

EDITOR'S COMMENTS

An Update and Extension to SEM Guidelines for Administrative and Social Science Research¹

By: **David Gefen**
Professor of Management Information Systems
Drexel University
gefend@drexel.edu

Edward E. Rigdon
Professor of Marketing
Georgia State University
erigdon@gsu.edu

Detmar Straub
Editor-in-Chief, *MIS Quarterly*
Professor of CIS
Georgia State University
dstraub@gsu.edu

Introduction and Objective

With the proliferation of structural equation modeling (SEM) methods in IS and other social science and business disciplines and the apparent varying fit indices and quality assurance methods used by researchers, it is time to at least minimally standardize the methodology and reporting requirements of articles using SEM in behavioral sciences. Besides aiding in the processes of writing and evaluating research reports applying these techniques, a guide can help authors plan and execute their study. (Once a study is completed, it is often difficult to address many SEM analytical issues, although some deviations from data distribution assumptions can be rectified after the fact. An example of this is deviation from acceptable levels of heteroskedasticity which can, as in linear regression (Neter et al. 1990), be addressed by taking a natural log transformation of the data.²)

At its best, a research report lets readers walk in the footsteps of the authors, not only observing results but understanding the alternatives confronting the authors at various stages and the rationales for the choices that the authors made. Such a report reveals potential confounds which may undermine substantive interpretation of the study findings and also provides food for thought for others who may be engaged in similar research. A guide that leads to such a research report will also help researchers produce research that meets high standards. At a bare minimum, a research report must avoid the partial or incomplete reporting of SEM statistics and assumptions which can mislead readers and reviewers. This need is especially urgent considering the apparent misuse of perceived leniencies such as assumptions about minimum sample size in partial least squares (PLS) (Marcoulides and Saunders 2006) and considering the differing opinions about fit indices in covariance-based SEM (CBSEM) (Marsh et al. 2004).

¹The appendices for this editorial comment are located in the "Online Supplements" section of the *MIS Quarterly*'s website (<http://www.misq.org>).

²Data transformations to make distributions more normal can be carried out with the cautionary note that one needs to interpret all statistics in light of how these transformations affect data values.

The objective of this guide, therefore, is to provide a comprehensive, organized, and contemporary summary of the minimal necessary reporting required of SEM research so that researchers and reviewers at *MIS Quarterly* and other top quality academic journals have at hand a standard checklist of what needs to be done and what needs to be reported. In doing so, the guide is presented as a handbook of applied statistics for researchers using SEM, providing a standardized presentation protocol and statistical quality control procedures. We do not claim to be saying anything radically new about SEM. Rather, this guide is a compilation and an organized summary of accepted SEM criteria and an extension of an earlier set of recommendations (Gefen et al. 2000). It extends the thinking of Gefen et al. (2000) in several ways, but is essentially an updating of what we know at the present time about the application of these statistical techniques in practice. It is part of a series on SEM that has been published in IS journals in recent years (Boudreau et al. 2001; Gefen 2003; Gefen and Straub 2005; Gefen et al. 2000; Straub et al. 2004).³ In that the logic and argumentation we offer is not unique to IS studies, these guidelines should be broadly applicable across research in the social and administrative sciences.

Why Use SEM and Which SEM Techniques to Use

Advantages of SEM

SEM has potential advantages over linear regression models that make SEM *a priori* the methods of choice in analyzing path diagrams when these involve latent variables with multiple indicators. Latent variables are theoretical constructs that, prior to neuroscience techniques, could not be measured directly (such as beliefs, intentions, and feelings); they could only be measured indirectly through those characteristics we attribute to them. At least in classical measurement theory (Churchill 1979), such latent variables should be based on relevant theory when they are expressed through measured variables like questionnaire scales. Not recognizing measurement error—the distinction between measures and the constructs being measured—leads to erroneous inference (Rigdon 1994), and so editors and reviewers in IS and related social sciences expect researchers to use statistical methods which either recognize this distinction or at least to some extent ameliorate the consequences of measurement error.

What SEM does is to integrate the measurements (the so-called measurement model) and the hypothesized causal paths (the so-called structural model) into a simultaneous assessment. SEM can analyze many stages of independent and dependent variables, including, in the case of CBSEM, the error terms, into one unified model. This one unified measurement and structural model is then estimated, either together as in CBSEM or iteratively as in PLS, and the results are presented as one unified model in which the path estimates of both the measurement and the structural models are presented as a whole. This process allows a better estimation of both measurement and structural relationships in both CBSEM (Anderson and Gerbing 1988) and PLS (Chin et al. 2008).⁴ This makes the estimates provided by SEM better than those produced by linear regression when the distribution assumptions hold. Even when the constructs of interest can be measured with limited ambiguity (such as price or weight),⁵ there are unique advantages to SEM over linear regression in that SEM allows the creation and estimation of models with multiple dependent variables and their interconnections at the same time. For a detailed discussion of this topic please refer to previous publications (Barclay et al. 1995; Chin 2010; Gefen et al. 2000).

³Previously suggested guidelines include PLS guidelines by Chin (2010), and CBSEM guidelines by McDonald and Ho (2002), as well as recommendations about older versions of CBSEM by Hatcher (1994) and Hoyle (1995). None of the CBSEM guidelines appear to have been universally adopted by IS researchers.

⁴As an example, if we use regression to test a simplified TAM data set, we must first run reliability and factorial validity tests to estimate the measurement error. If deemed to be acceptable, we create indices for the latent variables from the instrument scales using the original data (indices which still contain measurement error). Using only these indices subsequently in multiple regression runs, we examine the hypothesized effects of perceived usefulness and of perceived ease of use on use, but separately from the effects of perceived ease of use on perceived usefulness. One of the advantages of SEM in this case is that it also analyzes how the measurement errors of previous stages affect the estimation of the next stages in the model, as well as better analyzing the measurement errors at each stage. Because it lacks this capability, linear regression may either understate or overstate the path estimates. CBSEM solves this problem by considering all causal paths in a single and holistic statistical run and in doing so allows a better analysis of the measurement error.

⁵The standard procedure in running linear regression models assumes that the predictors are error free.

Choosing Which SEM to Use Based on Exploratory or Confirmatory Research Objectives

Although SEM is preferable to linear regression when multiple valid indicators are available, the two most widely used types of SEM in MIS research, namely PLS and CBSEM, are very different in their underlying philosophy, distributional assumptions, and estimation objectives.⁶ In a nutshell, PLS shines forth in exploratory research and shares the modest distributional and sample size requirements of ordinary least squares linear regression (Chin 1998a, 1998b, 2010; Thompson et al. 1995). PLS, however, does not allow the researcher to explicitly model the measurement error variance/covariance structure as in CBSEM. And so, while PLS can do anything linear regression can do, PLS yields biased parameter estimates.⁷

CBSEM, in contrast, addresses the problem of measurement error by explicitly modeling measurement error variance/covariance structures and relying on a factor analytic measurement model. The factor analytic measurement model isolates random measurement error (unreliability). Systematic measurement error, which is unique to individual observed variables, is also sequestered away from the latent variables, which are defined as the communalities of the observed variables (Jöreskog 1979). Still, systematic measurement error shared across observed variables, which can arise through common method effects or use of otherwise invalid measures, may still contaminate the latent variables and remains a potential confound. PLS path modeling, by contrast, addresses the measurement error problem by creating proxies for the latent variables, proxies which consist of weighted sum composites of the observed variables. Optimization of these weights aims to maximize the explained variance of dependent variables. Since random measurement error is by definition not predictable, maximizing explained variance will also tend to minimize the presence of random measurement error in these latent variable proxies, especially as compared to using simple unit-weighted composites, as might be used in a naïve regression analysis.

The roots of the CBSEM methodology as we know them today lie in Jöreskog's (1969) development of the inferential χ^2 test statistic. This, plus CBSEM's reliance on carefully developed measures based on strong theory, marks CBSEM as a primarily confirmatory methodology. Of course, there is an exploratory element to any study, including one that employs CBSEM (Jöreskog 1993) and lately CBSEM researchers have introduced more flexible measurement models, consistent with a more exploratory orientation (Asparouhov and Muthén 2010; Marsh et al. 2009). Nevertheless, CBSEM's limitations as an exploratory method have long been noted (Spirtes et al. 1990).

PLS path modeling, in contrast, was initially envisioned as a tool for situations that are "data-rich but theory-primitive" (Wold 1985), where the emphasis would be not on latent variables abstracted from reality but on prediction—on accounting for the observed dependent variables as they stand. PLS does not incorporate overidentifying constraints and lacks an overall inferential test statistic of the kind that CBSEM provides. This plus the ability of PLS to accommodate secondary data mark PLS as a tool well suited for exploratory research.⁸

Choosing Which SEM to Use Based on a Lack of a Strong Theory Base (e.g., Archival Data)

As a result of these differences, the factor analytic measurement model of CBSEM provides better protection from measurement error, but at a price. The sparse factor analytic measurement model typical of CBSEM requires that the covariances among the observed variables conform to a network of overlapping proportionality constraints (Jöreskog 1979). It is easy to show that if two observed variables are indeed two unidimensional measures of the same common factor, then the ratio of their covariances with any third observed variable is equal to the ratio of the two variables' loadings on the common factor, after allowing for random sampling error. If the loadings of the two variables are 0.7 and 0.8, for example, then the ratio of their covariances with any third variable must be equal to 0.7 / 0.8 in the population. To obtain a good fit in CBSEM, these constraints must hold for all pairs of reflective measures of each latent variable across all other observed variables in the model (even if the other observed variables are themselves formative measures (Franke et al. 2008)). Deviation from these overlapping proportionality constraints

⁶For a detailed discussion of PLS please refer to Chin (1998a, 1998b, 2010), Fornell and Bookstein (1982), Haenlein (2004), Lohmöller (1989), Wold (1975, 1982, 1985; 1988), and annotated examples in Gefen et al. (2000).

⁷Please see the reasoning below for how this occurs in PLS.

⁸We note that Chin (2010) argues that PLS path modeling is also useful for confirmatory research.

is expressed in the form of correlated measurement errors. Requiring that measurement errors be uncorrelated is equivalent to insisting that these constraints hold exactly (again, within random sampling error). Conformity to such constraints is unlikely to happen by chance. It is more likely that such conformity is the result of a long process of measure development and refinement where the aim from the beginning is to produce such observed variables. As described in Churchill's (1979) classic work and in MacKenzie et al. (2011) in this *MISQ* issue, this process will almost certainly involve multiple rounds of data collection, testing, and refinement. Such a process also requires strong and clear conceptualizations of the latent variables being measured, implying a highly developed theory base. Strong conceptualization is necessary for both initial item development and for ensuring content validity across multiple rounds of item refinement. Secondary or archival data, such as that typically found in corporate databases, which were not created and refined with the aim of conforming to these constraints, are unlikely to do so purely by chance. This is why secondary data usually performs poorly in CBSEM analysis and why using PLS with its less rigorous assumptions could be more appropriate.

In relying on weighted composites rather than common factors, PLS path modeling does not impose these proportionality constraints on the observed variables. Moreover, PLS path modeling does not require that measurement errors be uncorrelated—measurement error covariances are not part of the model at all, and so are unconstrained. So data sets which are not the result of long-term measurement development processes, and which are part of research that is in a more exploratory phase, may perform acceptably in PLS path modeling while they may produce unacceptable results in a CBSEM analysis.

Choosing Which SEM to Use Based on Avoiding Bias in the Estimations

Because the latent variable proxies in PLS path modeling are weighted composites of error-tainted observed variables, these proxies are themselves tainted with error. As a result, PLS path modeling parameter estimates are known to be biased (Chin 1998b; Hsu et al. 2006), except under a hypothetical limiting condition which Wold (1982) named “consistency at large.” This is an unrealistic situation that only occurs when both the sample size and the number of observed variables are infinite. Parameter estimates for paths between observed variables and latent variable proxies are biased upward in PLS (away from zero), while parameter estimates for paths between proxies are attenuated. By contrast, parameter estimates obtained from CBSEM are unbiased when distribution assumptions hold, and are even robust to mild violations of those assumptions.

Choosing Which SEM to Use Based on Formative Scales in the Research Model

Another reason frequently touted for preferring PLS is that a researcher is measuring a construct with formative scales.⁹ The use of formative scales is easily accomplished with PLS (Chin et al. 2003), but it presents challenges in CBSEM. Formative measurement creates identification problems in CBSEM models (see Treiblmaier et al. 2011), and some argue that formative measurement presents logical challenges (Edwards 2010; Edwards and Bagozzi 2000). Formative measurement negates, to some extent, CBSEM's ability to model measurement error (Jarvis et al. 2003; Petter et al. 2007),¹⁰ opening the door to parameter estimate bias, which in turn may induce lack of fit (Rigdon 1994). As a result, CBSEM faces problems in dealing with non-factor measurement models, including problems of statistical identification. These may thwart analysis entirely although there have been some attempts to resolve these issues (e.g., Diamantopoulos and Winklhofer 2001; Jarvis et al. 2003). Additional discussion on this issue can be found in Bagozzi (2011), Bollen (2011), and Diamantopoulos (2011). In contrast, because PLS path modeling works with weighted composites rather than factors, it is compatible with formative measurement.

⁹Formative measurement defines the construct of interest as resulting from or being the consequence of the behavior of its indicators (Chin 1998b). For example, “company performance” as a construct might be decomposed into components or aspects of performance. Individual companies might choose to emphasize different components (ROI, market share, or earnings per share). This creates less parsimonious models, to be sure.

¹⁰Petter et al. (2007) suggest that measurement error can be assessed through the zeta disturbance term in CBSEM. It can also be independently assessed through test–retest reliability estimates.

Obsolete Reasons: Choosing SEM Based on the need to Model Interactions/Moderation

Another consideration often mentioned in the literature is the need to model interactions. There are established methods for doing so in PLS (Chin et al. 2003; Goodhue et al. 2007). Although CBSEM originally did not seem well-suited to modeling interactions, CBSEM methods incorporating interactions with both continuous and discrete latent variables are coming of age. These approaches for modeling interactions in CBSEM range from the centering and recentering approaches of Lin et al. (2010) to more computationally sophisticated approaches, such as those incorporated into the Mplus package (Muthén and Muthén 2010). Currently, there is no compelling reason to choose PLS over CBSEM just because of the need to model latent variable interactions. Please refer to the Appendix A for a discussion of modeling interaction effects in CBSEM.

Obsolete Reasoning: Choosing SEM Based on Distribution Assumptions

Historically, there have been other data issues cited as reasons to prefer PLS path modeling over CBSEM. PLS path modeling makes only the very limited distributional assumptions of ordinary least squares (OLS) regression (Chin et al. 2008). Moreover, PLS path modeling uses bootstrapping to empirically estimate standard errors for its parameter estimates, again avoiding restrictive distributional assumptions. By contrast, CBSEM “grew up” relying on maximum likelihood estimation, which presumes a conditional multivariate normality among the observed variables, and derives estimated standard errors under this same distributional assumption.

Today, however, leading-edge CBSEM software, exemplified by Muthén and Muthén’s (2010) Mplus package, includes estimator options which yield correct results across a broad range of data distributions. These estimation methods work with observed variables whose distributions are non-normal, censored, or even discrete. However, CBSEM remains a large-sample methodology while PLS path modeling shares with least squares regression the ability to obtain parameter estimates at relatively lower sample sizes (Chin et al. 2008; Fornell and Bookstein et al. 1982). Although there is some evidence that analyzing interactions with a small sample size can bias the results in PLS (Goodhue et al. 2007), it should also be noted that the estimates of PLS and CBSEM tend to converge with large sample sizes, if the CBSEM model is correct and the distribution assumptions hold (Barroso et al. 2010).

Overall Judgment about SEM Practices

The choice of SEM should depend on the objectives and assumptions of the chosen SEM tool, and should be explained as such to the reader. It must not be that researchers just choose the SEM that best produces the model they wish to report or as a matter of merely applying defaults.¹¹

There are many aspects of statistical rigor that should be considered when using SEM. We detail these in Appendix B, “Recommended Rigor When Using SEM.”

What to Report in PLS

The objective of the final two sections of the paper is to present the minimal reporting of statistics necessary for the reader to be able to evaluate the conclusions drawn from data. These minimal requirements (see Majchrzak et al. 2005) are:

- (a) In the text:
 - i. The standard reporting of expectations and hypotheses.
 - ii. Why the researchers chose PLS.

¹¹We say this, however, recognizing that some journals have implicit or explicit policies preferring one SEM tool over another and this must be taken into account. MIS journals typically have no explicit preference for CBSEM over PLS. Other disciplines do.

- iii. If items are deleted to improve model fit, this must be reported because it increases the risk of over-specification and thus reduces the ability to generalize the findings or compare them to other studies. In this case it is advisable to add a footnote detailing what items were removed and whether the pattern of significant paths changed as a result.
 - iv. Comparison with the saturated model.¹²
- (b) Appendixes or Tables:
- i. Scales with their means, standard deviations, and correlation among each pair of scales.
 - ii. In the case of reflective scales, also add the reported PLS composite reliability,¹³ R^2 , and square root of the AVE.
 - iii. List of items in each scale with their wordings and loadings (in the case of Mode A measurement) or weights (in the case of Mode B measurement) and associated t-values. The loadings and weights of the measurement items associated with the same latent variable should be approximately the same, unless researchers have *a priori* theory-based expectations of substantial differences in performance across items.
 - iv. Show factorial validity, the equivalent of confirmatory factor analysis. See the example in Gefen and Straub (2005).
- (c) Recommended but Optional Additions:
- i. Common Method Bias analysis.
 - ii. Nonresponse bias analysis based on Armstrong and Overton (1977) or Sivo et al. (2006).
 - iii. Second order constructs, where applicable.
 - iv. Verification of the lack of interaction effects.
 - v. Supportive analyses in linear regression as a footnote indicating that the VIF and Durbin-Watson statistics are within their acceptable thresholds. VIF should be below 10.0 and Durbin-Watson should be close to 2.0 (Neter et al. 1990),¹⁴ and that there are no serious threats of outliers in the data.

What to Report in CBSEM

Background Reporting

The choice among different estimation methods may have a dramatic impact on CBSEM, but the researcher's choice should follow from the attributes of the data, keeping in mind that more familiar estimation methods like maximum likelihood will raise fewer questions with readers (as long as distributional assumptions seem to hold). Researchers should concisely explain their choice of estimation method.¹⁵

¹²This is rarely done in reported PLS research but it should, based on the same philosophical guidelines offered in Gerbing and Anderson (1988). It is mainly needed to compare the theoretical model, which includes only the hypothesized paths, with the saturated model, which includes all possible paths in order to verify (1) that the significant paths in the theoretical model also remain significant in the saturated model, and (2) that adding the paths via the saturated model does not significantly increase the f^2 , a standard measure of effect size. By convention, f^2 values of 0.02, 0.15, and 0.35 are labeled small, medium, and large effects, respectively. Based on Carte and Russell (2003), Chin et al. (2003), and Cohen (1988), f^2 is calculated as $(R^2_{\text{saturated model}} - R^2_{\text{theoretical model}}) / (1 - R^2_{\text{saturated model}})$.

¹³PLS composite reliability is typically higher than Cronbach's alpha, although the same standard of being above .80 or above .90 has been suggested for this statistic as well (Gefen et al. 2000) and seems to have become widely accepted.

¹⁴The significance level of the Durbin-Watson statistic, and therefore what value it should have, depends on the number of data points and the number of independent variables. One needs to look those up in an appropriate table (Durbin and Watson 1951; Neter et al. 1990).

¹⁵This is of a lesser concern in PLS, where users face a set of options regarding the weighting scheme, but experience shows that these choices have minimal impact on results. Researchers may obtain slightly higher R^2 with Mode B estimation (Dijkstra 2010), but may avoid unexpected signs on weights by using Mode A estimation, since each estimated loading is a mere restatement of the zero-order correlations between the individual indicator and the composite proxy. Different inner model weighting schemes also seem to have little effect.

Before recounting the actual data analysis, authors should also make clear whether their analysis is based on correlations (standardized data) or covariances. Of course, researchers need the observations themselves in order to estimate some types of models, to conduct bootstrapping, and to take fullest advantage of diagnostics. Standardization represents a loss of information, and within-group standardization can mislead researchers in multi-group analyses. Standardization is a poor choice in other specific circumstances. For example, a dichotomous variable can never have a variance of 1 (its variance is a function of proportions that can never produce “1” as an answer).

On the other hand, sometimes the scale of measurement (the variance of an observed variable) is completely arbitrary, or else scales of measurement differ to such a large degree that results in original metrics are hard to interpret. In such cases, working with standardized data may be the best choice.¹⁶ Standardization induces bias in factor analysis standard errors estimated under maximum likelihood (Lawley and Maxwell 1971). Although some CBSEM packages include a correction factor, there is good reason to encourage scholars using CBSEM to work with covariance matrices and unstandardized data. In fact, Cohen et al. (2003) do not endorse casually discarding original metrics because standardization carries substantial statistical risks, including the risk of producing nonsensical results. In the case of CBSEM, it would be simply better to adjust variances to reduce differences in scale, rather than standardizing *per se*. Covariances should be preferred even though large differences in variances across observed variables may induce convergence problems in some cases (Muthén and Muthén 2010, p. 415), a problem which researchers can address by adjusting the scales of observed variables to limit these differences. Related to this, authors should also report if they are using polychoric correlations to analyze ordinal data (Yang-Wallentin et al. 2010). Authors should be explicit about these choices because of their practical and statistical consequences.

Fit Indexes

A second issue of contention is the choice of overall fit indices. In reporting any of these indexes, it is important to remember that no mere set of fit indexes—and certainly no single measure of fit—can ensure high quality SEM research (Marsh et al. 2004). Such things are only outward signs of due diligence in the course of research, just as a handshake or a bow may be an outward sign of respect in a social situation. Moreover, fit indices happen late in the research process, when the damage is already done, and fit index values are affected by other design choices as well as by actual fit of the model to the data. Importantly, it must not be that authors pick and choose those indices that best produce what they think are acceptable fit values.

We recommend that authors who delete items in the course of refining their model also present the fit index values obtained before item deletion,¹⁷ to allow readers to evaluate the impact of the deletion. While removing items and adding paths such as covariances between error terms might improve overall fit indexes, doing so might compromise the content validity of the scale through over-fitting, which might also make comparison to other studies problematic.

Reporting about overall fit in CBSEM still begins with the χ^2 and its degrees of freedom, because these numbers are the starting point for so many other indices and evaluations. Historically, CBSEM was launched as an inferential technique with Jöreskog's (1969) development of the χ^2 statistic. If the constraints in the proposed model and all ancillary assumptions (homogeneity, distributional assumptions, and asymptotic sample size) hold true in the population, this statistic follows a central χ^2 distribution. Here, the null hypothesis is that the model and supporting assumptions are correct, and a large/significant χ^2 value amounts to evidence *against* this compound hypothesis. If this hypothesis does not hold exactly but only approximately, then the statistic follows a noncentral χ^2 distribution, and its size becomes a direct function of sample size (Muthén et al. 1997). To the extent that models are only simplifications of reality, it is likely that sufficient statistical power will tend to lead to rejection of models according to this strict χ^2 test. Acknowledging this failure of exact fit, researchers are left to determine whether the model may nevertheless represent a useful approximation. This has given rise to a variety of alternative fit indices.

Some of these alternative fit indices have been shown to be flawed and are not to be relied upon. This category includes the χ^2/df ratio (Carmines and McIver 1981). The χ^2/df ratio shares the χ^2 's dependence on sample size, except when the model and all

¹⁶This is less of a concern in PLS because PLS works primarily with correlation matrices.

¹⁷If authors are worried about the length of their papers, these can be reported in appendices that may or may not be published in the print version of the journal. Many journals now publish online supplements which allow greater page latitude for complete scientific reporting.

supporting assumptions are exactly correct in the population (Marsh et al. 1988). The statistical rationale for using the χ^2/df ratio as a fit index has never been clear (Bollen 1989). While researchers might use the χ^2/df ratio as a simplifying heuristic when comparing alternative models, just as one might properly use the Akaike information criterion (Akaike 1993), it should not be reported or offered as evidence affirming acceptable fit.

Another set of fit indexes about which there is some disagreement include the goodness of fit index (GFI) and the adjusted GFI (AGFI) statistics (Tanaka and Huba 1984). Although widely used in IS studies, GFI and AGFI are known to be upwardly biased (indicating better fit) when the sample size is large, and biased downward when degrees of freedom are large relative to the sample size (Steiger 1989). The necessary value of GFI was originally suggested based on intuition and some data to be above .90 (Chin and Todd 1995). The accepted practice in IS has been to have GFI at or above .90, although slightly lower values have been used when the other fit indices are good.

Regarding alternative fit indices, the most authoritative word comes from Hu and Bentler's (1998, 1999) simulations, although it should be added that not all authorities agree with their recommendations (Marsh et al. 2004). Unlike prior studies that recommended heuristic cut-off values for individual indices, Hu and Bentler argued for combinatorial rules involving multiple fit indices. Moreover, Hu and Bentler recommended different heuristic cut-off values depending on circumstances such as sample size and data distribution. We strongly encourage scholars using CBSEM to personally examine the range of Hu and Bentler's recommendations and determine for themselves how these best fit their own research.

RMR stands for root mean square residual. It is the positive root of the unweighted average of the squared fitted residuals (Jöreskog and Sörbom 2001). Because the scale of this index varies with the scale of the observed variables, researchers typically interpret a standardized RMR (SRMR). A high value of SRMR indicates that residuals are large on average, relative to what one might expect from a well-fitting model. An old rule of thumb says that well-fitting models will have SRMR less than 0.05—but again, these rules of thumb are largely superseded by the work of Hu and Bentler (1998, 1999).

The root mean square error of approximation (RMSEA) is an estimate of lack of fit per degree of freedom. Unlike *ad hoc* fit indices, RMSEA has a known statistical distribution, and thus can be used for hypothesis tests. RMSEA is associated with a “test of close fit” which stands as an alternative to the χ^2 -based test of exact fit (Browne and Cudeck 1993). Browne and Cudeck (1993) suggested an RMSEA value of .05 or less as indicating good approximate fit, while .08 or less indicated approximate fit, and values above 0.10 indicated room for improvement. A detailed discussion about why using a .05 cutoff for RMSEA is not empirically supported is presented by Chen et al. (2008), who also suggest that a value of .10 may be “too liberal” in some cases.

Measures of fit relative to a null or baseline model include the Tucker-Lewis Index (TLI), the comparative fit index (CFI), and relative noncentrality index (RNI). Of these three, CFI appears to be the most widely used—Hu and Bentler (1998, 1999) incorporated CFI within their combinatorial rule. All three should be above .90 (Marsh et al. 2004), although some, including Russell (2002) and Hu and Bentler (1998), recommend all three should be above .95.

It is important to add here that it is acceptable that not all fit indexes be within these threshold rules of thumb (Boudreau et al. 2001), as in Gefen et al. (2003).

What Estimation Method Is Used

By default, CBSEM as currently deployed in statistical packages relies mainly on maximum likelihood estimation, or ML. ML estimation leads to parameter estimates which maximize the likelihood of the data actually observed, conditional on the constraints in the model and supporting assumptions. Today, CBSEM estimation methods include many variations on ML which aim to make ML estimation robust to violations of distributional assumptions. Historically, CBSEM users could choose between ML estimation and generalized least squares (GLS) estimation. However, GLS estimation can produce counter-intuitive results in highly constrained models, and there is not a compelling reason to prefer GLS over ML. Alternatives to ML estimation include weighted least squares (WLS) methods, some of which only reach stability at very large sample sizes. No matter which method is used, the authors should specify which method they used and why.

What to Report

Minimal Requirements (see examples in Gefen et al. 2003; Pavlou and Gefen 2005):

- (a) Appendixes or Tables:
- i. Scales with their means, standard deviation, scale reliability, and squared multiple correlation (SMC) or R^2 of the constructs.¹⁸
 - ii. List of items in each scale with their wordings and loadings. It is advisable to add here also the SMC (R^2) of each measurement item, and if the SMC of the items of any given latent variable are not of the same order of magnitude (for example, it is .70 in one item and only .40 in another), then this must be reported too.
 - iii. Correlation or covariance tables (plus standard deviations) of all the items in the model so that basic CBSEM models can be reconstructed by interested readers. If ordered data, rather than interval scale data, are used, then polychoric correlations should be reported.
 - iv. If measurement items were removed for the model, then the authors should inform the readers about whether there were any significant cross loadings that prompted this action and what these cross loadings were. The authors should also inform the readers how removing these measurement items altered or did not alter the pattern of significant paths in the theoretical model.
- (b) Recommended, but Optional Additions:
- i. Common method bias analysis based on Podsakoff et al. (2003).
 - ii. For reflective scales, calculated square root of the AVE of each scale and its correlation to each other scale.¹⁹
 - iii. Consider whether each construct, based on theory, is better represented as a first order construct or as a second order construct.
 - iv. Nonresponse bias analysis based on Armstrong and Overton (1977) or Sivo et al. (2006).
 - v. Comparison of the theoretical model with the saturated model (confirmatory factor analysis or CFA model) based on the guidelines in Gerbing and Anderson (1988). These would show whether the saturated model, which allows all constructs to covary freely, has a significantly lower χ^2 from the structurally constrained theoretical model, given the difference in degrees of freedom.
 - vi. Since heteroscedasticity, influential outliers, and multicollinearity will bias the analysis in SEM just as they do in linear regression, it would be advisable to add supportive analyses in linear regression as a footnote indicating that the VIF and Durbin-Watson statistics are within their acceptable thresholds and that there are no serious threats of outliers in the data. CBSEM models often involve a few predictors for each dependent variable, but researchers should not dismiss this danger out of hand (Grewal et al. 2004).
 - vii. Informing the reader if this was a one step analysis, joining CFA and the structural model, or a two step analysis, doing the CFA first and the structural model afterward (Gerbing and Anderson 1988). Adding a footnote if different results were obtained with the other method is also recommended.

Conclusion

Today social and administrative sciences are examining critical phenomenon for causal relationships and, for this reason, scientists need to be particularly aware of the accepted community standards for using and applying SEM. More sophisticated statistical tools call for more sophisticated analyses and application of best current practices. We are not arguing for unanimity of practice when applying SEM inferential logic; we are arguing that our scientific community must engage in this analysis intelligently and conscientiously and not simply parrot past practice when that practice is, in many cases, weak or indefensible. We must continue to educate ourselves in which benchmarks, guidelines, and heuristics make the most sense in interpreting SEM theories and models. With attention to such details, we believe that the community can accomplish this noteworthy goal.

¹⁸Reliability in first-order factors is calculated as $(\sum \lambda_t)^2 / [(\sum \lambda_t)^2 + \sum \text{Var}(\epsilon_t)]$, where λ_t is indicator loading of item t , and $\text{Var}(\epsilon_t) = 1 - \lambda_t^2$. Reliability in second-order construct is calculated as multiplication of the standardized loadings of the first-order factors with the standardized loadings of the second-order factor. See Kumar et al. (1995) for a discussion; for an example of its use, see Pavlou and Gefen (2005).

¹⁹AVE is average variance extracted. It is calculated as $(\sum \lambda_i^2) / ((\sum \lambda_i^2) + (\sum (1 - \lambda_i^2)))$.

References

- Akaike, H. 1993. "Factor Analysis and AIC," *Psychometrika* (52:3), pp. 317-332.
- Anderson, J. C., and Gerbing, D. W. 1988. "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin* (103:3), pp. 411-423.
- Armstrong, J. S., and Overton, T. S. 1977. "Estimating Non-Response Bias in Mail Surveys," *Journal of Marketing Research* (14), pp. 396-402.
- Asparouhov, T., and Muthén, B. 2010. "Bayesian Analysis of Latent Variable Models Using Mplus," Working Paper (<http://www.statmodel.com/download/BayesAdvantages18.pdf>).
- Barclay, D., Thompson, R., and Higgins, C. 1995. "The Partial Least Squares (PLS) Approach to Causal Modeling: Personal Computer Adoption and Use an Illustration," *Technology Studies* (2:2), pp. 285-309.
- Bagozzi, R. P. 2011. "Measurement and Meaning in Information Systems and Organizational Research: Methodological and Philosophical Foundations," *MIS Quarterly* (35:2), pp. 261-292.
- Barroso, C., Cerrion, G. C., and Roldan, J. L. 2010. "Applying Maximum Likelihood and PLS on Different Sample Sizes: Studies on SERVQUAL Model and Employee Behavior Model," in *Handbook of Partial Least Squares: Concepts, Methods and Applications*, V. E. Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Berlin: Springer-Verlag, pp. 427-447.
- Bollen, K. A. 1989. *Structural Equations with Latent Variables*, New York: Wiley-Interscience.
- Bollen, K. A. 2011. "Evaluating Effect, Composite, and Causal Indicators in Structural Equation Models," *MIS Quarterly* (35:2), pp. 359-372.
- Boudreau, M., Gefen, D., and Straub, D. W. 2001. "Validation in IS Research: A State-of-the-Art Assessment," *MIS Quarterly* (25:1), pp. 1-16.
- Browne, M. W., and Cudeck, R. 1993. "Alternative Ways of Assessing Model Fit," in *Testing Structural Equation Models*, K. A. Bollen and J. S. Long (eds.), Newbury Park, CA: Sage Publications.
- Carmines, E. G., and McIver, J. P. 1981. "Analyzing Models with Unobserved Variables: Analysis of Covariance Structures," in *Social Measurement*, G. W. Bohmstedt and E. F. Borgatta (eds.), Thousand Oaks, CA: Sage Publications, pp. 65-115.
- Carte, T. A., and Russell, C. J. 2003. "In Pursuit of Moderation: Nine Common Errors and their Solutions," *MIS Quarterly* (27:3), pp. 479-501.
- Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., and Paxton, P. 2008. "An Empirical Evaluation of the Use of Fixed Cutoff Points in RMSEA Test Statistic in Structural Equation Models," *Sociological Methods Research*. (36:4), pp. 462-494.
- Chin, W. W. 1998a. "Issues and Opinion on Structural Equation Modeling," *MIS Quarterly* (22:1), pp. vii-xvi.
- Chin, W. W. 1998b. "The Partial Least Squares Approach to Structural Equation Modeling," in *Modern Methods for Business Research*, G. A. Marcoulides (ed.), Mahwah, NJ: Lawrence Erlbaum Associates, Inc., pp. 295-336.
- Chin, W. W. 2010. "How to Write Up and report PLS Analyses," in *Handbook of Partial Least Squares*, V. E. Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Berlin: Springer-Verlag, pp. 655-690.
- Chin, W. W., Marcolin, B. L., and Newsted, P. R. 2003. "A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study," *Information Systems Research* (14:2), pp. 189-217.
- Chin, W. W., Peterson, R. A., and Brown, S. P. 2008. "Structural Equation Modeling in Marketing: Some Practical Reminders," *Journal of Marketing Theory and Practice* (16:4), pp. 287-298.
- Chin, W. W., and Todd, P. A. 1995. "On the Use, Usefulness, and Ease of Use of Structural Equation Modeling in MIS Research: A Note of Caution," *MIS Quarterly* (19:2), pp. 237-246.
- Churchill, Jr., G. W. 1979. "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research* (16), pp. 64-73.
- Cohen, J. B. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J. B., Cohen, P., West, S. G., and Aiken, L. S. 2003. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.), Mahwah, NJ: Lawrence Erlbaum Associates.
- Diamantopoulos, A. 2011. "Incorporating Formative Measures into Covariance-Based Structural Equation Models," *MIS Quarterly* (35:2), pp. 335-358.
- Diamantopoulos, A., and Winklhofer, H. M. 2001. "Index Construction with Formative Indicators: An Alternative to Scale Development," *Journal of Marketing Research* (38:2), pp. 269-277.
- Dijkstra, T. K. 2010. "Latent Variables and Indices: Herman Wold's Basic Design and Partial Least Squares," in *Handbook of Partial Least Squares*, V. E. Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Berlin, Germany: Springer-Verlag, pp. 23-46.
- Durbin, J., and Watson, G. S. 1951. "Testing for Serial Correlation in Least Squares Regression, II," *Biometrika* (38:1-2), pp. 159-179.
- Edwards, J. E. 2010. "The Fallacy of Formative Measurement," *Organizational Research Methods* (forthcoming).
- Edwards, J. E., and Bagozzi, R. C. 2000. "On the Nature and Direction of Relationships Between Constructs and Measures," *Psychological Methods* (5:2), pp. 155-174.

- Franke, G. R., Preacher, K. J., and Rigdon, E. E. 2008. "Proportional Structural Effects of Formative Indicators," *Journal of Business Research* (61), pp. 1229-1237.
- Fornell, C., and Bookstein, F. L. 1982. "Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory," *Journal of Marketing Research* (19), pp. 440-452.
- Gefen, D. 2003. "Unidimensional Validity: An Explanation and Example," *Communications of the Association for Information Systems* (12:2), pp. 23-47.
- Gefen, D., Karahanna, E., and Straub, D. W. 2003. "Trust and TAM in Online Shopping: An Integrated Model," *MIS Quarterly* (27:1), pp. 51-90.
- Gefen, D., and Straub, D. W. 2005. "A Practical Guide to Factorial Validity Using PLS-Graph: Tutorial and Annotated Example," *Communications of the Association for Information Systems* (16), pp. 91-109.
- Gefen, D., Straub, D. W., and Boudreau, M.-C. 2000. "Structural Equation Modeling and Regression: Guidelines for Research Practice," *Communications of the Association for Information Systems* (4:7), pp. 1-70.
- Gerbing, D. W., and Anderson, J. C. 1988. "An Updated Paradigm for Scale Development Incorporating Unidimensionality and its Assessment," *Journal of Marketing Research* (25:May), pp. 186-192.
- Goodhue, D., Lewis, W., and Thompson, R. 2007. "Statistical Power in Analyzing Interaction Effects: Questioning the Advantage of PLS with Product Indicators," *Information Systems Research* (18:2), pp. 211-227.
- Grewal, R., Cote, J. A., and Baumgartner, H. 2004. "Multicollinearity and Measurement Error in Structural Equation Models: Implications for Theory Testing," *Marketing Science* (23:4), pp. 519-529.
- Haenlein, M. 2004. "A Beginner's Guide to Partial Least Squares Analysis," *Understanding Statistics* (3:4), pp. 283-297.
- Hatcher, L. 1994. *A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling*, Cary, NC: SAS Institute.
- Hoyle, R. H. 1995. *Structural Equation Modeling: Concepts, Issues, and Applications*, Thousand Oaks, CA: Sage Publications.
- Hsu, S.-H., Chen, W.-H., and Hsieh, M.-J. 2006. "Robustness Testing of PLS, LISREL, EQS and ANN-Based SEM for Measuring Customer Satisfaction," *Total Quality Management* (17:3), pp. 355-371.
- Hu, L.-T., and Bentler, P. M. 1998. "Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification," *Psychological Bulletin* (3), pp. 424-453.
- Hu, L.-T., and Bentler, P. M. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structural Analysis: Conventional Criteria Versus New Alternatives," *Structural Equation Modeling* (6:1), pp. 1-55.
- Jarvis, C. B., MacKenzie, S. B., and Podsakoff, P. M. 2003. "A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research," *Journal of Consumer Research* (30:2), pp. 199-218.
- Jöreskog, K. G. 1969. "A General Approach to Confirmatory Maximum Likelihood Factor Analysis," *Psychometrika* (34:2), pp. 183-202.
- Jöreskog, K. G. 1979. "Basic Ideas of Factor and Component Analysis," in *Advances in Factor Analysis and Structural Equation Models*, K. G. Jöreskog and D. Sörbom (eds.), New York: University Press of America, pp. 5-20.
- 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 Publications, pp. 294-316.
- Jöreskog, K. G., and Sörbom, D. 2001. *LISREL 8: User's Reference Guide*, Chicago: Scientific Software International.
- Kumar, N., Scheer, L. K., and Steenkamp, J.-B. 1995. "The Effects of Perceived Interdependence on Dealer Attitudes," *Journal of Marketing Research* (32), pp. 348-356.
- Lawley, D. N., and Maxwell, A. E. 1971. *Factor Analysis as a Statistical Method* (2nd ed.), London: Butterworths.
- Lin, G.-C., Wen, Z., Marsh, H. W., and Lin, H.-S. 2010. "Structural Equation Models of Latent Interactions: Clarification of Orthogonalizing and Double-Mean-Centering Strategies," *Structural Equation Modeling* (17:3), pp. 374-391.
- Lohmöller, J.-B. 1989. *Latent Variable Path Modeling with Partial Least Squares*, Heidelberg, Germany: PhysicaVerlag.
- MacKenzie, S. B., Podsakoff, P. M., and Podsakoff, N. P. 2011. "Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques," *MIS Quarterly* (35:2), pp. 293-334.
- Majchrzak, A., Beath, C. M., Lim, R. A., and Chin, W. W. 2005. "Managing Client Dialogues During Information Systems Design to Facilitate Client Learning," *MIS Quarterly* (29:4), pp. 653-672.
- Marcoulides, G. A., and Saunders, C. 2006. "PLS: A Silver Bullet?," *MIS Quarterly* (30:2), pp. iii-ix.
- Marsh, H. W., Balla, J. R., and McDonald, R. P. 1988. "Goodness of Fit Indexes in Confirmatory Factor Analysis: The Effect of Sample Size," *Psychological Bulletin* (103), pp. 391-410.
- Marsh, H. W., Hau, K.-T., and Wen, Z. 2004. "In Search of the Golden Rules: Comment on the Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings," *Structural Equation Modeling* (11:3), pp. 320-341.

- Marsh, H. W., Muthén, B., Asparouhov, A., Lüdtke, O., Robitzsch, A., Morin, A. J. S., and Trautwein, U. 2009. "Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching," *Structural Equation Modeling* (16), pp. 439-476.
- McDonald, R. P., and Ho, M.-H. R. 2002. "Principles and Practice in Reporting Structural Equation Analyses," *Psychological Methods* (7:1), pp. 64-82.
- Muthén, B., du Toit, S. H. C., and Spisic, D. 1997. "Robust Inference Using Weighted Least Squares and Quadratic Estimating Equations in Latent Variable Modeling with Categorical and Continuous Outcomes," Working Paper (http://gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf).
- Muthén, L. K., and Muthén, B. O. 2010. *Mplus 6 User's Guide*, Los Angeles, CA: Muthén and Muthén.
- Neter, J., Wasserman, W., and Kutner, M. H. 1990. *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Design* (3rd ed.), New York: McGraw-Hill, Inc.
- Pavlou, P. A., and Gefen, D. 2005. "Psychological Contract Violation in Online Marketplaces: Antecedents, Consequences, and Moderating Role," *Information Systems Research* (16:4), pp. 372-399.
- Petter, S., Straub, D. W., and Rai, A. 2007. "Specifying Formative Constructs in Information Systems Research," *MIS Quarterly* (31:4), pp. 623-656.
- Podsakoff, P. M., Lee, J. Y., and Podsakoff, N. P. 2003. "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of Applied Psychology* (88:5), pp. 879-903.
- Rigdon, E. E. 1994. "Demonstrating the Effects of Unmodeled Random Measurement Error," *Structural Equation Modeling* (1:4), pp. 375-380.
- Russell, D. W. 2002. "In Search of Underlying Dimensions: The Use (and Abuse) of Factor Analysis," *Personality and Social Psychology Bulletin* (28), pp. 1629-1646.
- Sivo, S. A., Saunders, C., Chang, Q., and Jiang, J. J. 2006. "How Low Should You Go? Low Response Rates and the Validity of Inference in IS Questionnaire Research," *Journal of the AIS* (7:6), pp. 361-414.
- Sirtes, P., Scheines, R., and Glymour, C. 1990. "Simulation Studies of the Reliability of Computer-Aided Model Specification Using the Tetrad II, EQS and LISREL Programs," *Sociological Methods & Research* (19:1), pp. 3-66.
- Steiger, J. H. 1989. *EzPATH: A Supplementary Module for SYSTAT and SYGRAPH*, Evanston, IL: Systat.
- Straub, D. W., Boudreau, M.-C., and Gefen, D. 2004. "Validation Guidelines for IS Positivist Research," *Communications of the Association for Information Systems* (13), pp. 380-427.
- Tanaka, J. S., and Huba, G. J. 1984. "Confirmatory Hierarchical Factor Analyses of Psychometric Distress Measures," *Journal of Personality and Social Psychology* (46), pp. 621-635.
- Thompson, R., Barclay, D. W., and Higgins, C. A. 1995. "The Partial Least Squares Approach to Causal Modeling: Personal Computer Adoption and Use as an Illustration," *Technology Studies: Special Issue on Research Methodology* (2:2), pp. 284-324.
- Treiblmaier, H., Bentler, P. M., and Mair, P. 2011. "Formative Constructs Implemented via Common Factors," *Structural Equation Modeling* (18), pp. 1-17.
- Wold, H. O. 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. Capocchi (eds.), New York: Academic Press, pp. 307-357.
- Wold, H. O. 1982. "Soft Modeling: The Basic Design and Some Extensions," in *Systems Under Indirect Observation: Causality, Structure, Prediction, Part II*, K. G. Jöreskog and H. Wold (eds.), Amsterdam: North-Holland.
- Wold, H. O. 1985. "Partial Least Squares," in *Encyclopedia of Statistical Sciences, Volume 6*, S. Kotz and N. L. Johnson (eds.), New York: Wiley, pp. 581-591.
- Wold, H. O. 1988. "Predictor Specification," in *Encyclopedia of Statistical Sciences, Volume 8*, S. Kotz and N. L. Johnson (eds.), New York: Wiley, pp. 587-599.
- Yang-Wallentin, F., Jöreskog, K. G., and Luo, H. 2010. "Confirmatory Factor Analysis of Ordinal Variables with Misspecified Models," *Structural Equation Modeling* (17:3), pp. 392-423.

EDITOR'S COMMENTS

An Update and Extension to SEM Guidelines for Administrative and Social Science Research

By: **David Gefen**
Professor of Management Information Systems
Drexel University
gefend@drexel.edu

Edward E. Rigdon
Professor of Marketing
Georgia State University
erigdon@gsu.edu

Detmar Straub
Editor-in-Chief, *MIS Quarterly*
Professor of CIS
Georgia State University
dstraub@gsu.edu

Appendix A: Modeling Interaction Effects in CBSEM

Among the several approaches proposed for estimating interactions in latent variable models in CBSEM, Bollen and Paxton (1998) and Little et al. (2006) describe procedures that do not require special-purpose software, and so are accessible to all CBSEM users. Bollen and Paxton described an approach using two-stage least squares (2SLS), an analytical approach used in regression to overcome estimation problems which would confound OLS. 2SLS associates each predictor in a regression model with an instrumental variable. In the first stage, each predictor is regressed on its instrument, then in the second stage the ultimate dependent variable is regressed on the expected or predicted portion of each predictor from the first stage.

In this case, the researcher wants to estimate an interaction model involving latent variables, using only observed variables, each of which is contaminated with error. Under assumptions, two indicators x_1 and x_2 , reflecting the same latent variable share variance due only to the latent variable, while their error terms are mutually uncorrelated. If x_1 is regressed on x_2 , then \hat{x}_1 , the portion of x_1 explained by x_2 , can be used as an error-free substitute for the latent variable. Thus, Bollen and Paxton's 2SLS approach involves 2SLS estimation with the ultimate dependent variable regressed on indicators of the main effect latents and the interaction latent variable, using other indicators of those latents as instruments. Klein and Moosbrugger (2000) demonstrated that the 2SLS approach is not as statistically efficient as more recent methods. Still, this approach can be used to estimate latent variable interaction models using most any standard statistics package.

Little et al. (2006) proposed an "orthogonalizing" strategy to overcome the problem of collinearity between main effects and the interaction term. First, create indicators of the interaction term by multiplying indicators of the main effect constructs. Then regress each product indicator on the main effect indicators used to form each product. The residuals from these regressions are retained as indicators of an interaction latent variable, which is completely orthogonal to the main effect latent variables. Marsh et al. (2007) showed that researchers using this approach could dispense with special and difficult parameter constraints, allowing estimation of the continuous latent variable interaction model using any standard CBSEM package.

References

- Bollen, K. A., and Paxton, P. 1998. "Interactions of Latent Variables in Structural Equation Models," *Structural Equation Modeling* (5:3), pp. 267-293.
- Klein, A., and Moosbrugger, H. 2000. "Maximum Likelihood Estimation of Latent Interaction Effects with the LMS Method," *Psychometrika* (65:4), pp. 457-474.
- Little, T. D., Bovaird, J. A., and Widaman, K. F. 2006. "On the Merits of Orthogonalizing Power and Product Terms: Implications for Modeling Interactions Among Latent Variables," *Structural Equation Modeling* (13:4), pp. 497-519.
- Marsh, H. W., Wen, Z., Hau, K.-T., Little, T. D., Bovaird, J. A., and Widaman, K. F. 2007. "Unconstrained Structural Equation Models of Latent Interactions: Contrasting Residual- and Mean-Centered Approaches," *Structural Equation Modeling* (14:4), pp. 570-580.

Appendix B: Recommended Rigor When Using SEM

SEM Quality Assurance Steps that Should Be Taken

Statistical rigor is crucial in SEM, as it is in linear regression. The following recommendations apply.

Number of Observed Variables/Measures Needed

How many measurement/indicators there should be for each reflective latent variable is still an open debate in the literature. With PLS, the bias in parameter estimates is an inverse function of the number of observed variables per construct (Dijkstra 2010; McDonald 1996). With CBSEM, the impact depends on whether measurement models are formative or reflective. With formative scales (Diamantopoulos and Winklhofer 2001) the main issue is completeness. Are all components of a hypothetical formatively measured construct represented in the data? With reflective, factor-analytic measurement models, multiple observed variables per construct contribute strongly to the model's degrees of freedom, which is a primary driver of statistical power (MacCallum et al. 1996). Three observed variables loading exclusively on one common factor make the individual factor measurement models statistically identified. More than three observed variables make each model over-identified, increasing the researcher's ability to detect lack of fit. Encompassing the measurement models of multiple correlated constructs within a single confirmatory factor model or imposing suitable constraints on parameter estimates will allow a model to achieve identification with fewer observed variables per construct, but with a more limited ability to detect poorly performing measures.

A design that includes only a few measures per construct also puts a study at risk: if one or two of a small number of measures behave in unexpected ways, researchers may be forced to choose methods based on their analytical feasibility rather than based on their ability to best address the research questions that motivated the study.

Common Method Bias

When developing data sources for a study, researchers must be wary of common method bias (CMB). CMB can inflate estimates of structural parameters in the model and may result in erroneous conclusions (Podsakoff et al. 2003). Addressing CMB is not an integral part of SEM and so auxiliary analysis needs to be carried out to assess it, although the cross-loadings CMB creates will be evident in CBSEM through observed lack of fit.

Several steps can be taken when designing the data collection to reduce CMB. One is collecting data with different methods or sources or at different points in time (Podsakoff et al. 2003). These different sources may include multiple informants (Kumar et al. 1993), corporate databases, transactional data, financial statements, or other independent sources.

CMB can also be addressed to some degree in questionnaires by including reverse-scored items to reduce acquiescence, and through statistical analysis such as Harman's single-factor test (Podsakoff et al. 2003). In this method, all of the items are examined in an exploratory factor analysis and the researchers verify that in the unrotated solution there is more than one dominant factor. A better way to test for CMB is to add a latent variable to the model, representing conceptually the shared variance of all the measurement items, and have all the measurement

items load also onto it in addition to the constructs onto which they theoretically load (Podsakoff et al. 2003). This method uses the ability of PLS and CBSEM, in contrast to linear regression, to model a loading of a measurement item on more than one latent construct.¹

Unidimensionality

Another crucial aspect of SEM that is almost always ignored (Malhotra et al. 2006) is distributional assumption checking (especially unidimensionality). Unidimensionality means that there is only one theoretically and statistically underlying factor in all of the measurement items associated with the same latent construct. Inadequate unidimensionality challenges reflective construct validity at its most fundamental level and results in parameter estimate biases (Anderson et al. 1987; Gefen 2003; Gerbing and Anderson 1988; Segars 1997). Researchers who design a scale to reflect multiple latent dimensions face special challenges. The simple proportionality test described previously is not directly applied, and standard tools like reliability indices also will not directly apply. Unidimensionality is assumed in linear regression and PLS, and can be tested for in CBSEM. See an example of how to do so and the consequences of ignoring it in the tutorial on unidimensionality by Gefen (2003). Like CBSEM, PLS path modeling as a method also has the ability to incorporate multidimensional constructs, although few researchers seem to exploit this opportunity.

Alternative Models

Another issue that has been around for a long time but is still not applied as rigorously as it should be is the comparison of the theoretical model with a saturated model. In a seminal paper, Anderson and Gerbing (1988) highlighted the need to compare alternative models, including the theoretical model containing the hypotheses with a saturated model containing paths among all pairs of latent variables, even those assumed to be unrelated in the model. In essence, this allows the researchers to verify that no significant path has been left out of the model. This is crucial because omitting significant predictors can bias other path estimates. In some cases, these omissions bias estimated paths leading to significance. While in CBSEM omitted paths may reveal themselves through lack of fit, with PLS and linear regression the consequences of these omissions may only be seen in careful diagnostics involving plots of residuals. It should be imperative, therefore, that authors include such analysis in their papers.

Possible additional checks related to alternative models include testing for moderating effects. In CBSEM this can be done using χ^2 nested model tests (Jöreskog and Sörbom 1994). Other tests include testing for a significant change in χ^2 when comparing the original model with an alternative nested model in which one pair of the latent variables are joined (Gefen et al. 2003). If the resulting difference in χ^2 is insignificant, then, at least statistically, there is no reason not to join the two latent constructs. If the model χ^2 fit index is not significantly improved by joining any two latent variables, then there is reason to believe that there is significant discriminant validity.

Sample Size and Power

Sample size has long been an important issue in PLS path modeling and in CBSEM, but regression has the most straightforward sample size guidance. Ordinary least squares, the most popular estimation method in regression, does not demand especially large sample sizes, so the choice of sample size is primarily a matter of statistical power (Cohen 1988). Statistical power in regression is a well-understood function of effect size (f^2), sample size, number of predictors and significance level. Cohen (1988) offers detailed guidance on how to use these variables to choose a minimum necessary sample size for regression users aiming to achieve a given level of statistical power. Across the social sciences, convention specifies 80 percent as the minimum acceptable power. Websites are available to calculate the minimum sample sizes needed to achieve adequate power under different conditions in linear regression (e.g., <http://www.danielsoper.com/statcalc/>).

With PLS and CBSEM, sample size plays a more complex role. PLS path modeling parameter estimates are biased, as noted previously, with the bias diminishing as both the number of indicators per construct and sample size increase. Researchers can calculate the expected degree of bias and determine the likely impact of investing in a larger sample size (Dijkstra 2010; McDonald 1996). It has been argued that PLS path modeling has advantages over OLS regression in terms of power (Chin et al. 2003) although this claim has been challenged (Goodhue et al. 2007). The core of the PLS estimation method—ordinary least squares—is remarkably stable even at low sample sizes. This gave rise to a

¹However, not all authorities endorse *ex post* techniques for dealing with common method variance (Richardson et al. 2009; Spector 2006).

rule of thumb specifying minimum sample size as 10 times the largest number of predictors for any dependent variable in the model (Barclay et al. 1995; Gefen et al. 2000). This is only a rule of thumb, however, and is not backed up with substantive research.

With CBSEM, the role of sample size is also complex, but in different ways. Statistical power itself takes on multiple meanings. On the one hand, there is power to reject the researcher's overall model. In contrast to regression, PLS path modeling, and most statistical methods in the social sciences, with CBSEM the model being proposed by the researcher is (typically, but not necessarily) equated with the null hypothesis, rather than being associated with the alternative hypothesis. In regression and PLS path modeling, however, power typically refers to the ability to reject a null hypothesis of no effect in favor of an alternative hypothesis which is identified with the researcher's proposed model. So, with CBSEM, increasing sample size means constructing a more stringent test for the model, while with regression and PLS path modeling, a large sample size may work in the researcher's favor. Power versus the overall model is related to the model degrees of freedom (MacCallum et al. 1996). A model with high degrees of freedom can obtain sufficient power at very low sample size. However, power can also be addressed in terms of individual path estimates, as in regression and PLS path modeling (Satorra and Saris 1985). In this setting, researchers may find a link between power and sample size that is more consistent with that observed in the other methods.

Additionally and entirely apart from the issue of power, CBSEM users must be concerned about stability of the estimation method. CBSEM long relied on maximum likelihood (ML) estimation, a method that achieves stable results only at larger sample sizes. Other estimation methods, designed to overcome ML shortcomings or violations of distributional assumptions, may require even larger sample sizes.

There is no guideline backed by substantial research for choosing sample size in CBSEM. At the absolute minimum, there should be more than one observation per free parameter in the model. Sample size should also be sufficient to obtain necessary statistical power. Experience argues for a practical minimum of 200 observations for a moderately complicated structural equation model with ML estimation. When other estimation methods, such as WLS, are used or when there are special features, such as mixture modeling and models with categorical latent variables, a larger sample may be needed.² Rules of thumb alone do not automatically tell us what the sample size for a given study should be.

Other Overall Issues of Statistical Rigor that Apply Also to SEM

Data Description

Researchers should be expected to share enough information with their readers about their empirical data that their readers can make informed judgments about the soundness of the choices made and conclusions reached by the researchers. Other researchers should find enough information in a research report to allow them to make comparisons to their own results, or to include the study as a data point in a meta-analysis. Inferior reporting limits contribution to knowledge and potential impact.

Certain reporting requirements are common across regression, PLS path modeling, and CBSEM. Sharing this basic summary information allows researchers to determine whether differences in results across studies may be due to differences at the univariate level. For example, restriction of range, which may be revealed in the standard deviations, can attenuate correlations in a particular sample (Cohen et al. 2003, p. 57). If the reader does not have the raw data available, then researchers have an even greater obligation to fully examine their data and report thoroughly. Researchers must demonstrate, for example, that data appear to be consistent with stated distributional assumptions. With regression and PLS path modeling, which both use ordinary least squares estimation, explicit distributional assumptions during estimation focus on error terms, not on the observed variables themselves, although violated assumptions such as multicollinearity will bias the estimation results. Still, skewed distributions can reduce correlations between variables, and this has direct implications for the results. Extreme violations of distributions assumptions or skewed data may suggest the need for transformation to achieve better fit, such as using a log transformation as is customary in linear regression in such cases (Neter et al. 1990), although obviously the results need to be interpreted accordingly because the slope of the relationship between Y and $\ln(X)$ is not the same as the slope of the relationship between Y and X.

²Muthén and Muthén (2002) recommend and illustrate the use of Monte Carlo simulations as a tool to help determine necessary sample size in a particular situation. Stability improves when path relationships are strong and when there are several indicators per construct. Researchers should use literature related to their particular choice of estimation method to support an argument regarding adequate sample size. For example, simulation research regarding the combination of polychoric correlation coefficients and weighted least squares estimation point to the need for sample sizes in the thousands (Yung and Bentler 1994). Even then, researchers must be alert for evidence of instability. This evidence may include unreasonable χ^2 values, parameter estimates, or estimated standard errors.

Researchers should always report means and standard deviations for all continuous observed variables, and proportions for all categorical variables, as a most basic requirement. Readers of studies that use CBSEM have additionally come to expect authors to provide the correlation or covariance matrix of the observed variables. In many cases, having these matrixes enables readers to literally reconstruct the analysis, allowing for fuller understanding; it also makes the report useful in the training of future researchers. These matrixes are also valuable when researchers use regression or PLS path modeling, although those methods rely more on the raw data. We strongly encourage researcher to add the correlation or covariance matrix as an appendix to their paper.

Missing Data

One issue too often ignored is proper handling of incomplete data. Partially missing observations are a fact of life. Unfortunately, some popular analytical programs still promote outdated procedures which at best are statistically inefficient and at worst bias results. When there are missing data and nothing can be done to rectify or simulate the missing data, we recommend the authors say so explicitly. Realizing that some methodologists do not approve of deleting data, researchers can choose and have chosen in the past listwise deletion. Discarding an entire case when only few observations are missing does not bias results under many circumstances, but it is wasteful of data as compared with currently available alternatives. The alternative pairwise deletion—estimating, say, each covariance based on all observations where both variables are present—conserves data but may, as in linear regression (Afifi et al. 2004), result in inaccurate estimations and introduce inconsistencies because of varying sample sizes within the data. Varying sample size across the model also complicates the analysis.

There may be cases, however, when there are not enough data to allow listwise deletion or there is a theoretical reason not to listwise delete these data points, such as when those data come from a unique category of respondents about whom these measurement items are irrelevant. In such cases, researchers can apply the EM algorithm with multiple imputation (Enders 2006), an approach which usually produces a complete data rectangle with minimal bias in the results. Standard statistical packages now include this approach, bringing it within reach of almost all researchers. Some statistical packages offer more sophisticated approaches such as the full information maximum likelihood approach, which is or will soon be standard in CBSEM packages. This approach models the missing data points along with the free parameters of the structural equation model. With this approach, estimation proceeds as if each missing observation is an additional free parameter to be estimated, although, unlike imputation, the method does not generate an actual set of replacement values. The modeling of missing values is specific to the larger model being estimated, so it would be inappropriate to establish specific replacement values as if they were model-independent. Techniques are available which can address missing data issues even in cases where missingness is tied to the phenomena under study under certain conditions (Enders 2011; Muthén et al. forthcoming).

Distribution Assumptions

With CBSEM, data distribution becomes an important factor in selecting the estimation method. The default maximum likelihood method assumes conditional multivariate normality (Muthén 1987), although the method is also robust against mild violations of this assumption (Bollen 1989). Therefore, researchers should report standard skewness and kurtosis measures for their observed variables. Substantial departures from conditional multivariate normality call for either transformation or the use of a different estimator.

Nonresponse Bias

Another topic related to missing data is nonresponse bias. This issue arises when more than a relatively slight proportion of the respondents did not return the survey or complete the experimental treatment. In these cases it cannot be ruled out that there is a selection bias among those who did complete the survey or treatment. There is nothing that can be done *ex post* to recreate these missing data within the data itself, but steps should be taken to verify at least at a minimal level that the nonresponders are not a unique group.³ This can be done partially by comparing the demographics of respondents and nonrespondents and by comparing the answers of waves of responders; presumably, if these are no significant differences among these groupings, then there is less reason for concern (Armstrong and Overton 1977).

³Although other steps can be taken such as follow up phone interviews to see if there is a difference between the respondents and nonrespondents. Also see Sivo et al. (2006).

Population: Appropriate and Homogeneous

As with any research project involving collected data, researchers must carefully describe and justify their choices of population, sampling frame, and sampling plan. Researchers must support the appropriateness of the population and address concerns about possible heterogeneity. Heterogeneity means that different members of the population are subject to either different models or different parameter values within the same model structure. In such cases, parameters estimated across the aggregated population are at best overall averages, which may not accurately represent even one single individual respondent, much less all of them. In some cases, the sampling frame itself presents a risk of heterogeneity, as when a researcher aggregates respondents across nations or cultures in order to reach a sufficiently large total sample size.

Currently, linear regression, PLS, and CBSEM (Rigdon et al. 2010) all offer tools for comparing across populations and for seeking multiple latent populations within a single data set. Researchers should evaluate their populations for the potential for heterogeneity. Where heterogeneity represents a plausible confound, researchers should investigate and report their comparative results.

References

- Affi, A. A., Clark, V., and May, S. 2004. *Computer-Aided Multivariate Analysis, Volume 62*, Boca Raton, FL: CRC Press.
- Anderson, J. C., and Gerbing, D. W. 1988. "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin* (103:3), pp. 411-423.
- Anderson, J. C., Gerbing, D. W., and Hunter, J. E. 1987. "On the Assessment of Unidimensional Measurement: Internal and External Consistency, and Overall Consistency Criteria," *Journal of Marketing Research* (24), pp. 432-437.
- Armstrong, J. S., and Overton, T. S. 1977. "Estimating Non-Response Bias in Mail Surveys," *Journal of Marketing Research* (14), pp. 396-402.
- Barclay, D., Thompson, R., and Higgins, C. 1995. "The Partial Least Squares (PLS) Approach to Causal Modeling: Personal Computer Adoption and Use an Illustration," *Technology Studies* (2:2), pp. 285-309.
- Bollen, K. A. 1989. *Structural Equations with Latent Variables*, New York: Wiley-Interscience
- Chin, W. W., Marcolin, B. L., and Newsted, P. R. 2003. "A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study," *Information Systems Research* (14:2), pp. 189-217.
- Cohen, J. B. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J. B., Cohen, P., West, S. G., and Aiken, L. S. 2003. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.), Mahwah, NJ: Lawrence Erlbaum Associates.
- Diamantopoulos, A., and Winklhofer, H. M. 2001. "Index Construction with Formative Indicators: An Alternative to Scale Development," *Journal of Marketing Research* (38:2), pp. 269-277.
- Dijkstra, T. K. 2010. "Latent Variables and Indices: Herman Wold's Basic Design and Partial Least Squares," in *Handbook of Partial Least Squares*, V. E. Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Berlin, Germany: Springer-Verlag, pp. 23-46.
- Enders, C. K. 2006. "Analyzing Structural Equation Models with Missing Data," in *Structural Equation Modeling: A Second Course*, G. R. Hancock and R. O. Mueller (eds.), Charlotte, NC: Information Age, pp. 315-344.
- Enders, C. K. 2011. "Missing Not at Random Models for Latent Growth Curve Analysis," *Psychological Methods* (16:1), pp. 1-16.
- Gefen, D. 2003. "Unidimensional Validity: An Explanation and Example," *Communications of the Association for Information Systems* (12:2), pp. 23-47.
- Gefen, D., Karahanna, E., and Straub, D. W. 2003. "Trust and TAM in Online Shopping: An Integrated Model," *MIS Quarterly* (27:1), pp. 51-90.
- Gefen, D., Straub, D. W., and Boudreau, M.-C. 2000. "Structural Equation Modeling and Regression: Guidelines for Research Practice," *Communications of the Association for Information Systems* (4:7), pp. 1-70.
- Gerbing, D. W., and Anderson, J. C. 1988. "An Updated Paradigm for Scale Development Incorporating Unidimensionality and its Assessment," *Journal of Marketing Research* (25:May), pp. 186-192.
- Goodhue, D., Lewis, W., and Thompson, R. 2007. "Statistical Power in Analyzing Interaction Effects: Questioning the Advantage of PLS with Product Indicators," *Information Systems Research* (18:2), pp. 211-227.
- Jöreskog, K. G., and Sörbom, D. 1994. *LISREL 8.12i and PRELIS 2.12i for Windows*, Chicago: Scientific Software International.
- Kumar, N., Stern, L. W., and Anderson, J. C. 1993. "Conducting Interorganizational Research Using Key Informants," *Academy of Management Journal* (36:6), pp. 1633-1651.
- MacCallum, R. C., Browne, M. W., and Sugawara, H. M. 1996. "Power Analysis and Determination of Sample Size in Covariance Structure Analysis," *Psychological Methods* (1:2), pp. 130-149.

- Malhotra, N. K., Kim, S. S., and Patil, A. 2006. "Common Method Variance in IS Research: A Comparison of Alternative Approaches and a Reanalysis of Past Research," *Management Science* (52:12), pp. 1865-1883.
- McDonald, R. P. 1996. "Path Analysis with Composite Variables," *Multivariate Behavioral Research* (31:2), pp. 39-70.
- Muthén, B. O. 1987. *LISCOMP: Analysis of Linear Structural Relations with a Comprehensive Measurement Model* (2nd ed.), Mooresville, IN: Scientific Software.
- Muthén, B. O., Asparouhov, T., Hunter, A., and Leuchter, A. forthcoming. "Growth Modeling with Non-Ignorable Dropout: Alternative Analyses of the STAR*D Antidepressant Trial," *Psychological Methods* (<http://www.statmodel.com/download/FebruaryVersionWide.pdf>).
- Muthén, L. K., and Muthén, B. O. 2002. "How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power," *Structural Equation Modeling* (9:4), pp. 599-620.
- Neter, J., Wasserman, W., and Kutner, M. H. 1990. *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Design* (3rd ed.), New York: McGraw-Hill, Inc.
- Podsakoff, P. M., Lee, J. Y., and Podsakoff, N. P. 2003. "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of Applied Psychology* (88:5), pp. 879-903.
- Richardson, H. A., Simmering, M. J., and Sturman, M. C. 2009. "A Tale of Three Perspectives: Examining Post Hoc Statistical Techniques for Detection and Correction of Common Method Bias," *Organizational Research Methods* (12:4), pp. 762-800.
- Rigdon, E. E., Ringle, C. M., and Sarstedt, M. 2010. "Structural Modeling of Heterogeneous Data with Partial Least Squares," *Review of Marketing Research* (7), pp. 257-298.
- Satorra, A., and Saris, W. E. 1985. "The Power of the Likelihood Ratio Test in Covariance Structure Analysis," *Psychometrika* (50), pp. 83-90.
- Segars, A. H. 1997. "Assessing the Unidimensionality of Measurement: A Paradigm and Illustration Within the Context of Information Systems Research," *Omega* (25:1), pp. 107-121.
- Sivo, S. A., Saunders, C., Chang, Q., and Jiang, J. J. 2006. "How Low Should You Go? Low Response Rates and the Validity of Inference in IS Questionnaire Research," *Journal of the AIS* (7:6), pp. 361-414.
- Spector, P. E. 2006. "Method Bias in Organizational Research," *Organizational Research Methods* (9:2), pp. 221-232.
- Yung, Y.-F., and Bentler, P. M. 1994. "Bootstrap-Corrected ADF Test Statistics in Covariance Structure Analysis," *British Journal of Mathematical and Statistical Psychology* (47), pp. 63-84.