LOWER BOUNDS ON SAMPLE SIZE IN STRUCTURAL EQUATION MODELING

Published in *Electronic Commerce Research and Applications*, forthcoming Dec 2010, PII:S1567-4223(10)00054-2, DOI: 10.1016/j.elerap.2010.07.003 (with software downloadable from Elsevier)

J. Christopher Westland

Professor, Information & Decision Sciences

University of Illinois, Chicago

601 S. Morgan Street, Chicago, IL 60607-7124

(312) 860-0587 email: westland@uic.edu

JULY 2010

ABSTRACT

LOWER BOUNDS ON SAMPLE SIZE IN STRUCTURAL EQUATION MODELING

Computationally intensive structural equation modeling (SEM) approaches have been in development over much of the 20th century, initiated by the seminal work of Sewall Wright. To this day, sample size requirements remain a vexing question in SEM based studies. Complexities which increase information demands in structural model estimation increase with the number of potential combinations of latent variables; while the information supplied for estimation increases with the number of measured parameters times the number of observations in the sample size – both are non-linear. This alone would imply that requisite sample size is *not* a linear function solely of indicator count, even though such heuristics are widely invoked in justifying SEM sample size. This paper develops two lower bounds on sample size in SEM, the first as a function of the ratio of indicator variables to latent variables, and the second as a function of minimum effect, power and significance. The algorithm is applied to a meta-study of a set of research published in five of the top MIS journals. The study shows a systematic bias towards choosing sample sizes that are significantly too small. Actual sample sizes averaged only 50% of the minimum needed to draw the conclusions the studies claimed. Overall, 80% of the research articles in the meta-study drew conclusions from insufficient samples. Lacking accurate sample size information, researchers are inclined to economize on sample collection with inadequate samples that hurt the credibility of research conclusions. Guidelines are provided for applying the algorithms developed in this study, and companion software encapsulating the paper's formulae is made available for download. *(261 words)*

Keywords: Structural equation modeling, SEM, Partial least squares, PLS, LISREL, AMOS, sample size, Gini correlation, common factor bias, rule of 10

1. Introduction

The past two decades have seen a remarkable acceleration of interest in structural equations modeling (SEM) methods in management research, including partial least squares (PLS) and implementations of Jöreskog's SEM algorithms (LISREL, AMOS, EQS). The breadth of application of SEM methods has been expanding, with SEM increasingly applied to exploratory, confirmatory and predictive analysis with a variety of *ad hoc* topics and models. SEM is particularly useful in the social sciences where many if not most key concepts are not directly observable. Because many key concepts in the social sciences are inherently latent, questions of construct validity and methodological soundness take on a particular urgency.

To this day, methodologies for assessing suitable sample size requirements remain a vexing question in SEM based studies. The number of degrees of freedom consuming information in structural model estimation increases with the number of potential combinations of latent variables; while the information supplied in estimating increases with the number of measured parameters (i.e., indicators) times the number of observations (i.e., the sample size) – both are non-linear in model parameters. This should imply that requisite sample size is *not* a linear function solely of indicator count, even though such heuristics are widely invoked in justifying SEM sample size. Monte Carlo simulation in this field has lent support to the non-linearity of sample size requirements, though research to date has not yielded a sample size formula suitable for SEM. This paper proposes a set of necessary conditions (thus lower bounds) for SEM sample adequacy.

The exposition proceeds as follows. Section 2 describes the historical context, commenting on how particular research objectives and computational limitations resulted in our current SEM toolsets. Section 3 summarizes the prior literature on sample adequacy results from Monte Carlo simulations. Section 4 develops an algorithm for computing the minimum sample size needed to detect a minimum effect at given power and significance levels in the structural equation model. Section 5 discusses these with an application research articles whose conclusions rest on confirmatory SEM analyses, and assesses whether the sample sizes used are adequate.

2. Prior Literature

SEM evolved in three different streams: (1) systems of equation regression methods developed mainly at the Cowles Commission; (2) iterative maximum likelihood algorithms for path analysis developed mainly at the University of Uppsala; and (3) iterative least squares fit algorithms for path analysis also developed at the University of Uppsala. Figure 1 provides a chronology of the pivotal developments in latent variable statistics in terms of method (pre-computer, computer intensive and SEM) and objectives (exploratory / prediction or confirmation).

INSERT FIGURE 1: DEVELOPMENT OF STRUCTURAL EQUATION MODEL ESTIMATION

Both LISREL and PLS were conceived as iterative computer algorithms, with an emphasis from the start on creating an accessible graphical and data entry interface and extension of Wright's (1921) path analysis. Early Cowles' Commission work on simultaneous equations estimation centered on Koopman and Hood's (1953) algorithms from the economics of transportation and optimal routing, with maximum likelihood estimation, and closed form algebraic calculations, as iterative solution search techniques were limited in the days before computers. Anderson and Rubin (1949, 1950) developed the limited information maximum likelihood estimator for the parameters of a single structural equation, which indirectly included the two-stage least squares estimator and its asymptotic distribution (Anderson, 2005) and Farebrother (1999). Two-stage least squares was originally proposed as a method of estimating the parameters of a single structural equation in a system of linear simultaneous equations, being introduced by Theil (1953a, 1953b, 1961) and more or less independently by Basmann (1957) and Sargan (1958). Anderson's limited information maximum likelihood estimation was eventually implemented in a computer search algorithm, where it competed with other iterative SEM algorithms. Of these, two-stage least squares was by far the most widely used method in the 1960s and the early 1970s.

LISREL and PLS path modeling approaches were championed at Cowles mainly by Nobelist Trygve Haavelmo (1943). Unfortunately underlying assumptions of LISREL and PLS were challenged by economists such as Freedman (1987) who objected to their "failure to distinguish among causal assumptions, statistical implications, and policy claims has been one of the main reasons for the suspicion and confusion surrounding quantitative methods in the social sciences" (see also Wold's (1987) response). Haavelmo's path analysis never gained a large following among U.S. econometricians, but was successful in influencing a generation of Haavelmo's fellow Scandinavian statisticians, including Hermann Wold, Karl Jöreskog , and Claes Fornell. Fornell introduced LISREL and PLS techniques to many of his Michigan colleagues through influential papers in accounting (Fornell and Larker 1981), and information systems (Davis, et al, 1989). Dhrymes (1971; Dhrymes, et al. 1974) provided evidence that PLS estimates asymptotically approached those of two-stage least squares with exactly identified equations. This point is more of academic importance than practical, because most empirical studies overidentify. But in one sense, all of the limited information methods (OLS excluded) yield similar results.

3. Sample Size and the Ratio of Indicators to Latent Variables

Structural equation modeling in MIS has taken a casual attitude towards choice of sample size. Since the early 1990s, MIS researchers have alluded to an *ad hoc* rule of thumb requiring the choosing of 10 observations per indicator in setting a lower bound for the adequacy of sample sizes. Justifications for this *rule of 10* appear in several frequently cited publications (Barclay, et al. 1995; Chin 1998; Chin, and Newsted 1999; Kahai and Cooper 2003) though none of these researchers refers to the original articulation of the rule by Nunnally (1967) who suggested (without providing supporting evidence) that in SEM estimation 'a good rule is to have at least ten times as many subjects as variables.'

Within the MIS field, Goodhue, et al. (2006, 2007) studied the *rule of 10* using Monte Carlo simulation to compare sample sizes of 40, 90, 150, and 200, along with varying effect sizes (large, medium, small and no effect) to determine the adequacy of this rule for a given significance and power of tests. They concluded that: *"In fact, for simple [SEM] models with normally distributed data and relatively reliable measures, none of the three techniques have adequate power to detect small or medium effects at small sample sizes … These findings run counter to extant suggestions in MIS literature"* (Goodhue, et al. 2006, p. 202b). This finding is not completely unexpected, as similar SEM rules of thumb have been investigated since Nunnally's (1967) proposal. The debate has evolved significantly since his 1967 publication.

The *rule of 10* couches the sample size question in terms of the ratio of observations (sample points) to free parameters – for example, Bollen (1989) stated that "though I know of no hard and fast rule, a useful suggestion is to have at least several cases per free parameter" and Bentler (1989) suggested a 5:1 ratio of sample size to number of free parameters. But is this the right question? Typically their parameters were considered to be indicator variables in the model, but unlike the early path analysis, structural equation models today are typically estimated in their entirety, and the number of unique entries in the covariance matrix is $\frac{p(p+1)}{2}$ when p is the number of indicators. It would be reasonable to assume that the sample size is proportional to $\frac{p(p+1)}{2}$ rather than p. Unfortunately, Monte Carlo studies conducted in the 1980s and 1990s showed that the problem is somewhat more subtle and complex than that, and sample size and estimator performance are generally uncorrelated with either $\frac{p(p+1)}{2}$ or p.

Difficulties arise because the p indicator variables are used to estimate the k latent variables (the unobserved variables of interest) in the SEM, and even though there may be $\frac{p(p+1)}{2}$ free parameters, these are not individually the focus of SEM estimation. Rather, free parameters are clustered around a much smaller set of latent variables which are the focus of the estimation (or alternatively, the correlations between these unobserved latent variables are the focus of estimation). Tanaka (1987) argued that sample size should be dependent on the number of estimated parameters (the latent variables and their correlations) rather than on the total number of indicators; a view mirrored in other discussions of minimum sample sizes (Browne and Cudeck 1989, 1993; Geweke and Singleton 1980; Gebring and Anderson 1985). Veicer and Fava (1987, 1989, 1994) went further, after reviewing a variety of such recommendations in the literature, concluding that there was no support for rules positing a minimum sample size as a function of indicators. They showed that for a given sample size, a convergence to proper solutions and goodness of fit were favorably influenced by: (1) a greater number of indicators per latent variable; and (2) a greater saturation (higher factor loadings).

Marsh and Bailey (1991) concluded that the *ratio of indicators to latent variables* rather than just the number of indicators, as suggested by the rule of 10, is a substantially better basis on which to calculate sample size, reiterating conclusions reached by Boomsma (1982) who suggested using a ratio $r = {}^{p}/_{k}$ of indicators to latent variables . Information input to the SEM estimation increases both with more indicators per latent variable, as well as with more sample observations. A series of studies (Ding, et al. 1995) found that the probability of rejecting true models at a significance level of 5% was close to 5% for $r = 2$ (where r is the ratio of indicators to latent variables) but rose steadily as r increased – for $r = 6$, rejection rates were 39% for sample size of 50; 22% for sample size of 100; 12% for sample size of 200; and 6% for sample size of 400.

Boomsma's (1982) simulations suggested that a ratio r of indicators to latent variables of $r = 4$ would require a sample size of at least 100 for adequate analysis; and for $r = 2$ would require a sample size of at least 400. Marsh et al (1988,1996, 1998)) ran 35,000 Monte Carlo simulations on LISREL CFA analysis, yielding data that suggested that: $r = 3$ would require a sample size of at least 200; $r = 2$ would require a sample size of at least 400; $r = 12$ would require a sample size of at least 50. Consolidation and summarization of these results suggest sample sizes:

$$n \geq 50r^2 - 450r + 1100$$

where r is the ratio of indicators to latent variables. Furthermore, Marsh et al. (1996) recommend $r = 6$ to 10 indicators per latent variable, assuming 25-50% of the initial choices add no explanatory power, which they found to often be the case in their studies. They note that this is a substantially larger ratio than found in most SEM studies, which tend to limit themselves to 3-4 indicators per latent variable. It is possible that a sample size rule of ten observations per indicator may indeed bias researchers towards selecting smaller numbers of indicators per latent variable in order to control the cost of a study or the length of a survey instrument.

4. Sample Size with Paired Latent Variables

This section develops an algorithm for computing the lower bound on sample size required to confirm or reject the existence of a minimum effect in an SEM at given significance and power levels. Where SEM studies are directed towards hypothesis testing for complex models, with some level of significance α and power $1 - \beta$, calculating the power requires first specifying the effect size δ you want to detect. Funding agencies, ethics boards and research review panels frequently request that a researcher perform a power analysis, the argument is that if a study is inadequately powered, there is no point in completing the research. Additionally, in the framework of SEM the assessment of power is affected by the variable information contained in social science data. Table 1 summarizes the notation used.

INSERT TABLE 1: NOTATION USED IN THE PAPER

DECONSTRUCTION

This research asks "What is the *lower bound on sample size* n for confirmatory testing of SEM as a function of these design parameters?" We want to detect a minimum correlation (effect) δ in estimating k latent (unobserved) variables, at significance and power levels $(\alpha^*, 1-\beta)$. In other words, devise an algorithm $f(\cdot)$ $such\ that\ n = f[k, \delta|\alpha^*, \beta]$.

In this section, we will adopt the standard targets for our required Type I and II errors under Neyman-Pearson hypothesis testing of $\alpha^* = .05$ and $\beta = .20$; but these requirements can be relaxed for a more general solution. Structural equation models are characterized here as a collection of pairs of canonically correlated latent variables, and adhere to the standard normalcy assumption on indicator variables. This leads naturally to a deconstruction of the SEM into an overlapping set of bivariate normal distributions. Make the assumption that an arbitrarily selected pair of latent variables, call them $\tilde{X}$ and $\tilde{Y}$, are bivariate normal with density function:

$$f(x, y|\mu_x, \mu_y, \rho, \sigma_x, \sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho}} exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right)\right]$$

and covariance structure is $\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$

COMBINATORICS OF HYPOTHESIS TESTS ON LINKS, AND SIGNIFICANCE LEVEL

It is typical in the literature to predicate an SEM analysis with the caveat that one needs to make strong arguments for the complex models constructed from the unobserved, latent constructs tested with the particular SEM, in order to support the particular links that are included in the model. This is usually interpreted to mean that each proposed (and tested) link in the SEM needs to be supported with references to prior research, anecdotal evidence and so forth. This may simply mean the wholesale import of a preexisting model (e.g., the Technology Acceptance Model) based on the success of that model in other contexts, but not specifically building on the particular effects under investigation. Unfortunately, it is uncommon to see any discussion of the particular links (causal or otherwise) or combinations of links that are *excluded* (either implicitly or explicitly) from the SEM model. Ideally, there should also be similarly strong arguments made for the inapplicability of omitted links or omitted combinations of links.

We can formalize these observations by letting i be the number of the potential links between latent variables. Extend the individual link minimum sample size to a minimum sample size for the entire SEM; building up from pairs of latent variables by determining the number of possible combinations of the i pairs, each with an 'effect' that needs detection. Each effect can be dichotomized:

$$link_i = \begin{cases} 0: & \rho_i < \delta \\ 1: & \rho_i \geq \delta \end{cases}$$

Our problem is to compute the number of distinct structural equation models that can exist in terms of the 0,1 values of their links using combinatorial analysis.

INSERT FIGURE 2: AN EXAMPLE OF A STRUCTURAL EQUATION MODEL WITH SIX LATENT VARIABLES AND FIVE CORRELATIONS

INSERT FIGURE 3: THE SEM EXAMPLE IN FIGURE 2 WITH ALL POSSIBLE PAIRED LINKS SHOWN

Then each combination of $\{0,1\}$ values for links which our tests of the SEM on the whole requires us to discriminate amongst provides us a set of $\frac{k(k-1)}{2}$ binary numbers (see figures 2 and 3) each representing a unique combination of latent variables. The unique model hypothesized in any particular study will be some model (binary number) which is exactly one out of the possible $2^{k(k-1)/2}$ ways of connecting these latent variables; testing must discriminate this path from the possible $2^{k(k-1)/2} - 1$ other paths which collectively define the alternative hypothesis.

For hypothesis testing with a significance of α^* (which we have by default set to $\alpha^* = .05$) on each link, it is necessary to correct for effective significance level α in differentiating one possible model from all other hypothesized structural equation models that are possible. The Šidàk correction is a commonly used alternative for the Bonferroni correction where an experimenter is testing a set of hypotheses with a dataset controlling the family-wise error rate. In the context of the current research the Šidàk correction provides the most accurate results. For the following analysis, a Šidàk correction gives $\alpha = \alpha(k) = 1 - (1-\alpha^*)^{2/k(k-1)}$ where the power of the test can be held at $1 - \beta = .8$ over the entire SEM with no modification.

MINIMUM EFFECT SIZE δ

Minimum effect, in the context of structural equation models, is the smallest correlation between latent variables that we wish to be able to detect with our sample and model. Small effects are more difficult to detect than large effects as they require more information to be collected. Information may be added to the analysis by collecting more sample observations, by adding parameters, and by constructing a better model.

INSERT FIGURE 4: SIGNIFICANCE AND POWER FOR THE MINIMUM EFFECT THAT NEEDS TO BE DETECTED

Sample size for hypothesis testing is typically determined from a *critical value* (see Figure 4) that defines the boundary between the rejection (set by α) and non rejection (set by β) regions. The minimum sample size that can differentiate between H_0 and H_A occurs where the *critical value that is exactly the same* under the null and alternative hypotheses. The approach to computing sample size here is analogous to standard univariate calculations (Cochran 1977; Kish 1955; Lohr 1999; Snedecor and Cochran 1989, Westland and See-to 2007) but using a formulation for variance customized to this problem.

In the context of structural equation models, canonical correlation between latent variables should be seen simply as correlation, the 'canonical' qualifier referring to the particulars of its calculation in SEM since the latent variables are unobserved, and thus cannot be directly measured. Correlation is interpreted as the strength of statistical relationship between two random variables obeying a joint probability distribution (Kendall and Gibbons 1990) like a bivariate normal. Several methods exist to compute correlation: the Pearson's product moment correlation coefficient (Fisher 1921, 1990), Spearman's rho and Kendall's tau (Kendall and Gibbons 1990) are perhaps the most widely used (Mari and Kotz 2001). Besides these three classical correlation coefficients, various estimators based on M-estimation (Shevlyakov and Vilchevski 2002) and order statistics (Schechtman and Yitzhaki 1987) have been proposed in the literature. Strengths and weaknesses of various correlation coefficients must be considered in decision making. The Pearson coefficient, which utilizes all the information contained in the variates, is optimal when measuring the correlation between bivariate normal variables (Stuart and Ord 1991). However, it can perform poorly when the data is attenuated by nonlinear transformations. The two rank correlation coefficients, Spearman's rho and Kendall's tau, are not as efficient as the Pearson correlation under the bivariate normal model; nevertheless they are invariant under increasing monotone transformations, thus often considered as robust alternatives to the Pearson coefficient when the data deviates from bivariate normal model. Despite their robustness and stability in non-normal cases, the M-estimator-based correlation coefficients suffer great losses (up to 63% according to Xu, et al. 2010) of *asymptotic relative efficiency* to the Pearson coefficient for normal samples, though such heavy loss of efficiency might not be compensated by their robustness in practice. Schechtman and Yitzhaki (1987) proposed a correlation coefficient based on order statistics for the bivariate distribution which they call Gini correlation (because it is related to *Gini's mean difference* in a way that is similar to the relationship between Pearson correlation coefficient and the variance).

INSERT FIGURE 5: BIVARIATE NORMAL SCATTERPLOTS FOR $\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}$ AND $\Sigma=\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}$ WITH $n=500$

As a measure of such strength, correlation should be large and positive if there is a high probability that large or small values of one variable occur (respectively) in conjunction with large of small values of another; and it should be large and negative if the direction is reversed (Gibbons and Chakraborti 1992). Figure 5 provides a rugplot of bivariate normal scatterplots generated by the R *mtvnorm* package that provide a visual description of the clustering and behavior of particular values of correlation ρ between the latent variables.

We will use a standard definition of minimum effect size to be detected – the strength of the relationship between two variables in a statistical population as measured by the correlation ρ for paired latent variables – following conventions articulated in Wilkinson (1999); Nakagawa et al. (2007) and Brand, et al. (2008). Where we are assessing completed research, we can substitute for δ the smallest correlation (effect size) on all of the links between latent variables in the SEM. Cohen (1988, 1992) provides the following guidelines for the social sciences: small effect size, $|\rho|$ = 0.1-.23; medium, $|\rho|$ = 0.24-.36; large, $|\rho|$ = 0.37 or larger. Figure 5 gives us a feel for Cohen's recommendations – $|\rho|$ = 0.37 still has a great deal of dispersion, and we might find it difficult to visually determine correlation merely by looking at a scatterplot where the variables on the two axes have correlation $|\rho|$ = 0.37 .

ESTIMATOR FOR CORRELATION IN A BIVARIATE NORMAL DISTRIBUTION

Let (X_i, Y_i) $i = 1,2, \ldots, n$ be a random sample of independent and identically distributed (i.i.d.) data pairs of size n from the bivariate normal population of (X, Y) population with continuous joint cumulative distribution function. Let $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ be the order statistics (where the first subscript is the rank, and the second the sample size) of the X_i sample values; let $Y_{1:n} \le Y_{2:n} \le \cdots \le Y_{n:n}$ be the order statistics of the Y_i sample values; and let $Y_{[i:n]}$ be the Y sample value associated with the $X_{i:n}$ sample value in the sample pairs (X_i, Y_i). $Y_{[i:n]}$ is called the concomitant of the i^{th} order statistic (Balakrishnan and Rao 1998*).* Reversing the roles of X and Y , we can also obtain the associated $X_{[i:n]}$. Extending the work of Schechtman and Yitzhaki (1987), Xu, et al. (2010) show that the two Gini correlations with respect to (X_i, Y_i) are

$$\hat{\rho}_{G_{XY}}(X,Y) = \frac{\frac{1}{n(n-1)}\sum_i^n (2i-n-1) X_{[i:n]}}{\frac{1}{n(n-1)}\sum_i^n (2i-n-1) X_{i:n}}$$

and

$$\hat{\rho}_{G_{YX}}(Y,X) = \frac{\frac{1}{n(n-1)}\sum_i^n (2i-n-1)Y_{[i:n]}}{\frac{1}{n(n-1)}\sum_i^n (2i-n-1)Y_{i:n}}$$

In general $\hat{\rho}_{G_{XY}}(X,Y)$ is not symmetric – that is, $\hat{\rho}_{G_{XY}}(X,Y) \neq \hat{\rho}_{G_{YX}}(Y,X)$. Such asymmetry violates the axioms of correlation measurement (Gibbons and Chakraborti 1992; Mari and Kotz 2001) which is assumed in SEM estimation. Xu, *et al*. (2010) provide a symmetrical estimator (which we use here) obtained from their linear combination:

$$\hat{\rho}_G(Y,X) = \tfrac{1}{2}\left[\hat{\rho}_{G_{YX}}(Y,X) + \hat{\rho}_{G_{XY}}(X,Y)\right]$$

Gini correlation $\hat{\rho}_G$ possesses the following general properties (Schechtman and Yitzhaki 1987):

1) $\hat{\rho}_G \in [-1,1]$

2) $\hat{\rho}_G(Y,X) = \hat{\rho}_G(Y,X) = \pm 1$ if Y is a monotone increasing (decreasing) function of X

3) $\hat{\rho}_G(Y,X)$ is asymptotically unbiased and the expectations of $\hat{\rho}_G(Y,X)$ and $\hat{\rho}_G(X,Y)$ are zero when Y is independent of X

4) $\hat{\rho}_G(+,+) = -\hat{\rho}_G(-,+) = -\hat{\rho}_G(+,-) = \hat{\rho}_G(-,-)$ for both $\hat{\rho}_G(Y,X)$ and $= \hat{\rho}_G(X,Y)$

5) $\hat{\rho}_G(Y,X)$ is invariant under all strictly monotone transformations of X

6) $\hat{\rho}_G(Y,X)$ is scale and shift invariant with respect to both $X\ and\ Y$ and

7) $\sqrt{n}\,(\hat{\rho}_G - \rho) \overset{\mathfrak{D}}{\rightarrow} \mathbb{N}(0,\sigma_G^2)$; i.e., converges in distribution to a normal distribution with mean zero and variance σ_G^2 (This is from Schechtman and Yitzhaki (1987) applying methods developed by Hoeffding (1948))

8) The Spearman rho measure of correlation is a special case of $\hat{\rho}_G\,(Y,X)$; Xu, et al (2010).

Xu, et al. (2010) showed that Gini correlations are asymptotically normal with the following mean and variance[1]:

$\hat{\mu}_G = \mu(\hat{\rho}_G) =$

$$\rho - \frac{2(n-2)}{n(n-1)}\left(arcsin\left({}^{\rho}/_{2}\right) + \rho\sqrt{4-\rho^2} - \rho\sqrt{3}\right) +$$

$$\frac{\pi}{3}\frac{\rho(n+1)}{n(n-1)} + \frac{2}{n(n-1)}\left(arcsin(\rho) + \rho\sqrt{1-\rho^2}\right) + o(n^{-1})$$

$\hat{\sigma}_G^2 = \sigma^2(\hat{\rho}_G) =$

$$\frac{(1-\rho^2)}{n(n-1)}\left(\sqrt{1-\rho^2} - \rho\ arcsin(\rho) + \frac{\pi(n+1)}{6}\right) +$$

$$\frac{(n-2)(1-\rho^2)}{n(n-1)}\left(\frac{(1-\rho^2)}{\sqrt{4-\rho^2}} - \rho\ arcsin\left({}^{\rho}/_{2}\right)\right) + o(n^{-1})$$

Xu, et al. (2010) used Monte Carlo simulations to verify these formulas' asymptotic results (using *asymptotic relative efficiency* and *root mean square error* performance metrics) showing that they are applicable for data of even relatively small sample sizes (down to around 30 sample points). Their simulations confirmed and extend Hea and Nagarajab's (2009) Monte Carlo simulations exploring the behavior of nine distinct correlation estimators of the bivariate normal correlation coefficient, including the estimator $\hat{\rho}_G$, the sample correlation for the bivariate normal, and estimators based on order statistics. The estimator $\hat{\rho}_G$ was found generally to reduce bias and improve efficiency as well or better than other correlation estimators in the study. Xu, et al. (2010) also compared $\hat{\rho}_G$ with three other closely related correlation coefficients: (1) classical Pearson's product moment correlation coefficient, (2) Spearman's rho, and (3) order statistics correlation coefficients. Gini correlation bridges the gap between the order statistics correlation coefficient and Spearman's rho, and its estimators are more mathematically tractable than Spearman's rho, whose variance involves complex elliptic integrals that cannot be expressed in elementary functions. Their efficiency analysis showed that estimator $\hat{\rho}_G$'s loss of efficiency is between 4.5% to 11.3%, much less than that of Spearman's rho which ranges from 8.8% to 30.5%.

Calculation of Sample Size on a Single Link

[1] $o(n^{-1})$ convergence implies that for the remaining terms $v(n)$ go to zero faster than n^{-1}; $nv(n) \xrightarrow[n\to\infty]{} 0$

Construct a hypothesis test to just detect the minimum effect size δ:

$H_0\colon\ \rho - \rho_0 = 0$

$H_A\colon\ \rho - \rho_0 = \delta$

The one-sample, two-sided formulation (see Figure 4) that reconciles the null and alternative hypothesis tests for the estimator $\hat{\rho}_G \equiv \hat{\rho}_G(n)$ is

$$0 + z_{1-\alpha/2}\hat{\sigma}_G(n) = \delta + z_{1-\beta}\hat{\sigma}_G(n)$$

Xu, et al. (2010) show that $|\hat{\mu}_G - \rho| \xrightarrow[n\to\infty]{} 0$ quickly: for $n > 30$ from a bivariate normal population they show that we can assume $|\hat{\mu}_G - \rho| = 0$. Similarly, for $n > 30$ we can assume that $z - values$ are adequate approximations for $t - values$ in the formula. Even under the very weak assumptions of the 'rule of 10' a sample of $n = 30$ implies a model of at most three variables – significantly simpler than the majority of published models. Rearranging to place all terms with n on the left hand side:

$$\hat{\sigma}_G^2(n) = \left(\frac{\delta}{z_{1-\alpha/2} - z_{1-\beta}}\right)^2 \equiv H$$

Thus to within little $o(n^{-1})$ and using the formula for $\hat{\sigma}_G^2$

$$H \cong f(n,\rho) = \frac{(1-\rho^2)}{n(n-1)}\left(\sqrt{1-\rho^2} - \rho\, arcsin(\rho) + \frac{\pi(n+1)}{6}\right) + \frac{(n-2)(1-\rho^2)}{n(n-1)}\left(\frac{(1-\rho^2)}{\sqrt{4-\rho^2}} - \rho\, arcsin(\rho/2)\right)$$

We want to restate this as some function that calculates sample size $n = g(H,\rho)$. Solve for n by simplifying in terms of:

$A = 1 - \rho^2$

$B = \rho\, arcsin(\rho/2)$

$C = \rho\, arcsin(\rho)$

$$D = \frac{A}{\sqrt{3-A}}$$

$$H = \left(\frac{\delta}{z_{1-\alpha/2} - z_{1-\beta}}\right)^2$$

Then $n = \frac{-E \pm \sqrt{E^2 - 4F}}{2}$ are the solutions for the quadratic equation that restates $H - f(n,\rho) = 0$:

$$n^2 - \frac{A\left(\frac{\pi}{6} - B + D\right) + H}{H} n - \frac{A\left(\frac{\pi}{6} + \sqrt{A} + 2B - C - 2D\right)}{H} = n^2 + En + F = 0$$

Or in terms of $A, B, C, D\ and\ H$ and taking the largest root

$$n = \frac{1}{2H}\left(A\left(\frac{\pi}{6} - B + D\right) + H + \sqrt{\left[A\left(\frac{\pi}{6} - B + D\right) + H\right]^2 + 4\,AH\left(\frac{\pi}{6} + \sqrt{A} + 2B - C - 2D\right)}\right)$$

5. Meta-study and Discussion

This research constructed two *necessary* conditions for sample adequacy:

1. Section 3 determined the sample size needed compensate for the ratio of number of indicator variables to latent variables (summarized from Monte Carlo simulations that have appeared in the literature); and

2. Section 4 determined the sample size required to assure the existence or non-existence of a minimum effect (correlation) on each possible pair of latent variables in the SEM (determined analytically).

Of course, neither of these conditions is *sufficient* to assure sample adequacy for a particular choice of (α, β) because there are so many other factors that can affect estimation and sample size – multicolinearity, appropriateness of data sets, and so forth. Additionally, the information contained in the sample and indicator variables must be adequate to compensate for variations in particular SEM estimation methodologies. For example, partial least square (PLS) approaches generate parameter estimates that lack consistency. Dhrymes (1970); Schneeweiß (1990, 1991, 1993); Thomas, et al. (2005); and Fèhèr (1989) all demonstrate that the IV/2SLS techniques converge to the same estimators, but are more robust. Joreskog (1967, 1970; Jöreskog and Sörbom 1996) suggests that departures from normal distribution for the indicators will demand larger samples, and that non-normal indicators require one, two or three magnitudes larger samples, depending on distribution.

From a practical viewpoint, sample size questions can take three forms:

1. *A priori:* will ask what sample size *will be sufficient* given the researchers *prior beliefs* on what the minimum effect is that the tests will need to detect

2. *Ex posteriori:* will ask what sample size *should have* been taken in order to detect the minimum effect that the researcher *actually detected* in an existing (either sufficient or insufficient) test. If the *ex posteriori* measured effect is *smaller* than the researchers prior beliefs about the minimum effect (in 1.) then sample size needs to be increased commensurately.

3. *Sequential test optimal-stopping:* is couched in a sequential test optimal-stopping context, where the sample size is incremented until it is considered sufficient to stop testing.

In this section, we report on an *ex posteriori* meta-study that applies the algorithms developed in this paper to a specific body of SEM research – studies published in five core journals in MIS and e-Commerce (*ISR, MISQ, Management Science, Decision Sciences* and *JMIS*) between 1989 (the date of the seminal study by Davis, et al. 1989) and 2007. We assumed that the link with the smallest effect actually observed in these studies determines δ – a conservative assumption, because the research would have been very likely to hold a bias in actually wanting to detect even smaller effects than those actually observed, but the model and data would have only had sufficient resolution to capture the minimum effect observed.

Additionally, many of the studies listed in Appendix A analyzed Likert scale data that is not distributed normally; nevertheless, the *assumption* of normalcy of data is a common one in SEM studies, even where the data is clearly not normal, for example where survey data returns discrete Likert scale data censored at 0, and derives from a mass function which is likely to be skewed. Because estimator behavior is best understood for normal data, we can assume that, in these non-normal data studies our lower bound on sample size needs at a non-normalcy risk premium for sample adequacy – departures from a normal weight matrix in LISREL suggest that this may be two to three orders of magnitude larger than sample size required for normal data.

Sample sizes actually used in drawing conclusions in the study were compared with our computed 'lower bound', and a difference taken as a percentage (the far right hand column of Appendix A). Histograms of sample adequacy $\frac{Actual\ SS - Required\ SS}{Actual\ SS}$ show a significant systematic bias towards too small a sample size in the papers surveyed. In the meta-study, the average sample was 770% too small; with the removal of three outliers, this dropped to 400% too small (figures 6 and 7). Actual sample sizes in these 74 research articles were on average only 50% of the minimum needed to draw the conclusions the studies claimed; median sample size was 38% of the minimum required, reflecting a substantial negative skewing in the undersampling, and standard deviation was 29%. Overall, 80% of the research articles in this meta-study drew conclusions from samples that were smaller than the lower bounds on sample size computed here. Because each additional observation increases the cost of the study in time, effort and monetary terms, an inclination to economize on data collection is understandable. The conclusion that seems most appropriate from our meta-study is that MIS researchers have been given inadequate guidance, and have not been well served by existing sample size heuristics. Lacking the sample size information they need, researchers may be inclined to skimp on sample collection. Unfortunately, when samples are too large, the studies were more costly than they needed to be in drawing particular conclusions; when samples are too small, the credibility of their conclusions is weakened.

INSERT FIGURE 6: PERCENT ERROR IN SAMPLE SIZE FOR 74 STUDIES IN THE ENTIRE META-STUDY (MEAN=-770, STANDARD DEVIATION = 25, SKEWNESS=-6.5, KURTOSIS = 47)

INSERT FIGURE 7: PERCENT ERROR IN SAMPLE SIZE FOR 74 STUDIES IN THE META-STUDY REMOVING OUTLIERS $< -2500\%$ (MEAN=-400, STANDARD DEVIATION = 642, SKEWNESS=-2.5, KURTOSIS = 7.6)

We should not be surprised, given our review of the prior literature, that existing sample size heuristics are misleading researchers in this area. Numerous studies have concluded that linear heuristics like 'the rule of 10' are poor guides to fit and explanatory power of the model or adequacy of the sample size. (Browne and Cudeck 1989, 1993, Geweke, and Singleton 1980; Gebring and Anderson 1985); Veicer and Fava 1987, 1989, 1994; Marsh and Bailey 1991; Boomsma 1982; Ding, et al. 1995)

As noted earlier, neither of the conditions developed here is *sufficient* to assure sample adequacy for a particular choice of (α, β) because there are so many factors that can affect estimation and sample size in something as complex as a structural equation model. Consequently, the necessary sample size for accurate estimation will in most cases exceed the lower bound computed here. But review of actual sample sizes summarized in figures 6 and 7 suggests that, at its most unambitious, this lower bound will insure against the very erratic under-sizing of samples that seems common in SEM analysis.

Future research on sample size choice should be conducted on lines specific to the various algorithms used to estimate SEM – PLS's principal components analysis algorithms; LISREL and AMOS's gradient search algorithms; and systems of equations regression algorithms. Indeed, seminal research in each of these areas alluded to this decades ago. Wold (1980, 1981) went even further in advising that PLS is more suitable for exploratory model specification searches rather than hypotheses testing, and introduced the concept of *plausible causality* for that very reason. Thus in PLS, the sample size question is probably both less relevant and less critical, because hypothesis testing is better left to LISREL and systems of equation approaches.

The problem in building the structural model completely on theory, without reference to the data is that the latent constructs chosen by the researcher may be substantially different than those that would drop out of an exploratory factor analysis. Researchers have developed a test for this called *Harmon's one factor test* (Podsakoff and Organ 1986) commonly used to check for common factor bias in SEM (and often conducted *ex posteriori*). Common factor bias appears because inherent clustering results from a particular distance measure used to position data points in n-dimensional space – for example, principal components analysis designs a distance measure to minimize the variance not explained by the main components (clusters). But SEM will impose prior beliefs on the data, in the form of the structure of latent variables. Thus data are assumed to cluster around the latent constructs – the factor loadings determine how this clustering occurs. SEM models are often constructed without reference to clustering in the underlying data given a particular distance measure; it is entirely theory-driven, though this is not in itself a bad thing. Common factor bias reflects this divergence in the model and the data, and if it is too extreme, may indicate that the data is incomplete, or that the model is misspecified.

Common factor bias can be avoided *a priori* through a pretest of the clustering of indicator data. Common factor bias occurs because procedures that should be a standard part of model specification are in practice left until after the data collection and confirmatory analysis. Jöreskog developed PRELIS for these sorts of pretests and model re-specifications. If this clustering shows that the indicators are providing information on fewer variables than the researchers' latent SEM contains, this is an indication that more indicators need to be collected that will provide (1) additional information about the latent constructs that don't show up in the cluster analysis; and (2) additional information to split one exploratory factor into the two or more latent constructs the research needs to complete the hypothesized model. In exploratory factor analysis, the two tests that are most useful for this are the *Kaiser (1960) criterion* that retains factors with eigenvalues greater than one (unless a factor extracts at least as much information as the equivalent of one original variable, we drop it) and the *scree test* proposed by Cattell (1966) that compares the difference between two successive eigenvalues and stops taking factors when this drops below a certain level. In either case, the suggested factors are not necessarily the latent factors that the researcher's theory would suggest – rather they are the information that is actually provided in the data, this information being the main justification for the cost of data collection. So in practice, either test would set a maximum number of latent factors in the SEM if that SEM is to be explored with one's own particular dataset.

When SEM are built around valid real world constructs (even if these are unobservable) the algorithms proposed in this paper impose only weak additional assumptions on the indicators and latent variables in order to compute sample sizes adequate for estimation. Our limited application to a window of IS and e-commerce publications has shown that concerns are warranted concerning existing SEM sample size calculations and we need to remain suspicious of conclusions reached in studies based on inadequate sample sizes. Furthermore, a large number of studies in our sample devised their tests without first committing to minimum effect size that they were trying to detect, or indicated in portion of non-response in surveys. It is clear that journal referees need to begin asking for survey response, minimum effect size δ and a justification of the sample size. By incorporating these suggestions, it is argued that the research community will enhance the credibility and applicability of their research, with a commensurate improved impact and influence in both industry and academe.

Note: *I want to thank the reviewers and editors who persevered through several revisions of this paper, and helped nurture it to completion. Any remaining errors are fully my own responsibility.*

Note on software: *You may download at Elsevier's ECRA site a software package that computes the lower bounds developed in this paper. This software is written in Windows C# Forms to run on Windows platforms; in addition to a number of the packages in the R language, it was used to calculate the results in this paper.*

Appendix A: Sample Adequacy in a Set of e-Commerce and MIS SEM Studies

**INSERT APPENDIX A **

References

Anderson, T. Origins of the limited information maximum likelihood and two-stage least squares estimators. *Journal of Econometrics* 127, 2005, 1–16.

Anderson, T. and Rubin, H. . Estimator of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20, 1949, 46–63.

Anderson, T. and Rubin, H. The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 21, 1950,570–82.

Balakrishnan N. and C. R. Rao *Order Statistics: Applications*, ser. Handbook of statistics; v. 17. New York: Elsevier, 1998.

Barclay, D. W., Higgins, C., & Thompson, R. . The partial least squares (PLS) approach to causal modeling: Personal computer adaptation and use as an illustration. *Technology Studies*, 2(2), 1995, 285-309.

Basmann, R. A generalized classical method of linear estimation of coefficients in a structural equation. *Econometrica* 25, 1957 77–83.

Bentler, P.M. *EQS, Structural Equations, Program Manual, Program Version 3.0,* Los Angeles: BMDP Statistical Software, Inc., 1989, p. 6

Bollen, K. A. *Structural equations with latent variables.* New York: Wiley,1989, p. 268

Boomsma, A Robustness of LISREL against small sample sizes in factor analysis models, in KG Joreskog and H Wold (eds) *Systems under indirect observations, Causality, structure, prediction* (part 1) 1982, pp 149-173, Amsterdam: North Holland.

Brand A, Bradley MT, Best LA, Stoica G Accuracy of effect size estimates from published psychological research". *Perceptual and Motor Skills* 106 (2) 2008. 645–649

Browne, M. W., and Cudeck, R. Alternative ways of assessing model fit. In K. A. Bollen & J. S.Long (Eds.), *Testing structural equation models,* 1993, pp. 136–16, Newbury Park, CA: Sage.

Browne, M. W.,and Cudeck, R. Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research, 24,*1989*,* 445–455.

Cattell RB *Handbook of multivariate experimental psychology* 1966 - Rand McNally Chicago

Chin, W. W. The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern Methods for business research* (pp. 295-336). Mahwah, 1998, New Jersey: Lawrence Erlbaum Associates.

Chin, W. W., and Newsted, P. R. Structural Equation Modeling analysis with Small Samples Using Partial Least Squares. In Rick Hoyle (Ed.), *Statistical Strategies for Small Sample Research*, Sage Publications, 1999, pp. 307-341

Cochran, WG *Sampling Techniques*, 3rd Edition 1977 New York: Wiley

Cohen, J *Statistical Power Analysis for the Behavioral Sciences* (second ed.) 1988, Lawrence Erlbaum Associates

Cohen, J A power primer, *Psychological Bulletin* 112, 1992, 155–159

Davis, F.D., Richard P. Bagozzi, Paul R. Warshaw User acceptance of computer technology: a comparison of two theoretical models, *Management Science* Volume 35 , Issue 8, 1989, 982 - 1003

Dhrymes P J., R. Berner, D. Cummins A Comparison of Some Limited Information Estimators for Dynamic Simultaneous Equations Models with Autocorrelated Errors *Econometrica*, Vol. 42, No.2, 1974, pp. 311-332

Dhrymes P. *Distributed Lags: problems of estimation and formulation*, San Francisco: Holden Day, 1971

Dhrymes, P J. *Econometrics – Statistical Foundations and Applications*, New York Evanston and London (Harper & Row), 1970, p. 53

Ding, L. , Belicer, W. F. and Harlow, LL The effects of estimation methods, number of indicators per factor and improper solutions on structural equation modeling fit indices, *Structural Equation Modeling*, 2, 1995, 119-144

Farebrother, R. *Fitting Linear Relationships: A History of the Calculus of Observations 1750–1900*. 1999, New York: Springer.

Fèhèr, K. Comparison of LISREL and PLS Estimation Methods in Latent Variable Models. Introducing Latent Variables into Econometric Models, *Manuscript SFB 303*, University of Bonn, Bonn, 1989

Fisher, R.A. On the `probable error' of a coefficient of correlation deduced from a small sample, *Metron* 1: 1921, 3–32

Fisher,R. A. *Statistical Methods, Experimental Design, and Scientific Inference*. New York: Oxford Univ. Press, 1990.

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

Freedman, D As others see us: A case study in path analysis (with discussion) *Journal of Educational Statistics*, 12, 1987, pages 101--223.

Gerbing, D. W.,&Anderson, J. C. The effects of sampling error and model characteristics on parameterestimation for maximum likelihood confirmatory factor analysis. *Multivariate Behavioral Research, 20,* 1985*,* 255–271.

Gibbons, J.D. and S. Chakraborti*, Nonparametric Statistical Inference*, 3rd ed. New York: Marcel Dekker, 1992.

Goodhue, Dale William Lewis, Ronald Thompson, Statistical Power in Analyzing Interaction Effects: Questioning the Advantage of PLS with Product Indicators (Research Note), *Information Systems Research* Vol. 18, No. 2, 2007, pp. 211-227

Goodhue,D. William Lewis, Ron Thompson, "PLS, Small Sample Size, and Statistical Power in MIS Research," HICSS, vol. 8, pp.202b, *Proceedings of the 39th Annual Hawaii International Conference on System Sciences* (HICSS'06) 2006

Haavelmo, T. The Statistical Implications of a System of Simultaneous Equations *Econometrica* 11, 1943, 1-12.

Hea, Q. and H. N. Nagarajab Correlation Estimation Using Concomitants of Order Statistics from Bivariate Normal Samples, *Communications in Statistics - Theory and Methods*, Volume 38, Issue 12, January 2009 , pages 2003 - 2015

Hoeffding, W. A class of statistics with asymptotically normal distribution. *Ann. Mathemat. Statist*. 19, 1948,293–325.

Jöreskog, K. G. Some contributions to maximum likelihood factor analysis. *Psychometrika* , 32(4),1967, 443-482.

Jöreskog, K. G. and Sörbom, D., *LISREL 8 User's Reference Guide*, Chicago: Scientific Software International, 1996.

Joreskog, K.G. A general method for analysis of covariance structures, *Biometrika*, 57, 1970, 239-251.

Kahai, S.S. and Cooper, R.B. Exploring the Core Concepts of Media Richness Theory: The Impact of Cue Multiplicity and Feedback Immediacy on Decision Quality, *Journal of Management Information Systems*, 20, 1, 2003 263-299

Kendall, M and J. D. Gibbons *Rank Correlation Methods*, 5th ed. New York: Oxford Univ. Press, 1990.

Kish, L *Survey Sampling*, 1995, New York: Wiley

Koopmans, T. and Hood, W. The estimation of simultaneous linear economic relationships. In Studies in Econometric Method, ed. W. Hood and T. Koopmans. *Cowles Foundation Monograph* 14. 1953 New Haven: Yale University Press.

Lohr, S L. *Sampling: Design and Analysis*. 1999 Duxbury

Mari D. D. and S. Kotz*,* *Correlation and Dependence*. London, U.K.:Imperial College Press, 2001.

Marsh, H. W., Balla, J. R., & McDonald, R. P. Goodness of fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,*1988 391–410.

Marsh, H. W., Balla, J. R.,&Hau, K.-T. An evaluation of incremental fit indices:Aclarificationof mathematical and empirical properties. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 315–353). 1996, Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Marsh, H. W., Hau, K.-T., Balla, J. R.,&Grayson, D.. Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33,*1998, 181–220.

Marsh, H. W. and M Bailey Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models *Applied Psychological Measurement*, Vol. 15, No. 1, 47-70 (1991)

Nakagawa, S. and Cuthill, I. C. Effect size, confidence interval and statistical significance: a practical guide for biologists *Biological Reviews Cambridge Philosophical Society* 82, 2007, 591–605

Nunnally, J.C. *Psychometric Theory*, New York: McGraw-Hill, 1967, p. 355

Podsakoff, P.M. and D. W. Organ Self-reports in organization research: Problems and prospects, *Journal of Management*, 12, 1986, 531-544

Sargan, J. Estimation of economic relationships using instrumental variables. Econometrica 67, 1958, 557–86.

Schechtman, E., Yitzhaki, S. A measure of association based on Gini’s mean difference. *Commun. Statist. Theor. Meth*. 16, 1987, 207–231.

Schneeweiß, H. Models with Latent Variables: LISREL versus PLS, in: *Contemporary Mathematics* Vol. 112 (1990), p. 33-40

Schneeweiß, H. Models with Latent Variables: LISREL versus PLS, in: *Statistica Neerlandica* 45 (1991), p. 145-157

Schneeweiß, H. Consistency at Large in Models with Latent Variables, in: K. Hagen, D. J. Barthdomew, M. Deistler, *Statistical Modelling and Latent Variables*, Elsevier (1993), p. 299-320

Shevlyakov G. L. and N. O. Vilchevski*,* *Robustness in Data Analysis: Criteria and Methods*, ser. Modern probability and statistics. *2002,* Utrecht, The Netherlands:, VSP, 2002.

Snedecor and Cochran *Statistical Methods*, 8th ed. 1989 Ames: Iowa U. Press

Tanaka, J. S. How big is big enough?”: Sample size and goodness of fit in structural equation models with latent variables. *Child Development, 58,*1987*,* 134–146.

Tanaka, J. S. Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 10–39), 1993, Newbury Park, CA: Sage.

Theil, H. *Repeated least-squares applied to a complete equation systems*. 1953a, The Hague: *Central Planning Bureau*.

Theil, H. *Estimation and simultaneous correlation in complete equation systems*. 1953b, The Hague: *Central Planning Bureau*.

Theil, H. *Economic Forecasts and Policy*, 2nd edn. 1961, Amsterdam: North-Holland.

Thomas, D. R., Lu, I. R. R. & Cedzynski, M. Partial least squares: A critical review and a potential alternative. *Proceedings of the Annual Conference of Administrative Sciences Association of Canada, Management Science Division*, Toronto, 2005

Velicer, W. F., and Fava, J. L. Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods, 3,*1998*,* 231–251.

Westland, J. C. and W.K. See-To The Short-run Price-Performance Dynamics of Microcomputer Technologies, *Research Policy*, Volume 36, Issue 5, 2007, Pages 591-604

Wilkinson, Leland; APA Task Force on Statistical Inference Statistical methods in psychology journals: Guidelines and explanations". *American Psychologist* 54, 1999, 594–604

Wold, H. "The Fix-Point Approach to Interdependent Systems: Review and Current Outlook," in H. Wold (Ed.), The Fix-Point Approach to Interdependent Systems, 1981, Amsterdam: North-Holland, 1-35.

Wold, Herman Response to D. A. Freedman, *Journal of Educational Statistics*, Vol. 12, No. 2, 1987, pp. 202-205

Wright, S. Correlation and causation, *Journal of Agricultural Research*, 20, 1921 557-585

Xu*, W*, Y. S. Hung,, M. Niranjan*,* and M. Shen Asymptotic Mean and Variance of Gini Correlation for Bivariate Normal Samples, IEEE Trans. On Signal Processing, V. 58(2) 2010

Appendix A: Sample Adequacy in a Set of e-Commerce and MIS SEM Studies

Studies	*Latent Variables*	*Indicator Variables*	*Sample Points*	*Minimum Effect Observed*	*Šidàk corrected α*	*Sample bound section 4*	*Sample bound section 3*	*Sample Size lower bound*	*Study sampl (-) or above*
Choudhury and Karahanna , MISQ, V.32(1)	6	35	499	0.12	0.0034	1177	176	1177	
Kanawattanachai and Yoo, MISQ, V.31(4)	3	11	146	0.21	0.0170	264	122	264	
Damien , et al. , MISQ, V.31(3)	6	43	31	0.08	0.0034	2694	443	2694	
Webster and Ahuja, MISQ, V.30(3)	6	38	207	0.17	0.0034	568	256	568	
Tanriverdi, MISQ, V.30(1)	10	16	356	0.11	0.0011	1661	508	1661	
Awad and Krishnan, MISQ, V.30(1)	8	24	532	-0.28	0.0018	206	200	206	
Van der Heijden nd Hans, MISQ, V.28(4)	4	15	1144	0.15	0.0085	628	116	628	
Barua, et al., MISQ, V.28(4)	12	45	1125	0.109	0.0008	1783	116	1783	
Straub, et al., MISQ, V.27(1)	8	34	213	0.1	0.0018	1886	91	1886	
Susaria, et al., MISQ, V.27(1)	8	32	256	0.067	0.0018	4253	100	4253	
Tarafdar, et al , JMIS, V.24 (1)	10	21	256	0.08	0.0011	3181	376	3181	
Fuller , et al , JMIS, V.23 (3)	5	22	318	0.3	0.0051	148	88	148	
Kearns, et al , JMIS, V.23 (3)	9	44	269	0.11	0.0014	1610	95	1610	
Wang, et al , JMIS, V.23 (2)	5	22	149	0.12	0.0051	1098	88	1098	
Lopes, et al , JMIS, V.23 (2)	4	13	392	0.24	0.0085	226	166	226	
Hess, et al, JMIS, V.22 (3)	8	36	233	-0.28	0.0018	206	88	206	
Johnson, et al , JMIS, V.22 (2)	7	16	202	-0.19	0.0024	472	333	472	
Bhatt, et al , JMIS, V.22 (2)	5	26	202	0.22	0.0051	303	112	303	
Chang et al , JMIS, V.22 (1)	3	12	476	0.65	0.0170	9	100	100	
Wallace, et al. Dec.Sci., V.35(2)	6	27	507	0.13	0.0034	997	88	997	
Rabinovich, et al., Dec.Sci., V.34(1)	2	14	840	0.126	0.0500	587	400	587	
Abdinnour-Helm, et al. Dec.Sci., V.36(2)	5	12	176	0.1	0.0051	1596	308	1596	
Escrig-Tena , et al., Dec.Sci., V.36(2)	9	37	231	0.1	0.0014	1957	95	1957	
Pullman and Gross, Dec.Sci., V.35(3)	3	15	400	0.52	0.0170	24	100	100	
Tu, et al., Dec.Sci., V.35(2)	3	24	303	0.22	0.0170	238	700	700	
Droge, et al., Dec.Sci., V.34(3)	4	17	437	-0.16	0.0085	548	91	548	
Janz and Prasarnphanich, Dec.Sci., V.34(2)	5	15	231	0.41	0.0051	65	200	200	
Hong and Tam, ISR, V.17(2)	5	27	1328	0.1	0.0051	1596	128	1596	
Dinev and Hart, ISR, V.17(1)	5	18	369	-0.15	0.0051	690	128	690	
Jarvenpaa, et al., ISR, V.15(3)	6	30	136	-0.28	0.0034	187	100	187	
Pavlou and Gefen, ISR, V.15(1)	10	33	274	0.1	0.0011	2020	160	2020	
Bassellier, et al., ISR, V.14(4)	7	38	404	0.2	0.0024	422	131	422	
Choo, et al., MS, V.53(3)	5	14	951	0.38	0.0051	81	232	232	

Sabherwal, et al., MS, V.52(12)	10	48	121	0.11	0.0011	1661	92	1661
Bagozzi and Dholakia, MS, V.52 (7)	14	24	402	0.17	0.0006	736	476	736
de Jong, et al., MS, V.51(11)	5	29	60	0.037	0.0051	11884	172	11884
Kim and Malhotra, MS, V.51(5)	4	8	189	0.14	0.0085	725	400	725
Vickery, et al., MS, V.50 (8)	4	14	113	0.212	0.0085	299	138	299
Balasubramanian, et al, MS, V.49(7)	5	22	428	-0.12	0.0051	1098	88	1098
Au, et al., MISQ, V.32(1)	6	20	922	0.17	0.0034	568	156	568
Hsieh, et al., MISQ, V.32(1)	12	31	451	0.21	0.0008	447	271	447
Nadkarni and Gupta, MISQ, V.31(3)	6	30	452	0.1	0.0034	1711	100	1711
Liang, et al., MISQ, V.31(1)	7	21	77	0.277	0.0024	202	200	202
Komiak and Benbasat, MISQ, V.30(4)	7	16	100	0.12	0.0024	1242	333	1242
Bhattacherjee and Sanford, MISQ, V.30(4)	7	24	81	0.26	0.0024	235	145	235
Karahanna, et al., MISQ, V.30(4)	8	29	278	0.1	0.0018	1886	126	1886
Srite and Karahanna, MISQ, V.30(3)	8	27	181	0.29	0.0018	190	151	190
Stewart and Gosain, MISQ, V.30(2)	8	63	51	0.13	0.0018	1099	657	1099
Moores and Chang, MISQ, V.30(1)	5	17	243	-0.47	0.0051	43	148	148
Pavlou and Fygenson, MISQ, V.30(1)	6	18	312	-0.12	0.0034	1177	200	1177
Ahuja and Thatcher, MISQ, V.29(3)	4	12	263	0.17	0.0085	482	200	482
Ko, et al. , MISQ, V.29(1)	5	14	96	0.134	0.0051	874	232	874
Bhattacherjee and Premkumar, MISQ, V.28(2)	7	27	77	0.1	0.0024	1806	108	1806
Gemino, JMIS, V.24 (3)	7	51	223	0.154	0.0024	739	476	739
Son, et al , JMIS, V.24 (1)	11	38	625	0.16	0.0009	783	142	783
Klein, et al., Dec.Sci., V.38(4)	5	30	91	0.2	0.0051	373	200	373
Wang and Wei, Dec.Sci., V.38(4)	4	46	150	0.22	0.0085	275	2538	2538
Keil, et al. , Dec.Sci., V.38(3)	3	14	178	0.15	0.0170	543	89	543
Ettlie and Pavlou, Dec.Sci., V.37(2)	4	31	72	-0.15	0.0085	628	616	628
Looney, et al., Dec.Sci., V.37(2)	5	31	414	0.153	0.0051	662	232	662
Brown and Chin, Dec.Sci., V.35(3)	7	20	240	0.14	0.0024	902	222	902
Teigland and Wasko, Dec.Sci., V.34(2)	9	19	83	-0.2	0.0014	458	373	458
Saraf, et al., ISR, V.18(3)	9	45	63	0.254	0.0014	268	100	268
Pavlou and Dimoka, ISR, V.17(4)	4	10	1665	0.05	0.0085	5904	288	5904
Nicolaou and McKnight, ISR, V.17(4)	6	32	69	0.247	0.0034	250	122	250
Pavlou and El Sawy, ISR, V.17(3)	4	10	507	0.14	0.0085	725	288	725
Pavlou and Gefen, ISR, V.16(4)	10	30	1031	-0.14	0.0011	1009	200	1009
Wixom and Todd, ISR, V.16(1)	7	17	465	0.1	0.0024	1806	302	1806

Zhu and Kraemer, ISR, V.16(1)	11	34	624	-0.04	0.0009	13213	187	13213
Malhotra, et al., ISR, V.15(4)	6	18	449	-0.12	0.0034	1177	200	1177
Karimi, et al., ISR, V.15(2)	5	20	286	0.04	0.0051	10163	100	10163
Mun and Davis, ISR, V.14(2)	7	54	95	0.18	0.0024	530	604	604
Venkatesh and Agarwal, MS, V.52(3)	5	21	757	0.1	0.0051	1596	92	1596
Ahuja, et al., MS, V.49(1)	4	5	1781	0.17	0.0085	482	616	616

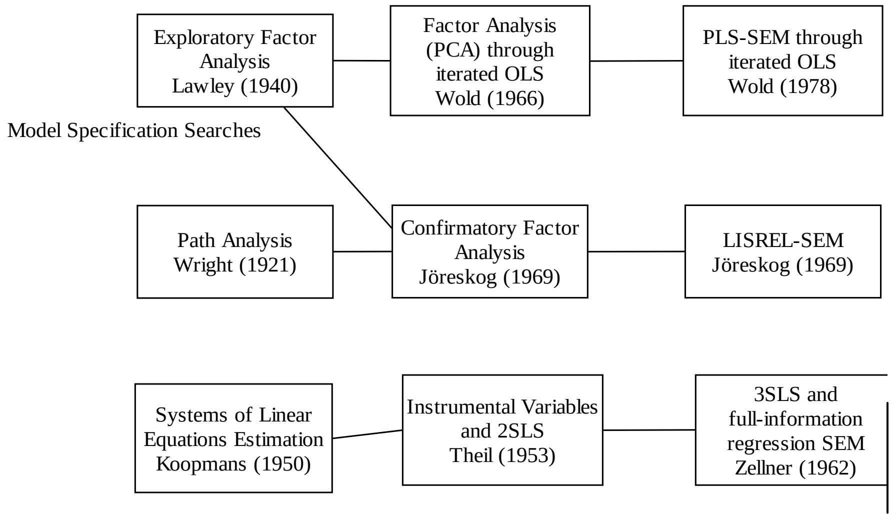

FIGURE 2: DEVELOPMENT OF STRUCTURAL EQUATION MODEL ESTIMATION

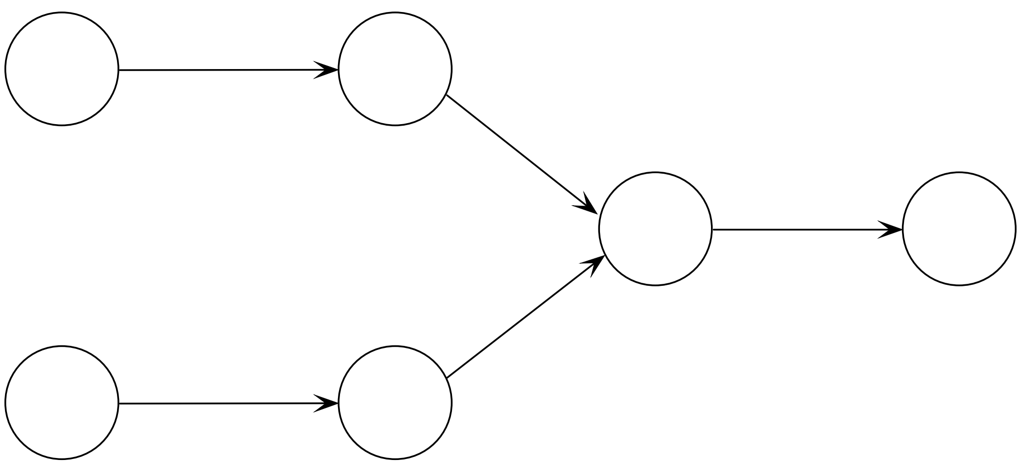

FIGURE 2: AN EXAMPLE OF A STRUCTURAL EQUATION MODEL WITH SIX LATENT VARIABLES AND FIVE CORRELATIONS

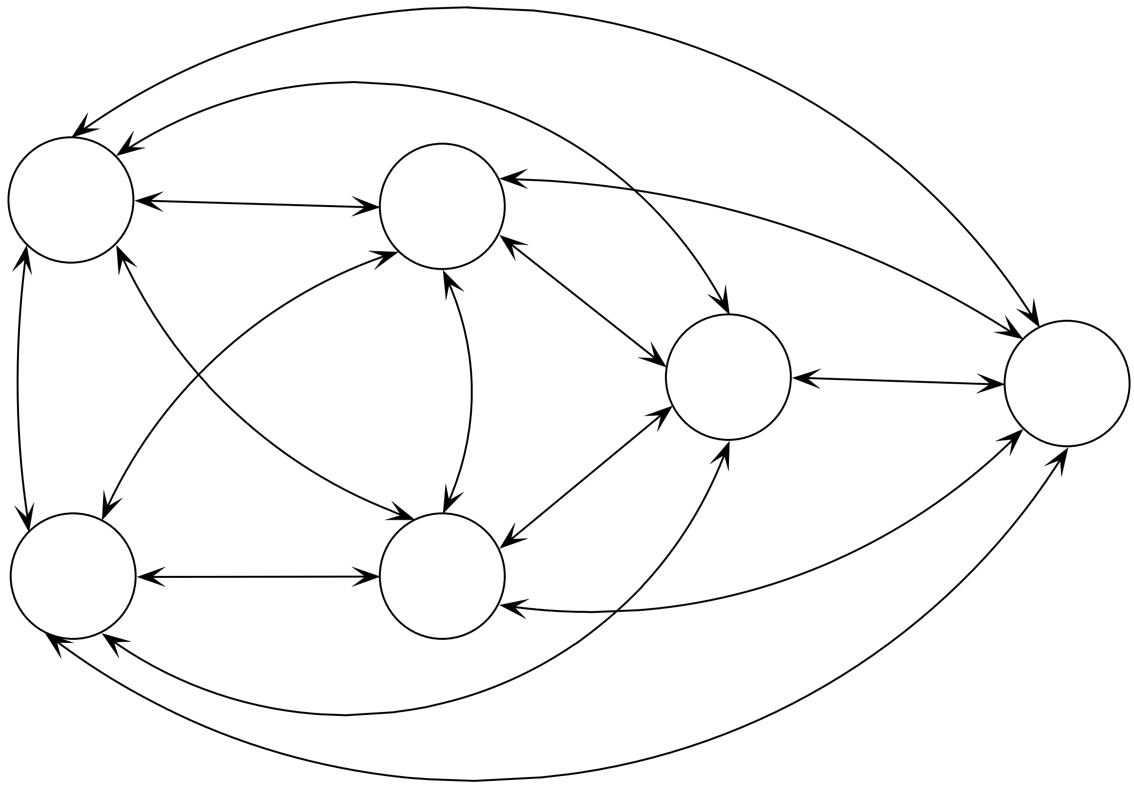

FIGURE 3: THE SEM EXAMPLE IN FIGURE 2 WITH ALL POSSIBLE PAIRED LINKS SHOWN

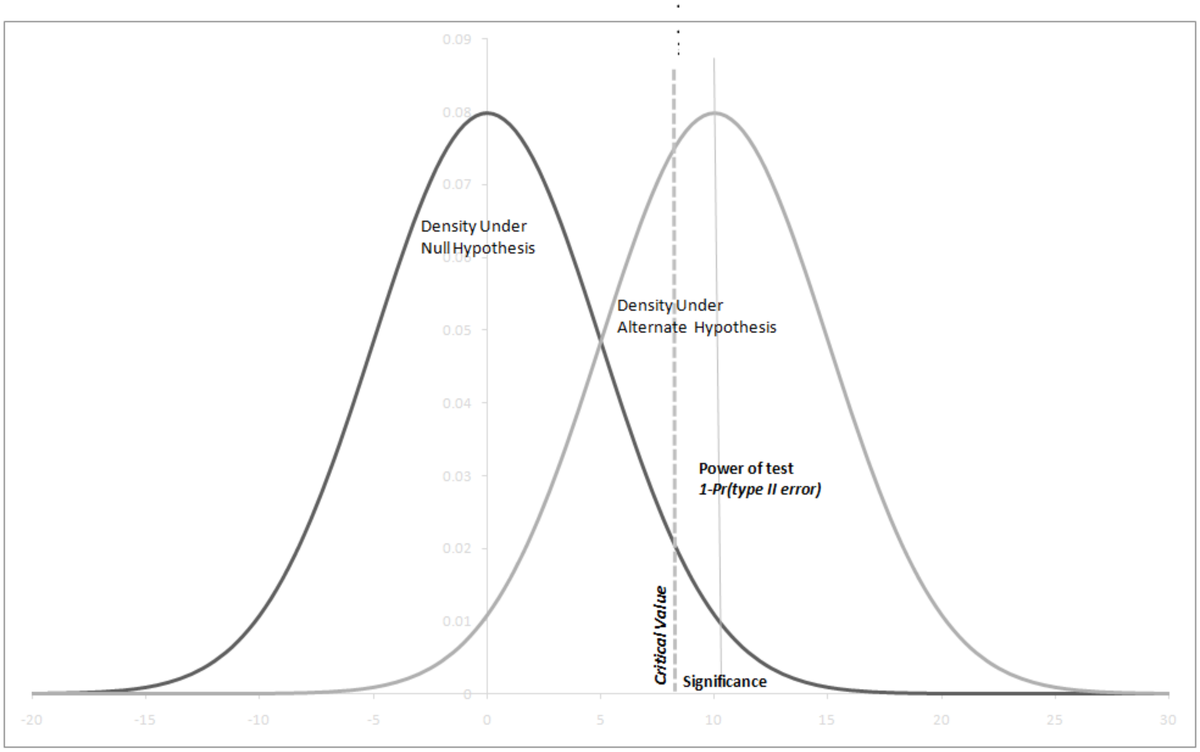

FIGURE 4: SIGNIFICANCE AND POWER FOR THE MINIMUM EFFECT THAT NEEDS TO BE DETECTED

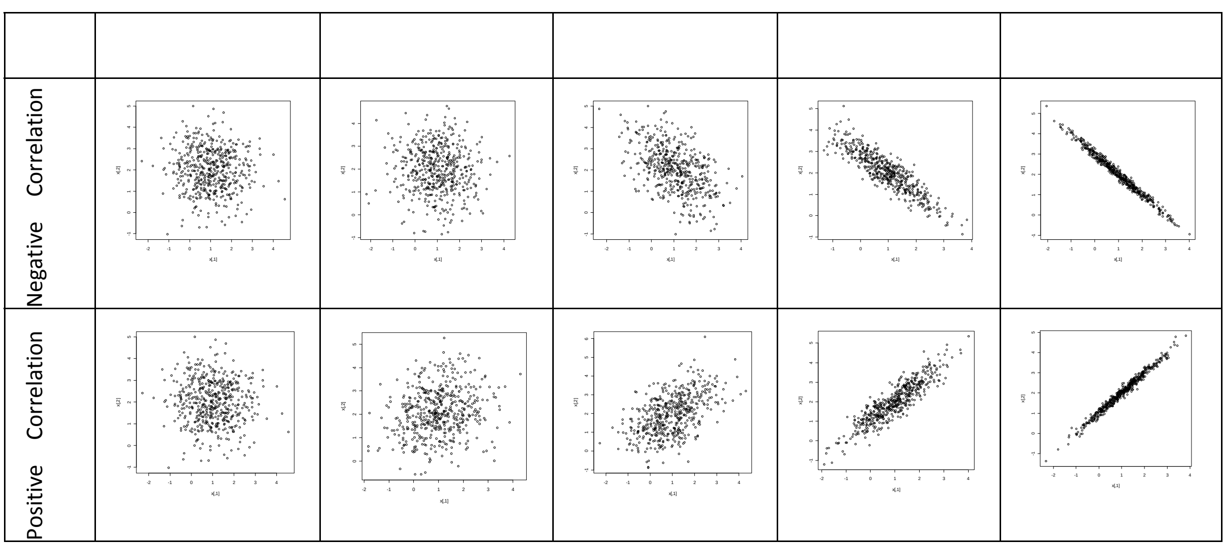

Figure 5: Bivariate Normal Scatterplots for $\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}$ *and* $\Sigma = \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}$ *with* $n = 500$

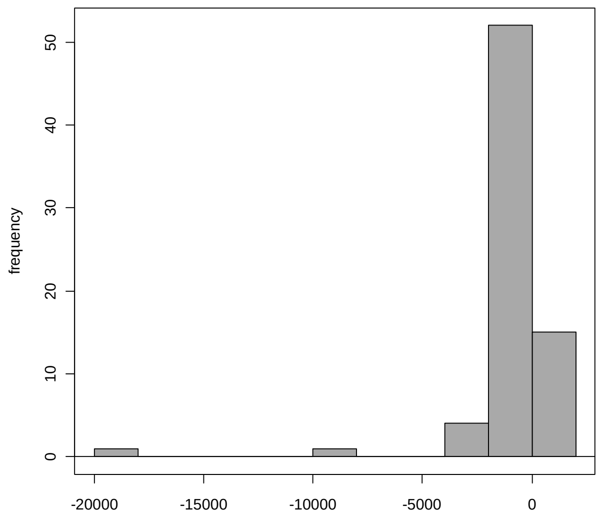

FIGURE 6: PERCENT ERROR IN SAMPLE SIZE FOR 74 STUDIES IN THE ENTIRE META-STUDY (MEAN=-770, STANDARD DEVIATION = 25, SKEWNESS=-6.5, KURTOSIS = 47)

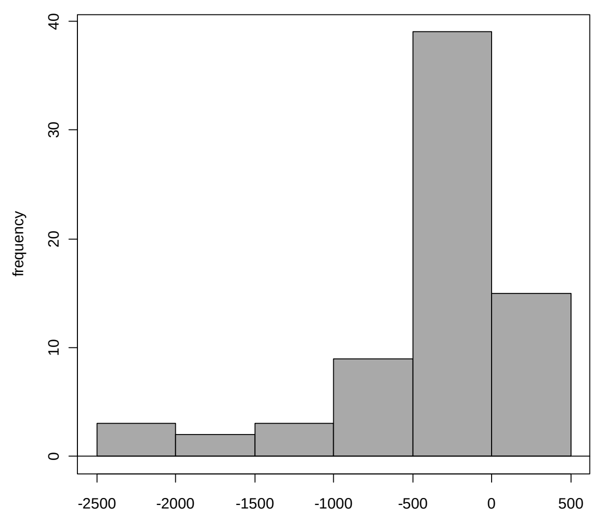

FIGURE 7: PERCENT ERROR IN SAMPLE SIZE FOR 74 STUDIES IN THE META-STUDY REMOVING OUTLIERS $< -2500\%$ **(MEAN=-400, STANDARD DEVIATION = 642, SKEWNESS=-2.5, KURTOSIS = 7.6)**

p	Number of parameters (indicators) in the SEM
k	Number of latent variables in the SEM
n	Computed sample size lower bound
$[\tilde{X}, \tilde{Y}]$ and $[X_i, Y_i]$	Bivariate Normal random latent variables (and their realization) in the SEM
$X_{1:n} \leq X_{2:n} \dots$ $Y_{1:n} \leq Y_{2:n} \dots$	order statistics of the (X_i, Y_i) sample values; the first index is rank, and the second is sample size
$Y_{[i:n]}$	concomitant of the i^{th} order statistic; $Y_{[i:n]}$ is the Y sample value associated with the $X_{i:n}$ sample value in the sample pairs (X_i, Y_i).
δ	Minimum effect size that our computed sample size can detect
ρ	Unknown correlation for a bivariate Normal random vector $[\tilde{X}, \tilde{Y}]$
$\hat{\rho}_G$	Estimator of Gini correlation ρ_G
$[\hat{\mu}_G; \hat{\sigma}_G]$	Mean and standard deviation estimators for Gini correlation
$[\alpha^*;\ 1-\beta]$	Significance and power of test
α	The Šidàk corrected significance for discriminations between possible SEM link combinations at a resolution of δ
$[z_{1-\alpha};\ z_{1-\beta}]$	Rejection bound at significance α and non-rejection bound at power $1-\beta$; we substitute the quantile function (inverse cumulative Normal) $\Phi^{-1}(x)$ for z_x in calculations

TABLE 2: NOTATION USED IN THE PAPER