1a into the result,

$$\Lambda_X = [\mathit{Var}(X)\rho_X / \text{Var}(X)]^{1/2} \approx \rho_X{}^{1/2} , \qquad \text{(A2}$$

where ≈ is approximate equality if *Var*(X) does not equal Var(X), and equality if *Var*(X) and Var(X) are 1 (i.e., if *Var*(X) and Var(X) are standardized).

Busemeyer and Jones (1983) showed that the reliability of XZ, ρ_{XZ}, is

$$\rho_{XZ} = (r_{XZ}{}^2 + \rho_X\rho_Z)/(r_{XZ}{}^2 + 1), \qquad \text{(A3}$$

where $r_{XZ}{}^2$ is the square of the disattenuated correlation of X and Z (i.e., available in a measurement model). Thus, using Equation A2

$$\Lambda_{XZ} \approx \rho_{XZ}{}^{1/2} = [(r_{XZ}{}^2 + \rho_X\rho_Z)/(r_{XZ}{}^2 + 1)]^{1/2} \qquad \text{(A4}$$

Using Equation A1a, A3, and the formula for r_{XZ} (= Cov(X,Z) / Var(X)Var(Z)), where Cov(X,Z) is the error disattenuated (i.e., measurement model) covariance of X and Z, along with the Kenny and Judd (1984) result for normal X and Z that *Var*(XZ) =

$Var(X)Var(Z) + Cov(X,Z)^2$,

$$\theta_{XZ} = Var(XZ)(1 - \rho_{XZ}) = Var(X)Var(Z)(1 - \rho_X\rho_Z) . \tag{A4a}$$

Similarly, based on Equation A3 the reliability of a quadratic such as XX, ρ_{XX}, is

$$\rho_{XX} = (1 + \rho_X^2) / 2 \tag{A5}$$

and using Equation A2

$$\Lambda_{XX} \approx \rho_{XX}^{1/2} = [(1 + \rho_X^2) / 2)^{1/2} \tag{A6}$$

Using Equation A1a, A5, and the Kenny and Judd (1984) result for normal X that $Var(XX) = 2Var(X)^2$,

$$\theta_{XX} = Var(X)^2(1 - \rho_X^2) . \tag{A6a}$$

Anderson and Gerbing (1988) pointed out that for unidimensional constructs there is little practical difference between latent variable reliability ρ and coefficient alpha (α). Thus for unidimensional constructs Equations A4, A4a, A6, and A6a can be approximated by

$$\Lambda_{XZ} \doteq [(r_{XZ}^2 + \alpha_X\alpha_Z)/(r_{XZ}^2 + 1)]^{1/2} , \tag{A7}$$

$$\theta_{XZ} \doteq Var(X)Var(Z)(1 - \alpha_X\alpha_Z) , \tag{A7a}$$

$$\Lambda_{XX} \doteq [(1 + \alpha_X^2) / 2)^{1/2} \text{ and} \tag{A8}$$

$$\theta_{XX} \doteq Var(X)^2(1 - \alpha_X^2) , \tag{A8a}$$

where $\doteq$ is approximate equality, and α is coefficient alpha. These approximate loadings and error variances could then be used in single indicator specifications to probe multiple interactions and quadratics as a labor-saving technique (also see Cortina, Chen and Dunlap, 2001).