Sociological Methods & Research
Volume 36 Number 1
August 2007 48-86
© 2007 Sage Publications
10.1177/0049124107301947
http://smr.sagepub.com
hosted at
http://online.sagepub.com

Latent Variable Models Under Misspecification

Two-Stage Least Squares (2SLS) and Maximum Likelihood (ML) Estimators

Kenneth A. Bollen
The University of North Carolina at Chapel Hill
James B. Kirby
Agency for Healthcare Research and Quality, Rockville, MD
Patrick J. Curran
The University of North Carolina at Chapel Hill
Pamela M. Paxton
The Ohio State University, Columbus
Feinian Chen
North Carolina State University, Raleigh

This article compares maximum likelihood (ML) estimation to three variants of two-stage least squares (2SLS) estimation in structural equation models. The authors use models that are both correctly and incorrectly specified. Simulated data are used to assess bias, efficiency, and accuracy of hypothesis tests. Generally, 2SLS with reduced sets of instrumental variables performs similarly to ML when models are correctly specified. Under correct specification, both estimators have little bias except at the smallest sample sizes and are approximately equally efficient. As predicted, when models are incorrectly specified, 2SLS generally performs better, with less bias and more accurate hypothesis tests. Unless a researcher has tremendous confidence in the correctness of his or her model, these results suggest that a 2SLS estimator should be considered.

Keywords: *2SLS; misspecification; latent variable models; structural equation models; FIML; specification error*

Structural equation modeling (SEM) with latent variables is a basic tool in social science research. The maximum likelihood (ML) estimator is by far the dominant estimator for these models. It is a *full information* estimator that simultaneously estimates all parameters while using information from the whole system of equations and is sometimes called the full information ML (FIML) estimator. We will use ML as synonymous with FIML in this article. Limited information estimators are less common in these general SEMs, though they are more common in simultaneous equations without latent variables. Bollen (1996a, 1996b) proposed a limited information two-stage least squares (2SLS) estimator for latent variable SEMs. Limited information estimators in other contexts are more robust to structural specification errors in models. An important question is whether this 2SLS estimator better isolates errors in models than does the ML, and if so, at what cost it does this. The purpose of this article is to compare the performance of the ML and 2SLS estimators of parameters in latent variable models under conditions of correctly and incorrectly specified models and across different sample sizes. Given the common situation of estimating imperfect models in small and moderate sample sizes, there is important practical usefulness in knowing the relative performance of these two estimators.

The ML estimator's dominant position in SEMs is due to several factors. One practical reason is that the ML estimator is the default estimator in SEM software. Another reason for its popularity is that under correct model specification and with observed variables that come from distributions with no excess multivariate kurtosis, the ML estimator is consistent, asymptotically unbiased, asymptotically efficient, and asymptotically normal, and we can estimate the asymptotic covariance matrix of the parameter estimator (e.g., see Browne 1984). The 2SLS estimator has the same properties except that it is asymptotically efficient among limited information estimators rather than full information estimators. As such, we would expect some large-sample efficiency advantage for the ML estimator when the underlying assumptions are met (Bollen 1996b).

The problem with this analytical description of the ML and 2SLS estimators is that it assumes ideal conditions that commonly fail in practice. For instance, in real applications the observed variables typically come from nonnormal distributions with excess kurtosis (Micceri 1989); the sample size might not be large; and the model is nearly certain to have structural specification

Authors' Note: An earlier version of this article was presented at the Research Methods 2003 Conference, Free University, June 25–27, 2003, Amsterdam, Netherlands. We wish to thank John Hipp for his research assistance and the reviewers for their comments. Kenneth Bollen gratefully acknowledges support from NSF SES 0617276. Bollen and Patrick Curran also are thankful for support from NIDA 1-R01-DA13148-01 and DA013148-05A2. Please address correspondence to Kenneth A. Bollen, CB 3210 Hamilton, Department of Sociology, The University of North Carolina, Chapel Hill, NC 27599-3210; e-mail: bollen@unc.edu.

errors. The 2SLS estimator is asymptotically "distribution free" so that in large samples its asymptotic standard errors and significance tests do not depend on normality. Bootstrapping techniques (Bollen and Stine 1990, 1993) or corrected asymptotic standard errors enable users to take account of nonnormality (or excess kurtosis) of the observed variables (Browne 1984; Satorra and Bentler 1988) when using the ML estimator. Under special conditions, some standard errors and the χ^2 test will be asymptotically robust to nonnormality (see review in Satorra 1990). However, the asymptotic efficiency of the ML estimator is established under the condition of no excess kurtosis so that even though we can correct the significance tests or we might meet the robustness conditions, this does not necessarily maintain the asymptotic efficiency of this full information estimator.

The impact of a structurally misspecified model is even more difficult to handle. The econometric literature on 2SLS in multiequation models with observed variables suggests that the 2SLS estimator is less sensitive to structurally misspecified models than are full information estimators such as the ML estimator (Cragg 1968). Yet this issue has not been examined in latent variable models. Bollen (2001) provides the conditions for the 2SLS estimator to be robust to specification errors in both the simultaneous equation and the latent variable models, but the analogous conditions are not available for the ML estimator. In addition, we do not have sufficient evidence on the behavior of 2SLS and ML estimators of latent variable models across different sample sizes.

In sum, both ML and 2SLS offer many promising advantages. However, the specific experimental conditions under which these advantages hold are not well understood, thus making an informed selection of one estimator over the other difficult in practice. To understand these complex issues better, we use Monte Carlo simulations to examine the relative performance of the 2SLS and ML estimators. Specifically, we study sample sizes ranging from 50 to 1,000; models with varying degrees of structural specification error; and three versions of the 2SLS estimator that are distinguished by the number of instrumental variables (IVs) that are included in the estimation. We vary the number of IVs since analytic work with the 2SLS estimator in simultaneous equation models without latent variables suggests that this can affect the finite sample properties of the estimator (e.g., Nagar 1959). For each estimator, we examine the percentage of bias in parameter estimates, the standard deviation of the estimate, and the accuracy of hypothesis tests.

The next section of the article presents the general structural equation model with latent variables and the ML and 2SLS estimators. In this section, we briefly summarize key properties of these estimators and their behavior in different sample sizes and with different degrees of model misspecification. We follow this with a discussion of the research issues about the properties of the ML and 2SLS estimators in correct and incorrect models. Next is a presentation of the design of our simulation experiment including the primary model and the design factors. After this section we present our findings on bias, standard deviations, and hypothesis testing using the ML and 2SLS estimators. Finally, we summarize the key findings and discuss their implications.

Model and Estimators

We use a slight modification of the LISREL notation (Bollen 2001) to represent the SEM:

$$\begin{aligned}\boldsymbol{\eta} &= \alpha_\eta + \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta},\\ \boldsymbol{y} &= \alpha_y + \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon},\\ \boldsymbol{x} &= \alpha_x + \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta},\end{aligned} \tag{1}$$

where $\boldsymbol{\eta}$ is the vector of latent endogenous variables,[1] $\boldsymbol{\xi}$ is the vector of latent exogenous variables, and $\boldsymbol{\zeta}$ is the vector of disturbances. The $\boldsymbol{B}$ matrix gives the effect of the endogenous latent variables on each other, and $\boldsymbol{\Gamma}$ is the matrix of coefficients for the effects of the latent exogenous variables on the latent endogenous variables. Intercepts are indicated by α. The $\boldsymbol{y}$ and $\boldsymbol{x}$ vectors are the observed variables affected by $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, respectively, with their coefficients in $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$. In the latent variable equation of (1), we assume that $E(\boldsymbol{\zeta}) = 0$ and the COV($\boldsymbol{\xi}$, $\boldsymbol{\zeta}'$) $= 0$. In the measurement model equations of (1), we assume that $E(\boldsymbol{\varepsilon}) = 0$, $E(\boldsymbol{\delta}) = 0$, and these unique factors are uncorrelated with $\boldsymbol{\xi}$, $\boldsymbol{\zeta}$, and each other. As is well known, we can represent simultaneous equations with no measurement error, confirmatory factor analysis, and numerous other common statistical models as special cases of these equations (e.g., see Bollen 1989).

ML Estimator

The most common estimator of the parameters for this model is the ML estimator:

$$\begin{aligned}F_{ML} = \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \mathrm{tr}[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{S}] - \ln|\boldsymbol{S}| - p\\ + [\bar{z} - \boldsymbol{\mu}(\boldsymbol{\theta})]'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})[\bar{z} - \boldsymbol{\mu}(\boldsymbol{\theta})],\end{aligned} \tag{2}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the model-implied covariance matrix, $\boldsymbol{S}$ is the sample covariance matrix, $\boldsymbol{\mu}$ is the vector of population means, $\boldsymbol{\mu}(\boldsymbol{\theta})$ is the model-implied mean vector, $\bar{z}$ is the vector of sample means of the observed variables, and p is the number of observed variables. The $\boldsymbol{\theta}$ vector contains the parameters of the model that are to be estimated.

Asymptotic properties of ML. . The ML estimator is a full information estimator in that it simultaneously estimates all model parameters and takes account of the full system of equations including constraints and restrictions when developing the estimates. When the observed variables come from distributions with no excess multivariate kurtosis and when the model is correct, then the $\hat{\boldsymbol{\theta}}$ that minimizes F_{ML} is consistent, asymptotically unbiased, asymptotically efficient, and asymptotically normal, and we can consistently estimate its covariance matrix (Jöreskog 1973; Browne 1984). Some research is available on the behavior of $\hat{\boldsymbol{\theta}}$ when there is excess kurtosis (Boomsma 1982; Browne 1984; Curran 1994). The main consequence is that the significance tests are likely to be inaccurate, though the consistency of the ML estimator is not affected. There also are special conditions under which some of the significance tests are asymptotically robust (Satorra 1990).

Properties of ML for structurally misspecified models.. Much less is known of ML's properties when the model is structurally misspecified, particularly in models with latent variables. Kaplan (1988) and Curran (1994) use simulations to illustrate that the ML estimator is susceptible to spreading the bias due to specification error in one equation to other correctly specified parts of the model. These results are consistent with research from econometrics. The econometric literature has examined the behavior of the ML estimator in simultaneous equations without measurement error or latent variables. If a simultaneous equation model suffers from misspecifications such as omitted variables, then the misspecification in one part of the system can spread bias to coefficient estimates from other equations even when the other equations are well specified.[2] Hausman's (1983:414) description of this is typical: "With system estimation misspecification in any equation in the system will generally lead to inconsistent estimation of *all* equations in the system." Furthermore, the asymptotic efficiency and asymptotic unbiasedness properties of ML are called into question when the model is structurally misspecified.

Finite sample properties of ML. . Research on the finite sample properties of the ML estimator in latent variable models also is sparse (Phillips

1983:490). Finite sample properties are typically approached in two ways. One is via analytical results where exact or approximate expressions are derived that provide the bias and variance of the ML estimator at different finite sample sizes. The second approach is to use Monte Carlo techniques to provide finite sample bias and variance using simulated data for particular models, parameter values, and sample sizes.

There is little analytical or simulation research on the properties of the parameter estimates from ML in latent variable SEM. Turning again to the econometric literature and simultaneous equation models, we see that Sargan (1970) has shown that the ML estimator does not always have finite moments (see also Mariano 1982:512–13; Phillips 1983:491–2). In special cases such as multiple regression with a normal disturbance term, where the ML estimator is equivalent to ordinary least squares (OLS), there are well-known finite moments. However, we cannot always determine the exact expected value or variance of the ML coefficient estimator in finite samples. Although this work on nonexistence of finite moments for the ML is in the context of simultaneous equations, it has implications for the ML estimator in the more general SEM as well since simultaneous equations are a special case of the general SEM. Furthermore, the same analytics that establish the nonexistence of finite moments in simultaneous equations (e.g., see Phillips 1983:491–2) are likely to apply to latent variable SEMs that use ML.

Lacking analytical expressions for finite moments, Monte Carlo simulations provide evidence on the simulation mean bias and variance of the ML. Boomsma (1982) studied the ML estimator in a confirmatory factor analysis model where he varied the magnitude of factor loadings, the correlation between the factors, the number of indicators per factor, and the sample size. Boomsma (1982) concluded that though there were biases in the parameter estimates and standard errors for the smaller sample sizes ($N < 100$), these were negligible in samples larger than 100 (see also Boomsma and Hoogland 2001).

Gerbing and Anderson (1985) also studied a confirmatory factor analysis model. Their sample sizes varied from 50 to 300, and they varied the number of indicators per factor, the numbers of factors, the magnitude of the correlations of the factors, and the reliability of the indicators. Gerbing and Anderson (1985) found little bias in the ML estimator with the exception of the parameter relating factors that were defined by only two indicators. Curran's (1994) Monte Carlo simulation of confirmatory factor analysis models found the ML estimator to exhibit little bias in correctly specified models across his sample sizes of 100, 200, and 1,000. The evidence from these factor analysis simulations are consistent in their finding of little bias in the ML estimator for samples sizes of at least 100, though bias in smaller sample sizes is less clear. We were not able to locate any Monte Carlo simulation studies that examined a general SEM with structural relations between latent variables, so we have no evidence on the bias in these types of models.

2SLS Estimator

The term *2SLS estimator* applies to a family of different methods that are in use in latent variable SEMs. For instance, Jöreskog (1983) builds on the factor analysis work of Hägglund (1982) and Madansky (1964) to propose a 2SLS estimator that provides the starting values in the Jöreskog and Sörbom (1993) LISREL software. The Jöreskog–Hägglund–Madansky 2SLS estimator first estimates a factor analysis model that assumes no correlated errors, then estimates the covariance matrix of the factors, and last applies another variant of the 2SLS estimator to this covariance matrix of the factors to estimate a latent variable structural model. This 2SLS estimator differs from the 2SLS estimator of Lance, Cornwell, and Mulaik (1988) in that the latter authors estimate the factor analysis model with the ML estimator and then use a 2SLS estimator in a similar manner to that of Jöreskog, Hägglund, and Madansky.

The 2SLS estimator for latent variable SEM in Bollen (1996a, 1996b, 2001) differs from both of these in that it permits correlated errors across equations, does not require that the measurement model be estimated first (or at all), estimates intercepts, and provides the asymptotic covariance matrix of the estimator for significance testing. Significance tests of multiple coefficients in the same or different equations are possible using the estimated asymptotic covariance matrix.[3] Furthermore, when there are no latent variables or measurement error, Bollen's (1996a, 2001) 2SLS estimator is equivalent to the original Theil (1953, 1961)–Basmann (1957) 2SLS estimator for simultaneous equation models. This equivalence is valuable since it enables us to draw on the econometric literature that has studied the 2SLS estimator in the context of simultaneous equation models.

Alternatively, if there is a factor analysis model with no correlated errors and no intercepts, then Bollen's (1996b, 2001) 2SLS estimator is equivalent to the Jöreskog–Hägglund–Madansky 2SLS estimator. For the remainder of the article, we examine Bollen's version of the 2SLS estimator while recognizing that other versions of 2SLS have been proposed. This version of the 2SLS estimator has been applied to estimation of SEMs that are nonlinear in the latent variables (Bollen 1995; Bollen and Paxton 1998), to comparison of nonnested latent variable models (Oczkowski 2002), to models with heteroscedastic disturbances (Bollen 1996a), and to higher order factor analysis models (Bollen and Biesanz 2002).

To apply this 2SLS to equation (1), each latent variable must have a single observed variable to scale it such that

$$\begin{aligned} \boldsymbol{y}_1 &= \boldsymbol{\eta} + \boldsymbol{\varepsilon}_1, \\ \boldsymbol{x}_1 &= \boldsymbol{\xi} + \boldsymbol{\delta}_1, \end{aligned} \tag{3}$$

where $\boldsymbol{y}_1$ and $\boldsymbol{x}_1$ are the vectors of scaling indicators. We can then reexpress equation (3) as

$$\begin{aligned} \boldsymbol{\eta} &= \boldsymbol{y}_1 - \boldsymbol{\varepsilon}_1, \\ \boldsymbol{\xi} &= \boldsymbol{x}_1 - \boldsymbol{\delta}_1. \end{aligned} \tag{4}$$

Following Bollen (2001:122–4), we can rewrite the latent variable and measurement models as

$$\begin{aligned} \boldsymbol{y}_1 &= \alpha_\eta + \boldsymbol{B}\boldsymbol{y}_1 + \boldsymbol{\Gamma}\boldsymbol{x}_1 + \boldsymbol{\varepsilon}_1 - \boldsymbol{B}\boldsymbol{\varepsilon}_1 - \boldsymbol{\Gamma}\boldsymbol{\delta}_1 + \boldsymbol{\zeta}, \\ \boldsymbol{y}_2 &= \alpha_{y2} + \boldsymbol{\Lambda}_{y2}\boldsymbol{y}_1 - \boldsymbol{\Lambda}_{y2}\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2, \\ \boldsymbol{x}_2 &= \alpha_{x2} + \boldsymbol{\Lambda}_{x2}\boldsymbol{x}_1 - \boldsymbol{\Lambda}_{x2}\boldsymbol{\delta}_1 + \boldsymbol{\delta}_2, \end{aligned} \tag{5}$$

where $\boldsymbol{y}_2$ and $\boldsymbol{x}_2$ are the vectors of the remaining nonscaling indicators. To simplify the presentation of the 2SLS estimator, we focus on a single equation. Consider the jth equation from $\boldsymbol{y}_1$ as

$$y_j = \alpha_{\eta j} + \boldsymbol{B}_j\boldsymbol{y}_1 + \boldsymbol{\Gamma}_j\boldsymbol{x}_1 + u_j, \tag{6}$$

where y_j is the jth y from $\boldsymbol{y}_1$, $\alpha_{\eta j}$ is the corresponding intercept, $\boldsymbol{B}_j$ is the jth row from $\boldsymbol{B}$, $\boldsymbol{\Gamma}_j$ is the jth row from $\boldsymbol{\Gamma}$, and u_j is the jth element from $\boldsymbol{u}$ where $\boldsymbol{u} = \boldsymbol{\varepsilon}_1 - \boldsymbol{B}\boldsymbol{\varepsilon}_1 - \boldsymbol{\Gamma}\boldsymbol{\delta}_1 + \boldsymbol{\zeta}$. Following Bollen (1996b:113), we define $\boldsymbol{A}_j$ to be a column vector that contains $\alpha_{\eta j}$ and all the nonzero elements of $\boldsymbol{B}_j$ and $\boldsymbol{\Gamma}_j$. Let N represent the number of cases in a sample. Form $\boldsymbol{Z}_j$ as an N-row matrix that has the sample elements of all of the $\boldsymbol{y}_1$ and $\boldsymbol{x}_1$ variables that have nonzero coefficients in equation (6), form $\boldsymbol{y}_j$ as an $N \times 1$ vector of the sample values of y_j, and make $\boldsymbol{u}_j$ an $N \times 1$ vector of the values of u_j. Using these definitions, we rewrite equation (6) as

$$\boldsymbol{y}_j = \boldsymbol{Z}_j\boldsymbol{A}_j + \boldsymbol{u}_j. \tag{7}$$

The 2SLS estimator provides a consistent estimator of these parameters, but it requires IVs. The IVs are observed variables that must satisfy several conditions. First, the observed variable must not be directly or indirectly influenced by any of the disturbances in (7). Second, if the observed variable is endogenous, its disturbance must be uncorrelated with all of the disturbances in (7). Finally, the observed variable must correlate with the observed variables that they will be predicting. These model-implied IVs are determined by the structure of the model. In other words, if the model is correct, there is no ambiguity as to which variables are uncorrelated with the disturbances of an equation and hence a potential IV. In contrast to most treatments of IVs, we do not assume that the researcher searches for IVs outside of the model. Rather, all IVs are found among the observed variables that are part of the model and that satisfy the conditions of being an IV for a particular equation. This approach also differs from the typical simultaneous equation approach in that endogenous observed variables are sometimes suitable as IVs.

So that we can identify the equation parameters, there must be at least as many IVs as there are variables that they will be replacing, and the squared multiple correlation coefficients from the regression of the variables to be replaced on the IVs should be reasonably high (e.g., $> .10$). Bollen (1995, 1996b) and Bollen and Paxton (1998) have more details on the selection of IVs along with several examples. Bollen and Bauer (2004) provide an automated procedure that chooses the model-implied IVs for an equation, but to conserve space we do not repeat these discussions here.

If we use $\boldsymbol{V}_j$ to represent an N-row matrix containing the IVs,[4] then the 2SLS estimator is

$$\hat{\boldsymbol{A}}_j = (\hat{\boldsymbol{Z}}_j'\hat{\boldsymbol{Z}}_j)^{-1}\hat{\boldsymbol{Z}}_j'\boldsymbol{y}_j, \tag{8}$$

where

$$\hat{\boldsymbol{Z}}_j = \boldsymbol{V}_j(\boldsymbol{V}_j'\boldsymbol{V}_j)^{-1}\boldsymbol{V}_j'\boldsymbol{Z}_j. \tag{9}$$

The 2SLS estimator, $\hat{A}_j$, has an *estimated* asymptotic covariance matrix equal to

$$\text{acov}(\hat{\boldsymbol{A}}_j) = \hat{\sigma}^2_{u_j}(\hat{\boldsymbol{Z}}_j'\hat{\boldsymbol{Z}}_j)^{-1} \tag{10}$$

with $\hat{\sigma}^2_{u_j} = (\boldsymbol{y}_j - \boldsymbol{Z}_j\hat{\boldsymbol{A}}_j)'(\boldsymbol{y}_j - \boldsymbol{Z}_j\hat{\boldsymbol{A}}_j)/N$. An important characteristic of 2SLS is that this is a noniterative procedure. All parameter estimates and standard errors are computed in closed form, and a solution always exists. We will later contrast this aspect of 2SLS with the iterative process that is required for ML estimation. We need not assume normality of the observed variables or disturbances for the asymptotic covariance matrix to be accurate (Bollen 1996b).

Asymptotic properties of 2SLS. The 2SLS estimator shares several desirable asymptotic properties with the ML estimator. The 2SLS estimator is consistent (Basmann 1957; Theil 1958), asymptotically unbiased (Theil 1958; Richardson 1970), and asymptotically normally distributed (Basmann 1960). As equation (10) shows above, 2SLS has an asymptotic covariance matrix that we can use in statistical significance testing. Furthermore, 2SLS is an asymptotically efficient estimator among the single-equation limited information estimators (Bowden and Turkington 1984:110–11).

Finite sample properties of 2SLS. The number of finite moments of the 2SLS estimator is equal to the degree of overidentification of an equation (Mariano 1972; Hatanaka 1973; Kinal 1980). This means that the 2SLS estimator for an equation that has two more IVs than required to identify the equation will have first and second finite moments. However, there will be no finite moments of the 2SLS estimator if an equation is exactly identified.

In finite samples, the 2SLS estimator is influenced by the degree of overidentification (number of extra IVs) for an equation such that bias increases but the variability decreases with the number of additional IVs (e.g., Sawa 1969:936; Mariano 1982:523). For instance, the bias is less when an equation has one additional IV compared to two additional IVs. However, Buse (1992) suggests that the relation to bias is more complicated than just the number of IVs. He finds that if the additional IVs improve the squared multiple correlation of the first-stage regression, then the bias need not occur.

Other things being equal, the precision of the 2SLS estimator is lower in an equation as the number of endogenous variables increases (Phillips 1983). Also, the approach to normality of the estimator is influenced by the magnitude of the coefficient, the sample size, and the degree of overidentification such that larger coefficients and greater overidentification slow the approach to normality while larger sample sizes accelerate the approach to normality (Phillips 1983).

Summary of Properties of 2SLS and ML

Our brief review of the properties of 2SLS and ML show that both estimators have desirable asymptotic properties: consistency, asymptotic unbiasedness, asymptotic normality, and estimable asymptotic covariance matrix. The 2SLS estimator is an asymptotically efficient estimator among single-equation estimators, and the ML estimator is an asymptotically efficient estimator among system-wide, asymptotically unbiased estimators. As a full information estimator that utilizes information from the whole system of equations, we expect the asymptotic efficiency of the ML estimator to be better than 2SLS provided that the model is correct and that there is no excess multivariate kurtosis in the data. But practically speaking, we need to interpret these asymptotic results with caution. We do not know how well they characterize the sample sizes and models that are typical in practice. As we noted earlier, there is little information on the finite sample properties of the ML. There is more research on the 2SLS estimator, but still some gaps in knowledge. Finally, the relative efficiency is uncertain when there is excess multivariate kurtosis or misspecification in the model.

Monte Carlo simulations that examine both 2SLS and ML estimators in latent variable SEMs are rare. The only exceptions we have found are two simulation studies by Hägglund (1983) and Brown (1990). Both studies include the ML and the Jöreskog–Hägglund–Madansky 2SLS estimators and use confirmatory factor analysis models with uncorrelated disturbances. The Jöreskog–Hägglund–Madansky 2SLS and Bollen's (1996b) 2SLS estimators are equivalent under these restrictive conditions and assumptions.[5] Hence, Hägglund's (1983) and Brown's (1990) results are relevant to our study.[6]

Hägglund (1983) studied two-factor analysis models, one with two factors and 6 indicators and another with three factors and 12 indicators. The factor loadings were moderately high (.889 and .714), and the sample sizes were 100, 200, 400, and 1,000. Considering a perfectly specified model, he found that the biases were small for ML and 2SLS. At sample size 100, there tended to be a slight negative bias for the factor loadings for 2SLS and a slight positive bias for ML. Hägglund also found small differences in the standard deviations of the factor loadings across the estimators. The ML gave somewhat smaller standard deviations than 2SLS for $N > 100$. Overall, Hägglund concluded that the differences across estimators were small.

Brown's (1990) factor analysis model had eight indicators and two factors. The factor loadings ranged from .4 to .9, and he looked only at a sample size of 500. He varied the distribution of the indicators from normal to highly skewed. When the indicators came from normal to moderately nonnormal distributions, Brown found that 2SLS and ML performed similarly under correct specification. In the cases with the highest degrees of nonnormality, Brown found that the ML factor loadings had less bias than

2SLS, but the ML estimate of the covariance of the factors was more biased than the covariance estimate of 2SLS.

Kaplan (1988) conducted a small simulation study that included a general latent variable SEM rather than just a factor analysis model and looked at misspecification in the model. He found that the ML estimator tended to spread bias throughout a model and 2SLS was better at isolating specification error. However, Kaplan used the Jöreskog–Hägglund–Madansky 2SLS estimator, and for his model this estimator differs from Bollen's (1996b) 2SLS estimator. Lance et al. (1988) also looked at a general SEM with latent variables and compared the ML estimator to their version of a 2SLS model. They also found greater robustness to misspecified models for 2SLS than for ML. But we caution that their 2SLS estimator differs from Bollen's (1996b) for their models.

The econometric literature on simultaneous equations, a special case of SEM, has the largest number of Monte Carlo simulations that looked at the ML and 2SLS estimators. Two things stand out about these econometric simulation results. One is that the ranking of the relative performance of the two estimators varies across the studies. The other is that the overwhelming majority of these simulations look at sample sizes that are much smaller than is typical in latent variable SEMs.

Consider first the inconsistency in results of the simulation studies. Donatos and Michailidis (1996) simulated a two-endogenous-variable model with four exogenous variables. They found that 2SLS had less bias than ML (or than OLS) for normal and nonnormal disturbances. Compare this to Cragg (1967), who had three endogenous variables and seven exogenous variables in a simultaneous equation model and found that ML had lower median bias than 2SLS. In Mikhail's (1975) simulation study of two models that differ in the degree to which the disturbances are correlated, Mikhail found differences in the 2SLS and ML estimators. The ML estimator showed slightly less bias than 2SLS for some coefficients when the disturbance correlation was low, whereas these estimators showed no significant differences in the magnitude of bias when the disturbance correlation was high. Johnston (1972:410) surveyed simulation studies of ML, 2SLS, and other estimators for simultaneous equations and concluded that

> . . . the consistent estimators [e.g., 2SLS and ML] show some finite sample bias, but the means of the sampling distributions are not usually significantly different from the true values, and the variation in the bias of the consistent estimators is neither large nor systematically in favor of one consistent estimator vis-*à*-vis another. (P. 410)

Thus, the studies do not provide compelling evidence for ranking one estimator above the other.

The other aspect of these econometric Monte Carlo simulation studies comparing the ML and 2SLS estimators is that the majority used extremely small sample sizes, particularly when considering sample sizes commonly encountered in applied social science research. Indeed, the modal sample size from these prior simulation studies was $N = 20$, and it is rare to find studies with sample sizes beyond 100. In contrast, most latent variable SEMs have sample sizes of 100 or more. This small sample size in the simulation studies is important to remember since we know from the analytic work reviewed above that in some cases the 2SLS estimator's bias increases with the degree of overidentification of an equation when the sample is small. Thus, the typically small N in the econometric simultaneous equation simulations raises questions about the generalizability of their findings to larger sample sizes and to the latent variable models that are of primary concern to us.

Research Issues

Based on the analytical work and simulation research that we have reviewed, we briefly outline the research issues and our expectations for this simulation experiment:

1. *When the model is correctly specified, we expect the bias of all estimators to be negligible in the larger sample sizes, but we expect some degree of bias at the smaller sample sizes.* Both the ML and 2SLS estimators are consistent and asymptotically unbiased estimators for a well-specified model. Although it is difficult to determine at what sample size the estimators will be essentially unbiased, based on previous simulations we expect that this will occur when the sample contains several hundred cases. Simulations of the ML estimator suggest some bias in the smaller sample sizes (Boomsma 1982). Analytic and simulation work suggests a negative bias for the 2SLS with a high degree of overidentification at the smaller sample sizes, though Buse's (1992) work suggests that this will not always occur. We expect the 2SLS estimators with lower degrees of overidentification to exhibit less bias at the smaller sample sizes.
2. *When the model is incorrectly specified, we expect the bias of the ML estimator to be greater than that of the 2SLS estimators and that this bias will remain present across all sample sizes.* Analytic work

on the bias in the ML estimator in latent variable models is not developed sufficiently to predict the exact conditions under which bias will spread beyond a misspecified equation. However, in the first part of our analysis, we will present results based on the population covariance matrix that reveal the bias of each estimator under different structural misspecifications when sampling variability is removed. Furthermore, simulation work by Cragg (1968) suggests that the ML estimator will be more susceptible to spreading bias in one part of a multiequation system to other parts. Analytical work presented in Bollen (2001) shows that if a given equation is not misspecified, it will be robust to misspecification in other equations as long as the misspecifications elsewhere in the model do not change the model-implied IVs for the given equation. Among the 2SLS estimators, we expect that the ones that use fewer IVs will exhibit greater robustness than those with many more IVs.

3. *When the model is correctly specified, we expect the variance of the ML estimator to be smaller than that of the 2SLS estimators in the larger sample sizes. We have no prediction at the smaller sample sizes.* Analytical work on the ML shows that it is asymptotically efficient among asymptotically unbiased estimators when the model is correctly specified (Browne 1984). The 2SLS estimators are asymptotically unbiased, but as limited information estimators, they use less information in developing the estimates than does the FIML estimator. In large samples we expect that the ML estimator will have lower variances. However, we have little to guide us on the magnitude of the differences in variances and how large the sample must be to see these differences. Furthermore, we do not know the relative variances of these estimators in smaller sample sizes.
4. *When the model is incorrectly specified, we have no prediction about the relative size of the variances of the estimators.* The asymptotic variance properties of the ML and 2SLS estimators are derived under the assumption that the model is correctly specified. Although it is tempting to assume that the full information nature of the ML estimator will still lead to lower asymptotic variances than the 2SLS estimators, we have no evidence to support this hypothesis. We have no prediction about the variances in misspecified models at the smaller sample sizes.
5. *When the model is correctly specified, we expect that the Type I error rates for the null hypothesis that the estimate equals the true parameter will be accurate in the larger sample sizes for both ML*

and 2SLS. The accuracy of the Type I error rate of testing a true null hypothesis about coefficients depends on the accuracy of the asymptotic standard errors, the asymptotic unbiasedness of the estimator, and the accuracy of the asymptotic normality of the estimators. Analytical results suggest that the Type I error rates will be accurate in large samples, though at what sample size this occurs is not known. We would expect that the accuracy of the Type I error rates for the ML and 2SLS estimators will be good in samples of several hundred cases or more. We have no evidence to predict the accuracy in smaller sample sizes.

6. *When the model is incorrectly specified, we expect that the Type I error rates will be inaccurate for those coefficients that are not robust to the specification error. Generally, we expect the ML estimator to have less accurate Type I error rates than 2SLS.* When a model is misspecified, at least some of the coefficient estimators will be biased even in large samples. As a result, hypothesis tests that use the population parameter as the null hypothesis will reject the null hypothesis at the wrong error rate. Given our expectation that the 2SLS estimator will be more robust to specification errors than the ML, we expect that the 2SLS estimator Type I error rate will be accurate for more coefficient hypothesis tests than is true for the ML.

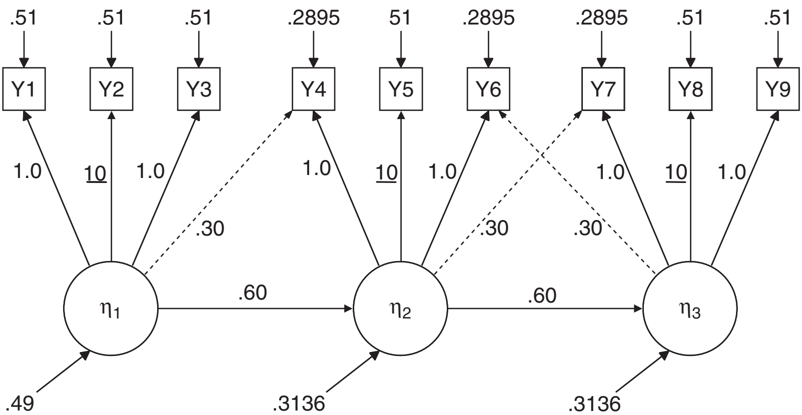
We used a Monte Carlo simulation experimental design to examine the preceding six issues empirically. To isolate the specific impact of model misspecification across a large number of conditions, we focused on data drawn from a multivariate normal distribution.[7]

Design of the Monte Carlo Experiment

Model

One of our guiding goals for this study was to identify population models that would allow us to maximize the external validity of resulting findings (for further details, see Paxton et al. 2001). To accomplish this, we reviewed key journals in several areas of social science research to catalog the most common types of SEM applications. Using this information in combination with our own modeling experience, we selected our model in Figure 1. We designed the model to represent features that are commonly encountered in social science research. It is based on the frequent situation of having applications with relatively few latent variables and few indicators

Figure 1
Simulation Model



Note: Dashed lines represent paths that are omitted in misspecified models.

per latent variable. The model contains nine measured variables and three latent factors with three or four indicators per factor.[8] Furthermore, we use one correct and three incorrect specifications of the model.

All our models were fitted separately for one correct specification of the model and three misspecifications. Our first specification is *properly* specified such that the estimated model matches the population model (Specification 1). The three structurally misspecified models were Specification 2, omitting only the path from η_2 to variable 7 (Y7); Specification 3, jointly omitting the paths from η_2 to variable 7 (Y7) and from η_3 to variable 6 (Y6); and Specification 4, jointly omitting the paths from η_2 to variable 7 (Y7), from η_3 to variable 6 (Y6), and from η_1 to variable 4 (Y4).

In the first part of the Results section, we report the coefficient values when analyzing the population covariance matrix for all four specifications. This will reveal which coefficients are biased, even when the population is available, and the degree of bias. Since we do not have population information in practice, we then look at the sampling results across a range of sample sizes. Due to the high degree of similarity of results across many experimental conditions combined with space constraints, we report only the results of the first, correctly specified model (Specification 1) and another moderately misspecified model (Specification 3) in graphical and tabular form for the Monte Carlo simulations. The pattern of results is similar in the auxiliary misspecifications that have fewer (Specification 2) or more (Specification 4) omitted paths.

Model Parameterization

Parameter values were selected to result in a range of effect sizes (e.g., communalities and R^2 values ranging from 49 percent to 72 percent) and for the misspecified conditions to lead to both a wide range of power to detect the misspecifications (e.g., power estimates computed using the method of Satorra and Saris [1985] ranged from .07 to 1.0 across all sample sizes) and a range of bias in parameter estimates (e.g., absolute bias ranged from 0 percent to 37 percent). See Paxton et al. (2001) for a comprehensive description of our model parameterization.

Estimators

We examined four estimators: the ML estimator and three 2SLS estimators. We use the ML estimator from equation (2). The three 2SLS estimators differ only in the number of IVs used. We vary the number of IVs because our earlier review of the analytic work suggested that the degree of bias and the variance of the 2SLS in simultaneous equation models can differ depending on the number of IVs, especially when N is small to moderate. The 2SLS-ALLIV uses all possible IVs for an equation. This leads to the highest degree of overidentification for an equation. The 2SLS-OVERID1 and 2SLS-OVERID2 use reduced sets of IVs such that the former has one more IV than is needed for identification and the latter uses two more IVs than required for identification. In these latter two cases, we choose IVs from the list of all eligible IVs, selecting those that lead to the largest increment in the R^2 from the first-stage regression.

Sample Size

We chose seven sample sizes to represent those commonly encountered in applied research, and these range from very small to large: 50, 75, 100, 200, 400, 800, and 1,000.

Data Generation and Estimation

We used the simulation feature in Version 5 of EQS (Bentler 1995) to generate the raw data and EQS's ML estimation to fit the sample models.

Population values for each parameter were used as initial start values, and a maximum of 100 iterations was allowed to achieve convergence.

Distribution

We generated data from a multivariate normal distribution.

Replications

There were a total of 28 experimental conditions (four specifications and seven sample sizes), and we generated up to 500 replications for each condition.

Convergence

We eliminated any replication that failed to converge within 100 iterations or that did converge but resulted in an out-of-bounds parameter estimate (e.g., "Heywood cases"). We adopted this strategy because the research hypotheses were directly related to proper solutions in SEM, and the external validity of findings would be threatened with the inclusion of improper solutions. To maintain 500 replications per condition, we generated an initial set of 650 replications for the two smallest sample sizes and 550 for the larger sample sizes. We then fit the perfectly specified model to the generated data and selected the first 500 proper solutions. The samples that generated these solutions were used to estimate the remaining misspecified models. This resulted in 500 proper solutions for all properly specified and most misspecified experimental conditions, but there were several misspecified conditions that resulted in fewer than 500 proper solutions at the smaller sample sizes. See Table 1 for details on the samples generated, the samples used, and mean values for a variety of common fit indicators.

Only the ML solutions had any failures to converge or have proper solutions. The 2SLS estimators are noniterative and as such do not face this difficulty. This does raise a limitation of our simulation that should be kept in mind: Nonconvergence or improper solutions are likely to represent some of the worst estimates for the ML estimator. By removing them from the simulation, we are in essence trimming the ML outliers, while no such trimming is done for the 2SLS estimators. As with all prior research in comparing 2SLS and ML, our empirical results tend to favor the ML

Table 1
Number of Replications and Mean Fit Statistics by Sample Size and Specification

	Sample Size						
	50	75	100	200	400	800	1,000
Total samples generated	650	650	550	550	550	550	550
Samples generating improper ML solutions	104	28	16	0	0	0	0
Percentage of samples rejected	16	4	3	0	0	0	0
Specification 1							
Number of samples	500	500	500	500	500	500	500
Mean model χ^2	23.64	23.01	23.32	22.58	22.18	21.89	21.74
Mean model χ^2 p value	.44	.45	.45	.47	.49	.50	.51
Specification 2							
Number of samples	471	493	497	500	500	500	500
Mean model χ^2	25.70	25.30	25.97	26.69	30.14	36.59	38.49
Mean model χ^2 p value	.40	.41	.39	.36	.26	.12	.11
Specification 3							
Number of samples	463	492	495	500	500	500	500
Mean model χ^2	27.50	27.60	28.95	31.84	39.32	54.26	60.19
Mean model χ^2 p value	.37	.36	.32	.24	.11	.02	.01
Specification 4							
Number of samples	467	492	495	500	500	500	500
Mean model χ^2	31.32	32.42	35.84	44.17	63.50	100.01	117.56
Mean model χ^2 p value	.29	.25	.18	.07	.01	.00	.00

Note: ML = maximum likelihood.

estimates relative to the 2SLS estimates to an unknown degree. We discuss the potential implications of this in greater detail later.

Results

Analysis of Population Covariance Matrix

The simulation methodology gives us the advantage of knowing the population covariance matrix of the observed variables. Analysis of this population covariance matrix with the ML and 2SLS estimators reveals the biases of the estimators free of sampling error. Table 2 presents the parameter values when using the ML and 2SLS estimators on the population covariance matrix under all four specifications. The 2SLS-ALLIV, 2SLS-OVERID1, and 2SLS-OVERID2 all give the same values of the

Table 2
Percentage of Biases When ML and 2SLS Estimators Are Applied to Population Covariance Matrix of Observed Variables

	ML				2SLS			
	Specification				Specification			
Parameters	1	2	3	4	1	2	3	4
λ_{21}								
λ_{31}								
λ_{41}			+23%	—			0%	—
λ_{42}			−6%	+20%			0%	+21%
λ_{62}		−4%	+28%	+23%		0%	+23%	+23%
λ_{63}		+10%	—	—		0%	—	—
λ_{72}		—	—	—		—	—	—
λ_{73}		+33%	+33%	+34%		+28%	+28%	+28
λ_{93}								
β_{21}			−7%	+5%			0%	0%
β_{32}		+8%	+22%	+18%		0%	+13%	+13%

Note: ML = maximum likelihood; 2SLS = two-stage least squares; blank cells = no bias in either estimator; — = omitted parameter in specification; shading = parameter from misspecified equation.

parameters when analyzing the population covariance matrix. Therefore, we can treat all the 2SLS results the same, and we show these results on the right side of Table 2. The main cell entries are the percentage of bias in the parameter value. Blank cells represent cells where both estimators are unbiased. The "—" stands for parameters that are omitted within a model specification. Shaded cells represent parameters that are part of a structurally misspecified equation. The first column under Specification 1 has neither bias nor any shaded area for either estimator since Specification 1 is the correct model and both estimators should and do provide the population parameters.

The rest of the table shows the results in models that contain structural misspecifications. There are several interesting findings. First, the nonzero percentage biases that occur outside the shaded cells are biases that occur in equations that are correctly specified. That is, as predicted, there is evidence of bias that spreads beyond the misspecified equation. It also is clear from Table 1 that the ML estimator spreads bias throughout the model to a greater extent than the 2SLS estimator, as anticipated from past research. For instance, Specification 3 omits one cross-loading from the sixth indicator and one from the seventh indicator equations. Even though the fourth indicator's equation is correct and the latent variable equations are correct, their parameter values are biased by the misspecification of the sixth and seventh indicator equations. In contrast to these four biased coefficients from the ML estimator, in the 2SLS estimator only the β_{32} parameter is biased. In addition, its bias in 2SLS is 13 percent compared to 22 percent for the same parameter using ML. More generally, when the 2SLS estimator shows bias, it typically is smaller than the ML estimator for the same parameter. And we know from the analytic conditions in Bollen (2001) that this bias in 2SLS occurs only when the misspecification leads a researcher to use a set of IVs that differs from the model-implied IVs of the true model.

Overall, this analysis of the population covariance matrix reveals two general results. First, the 2SLS estimator better isolates biases caused by structural misspecifications to the misspecified equation than does ML. Second, even when there is bias in the 2SLS estimator, its magnitude is generally smaller than the ML estimator.

Of course, in practice we do not have the population covariance matrix. These results do not tell us the relative bias and variability of the 2SLS or ML estimator in small to moderate samples. Nor do these population results give us the relative bias of the 2SLS estimators when different numbers of IVs are used. To address these important and practical issues, we examine the behavior of these estimators in finite samples using Monte Carlo simulations.

Bias of Estimators

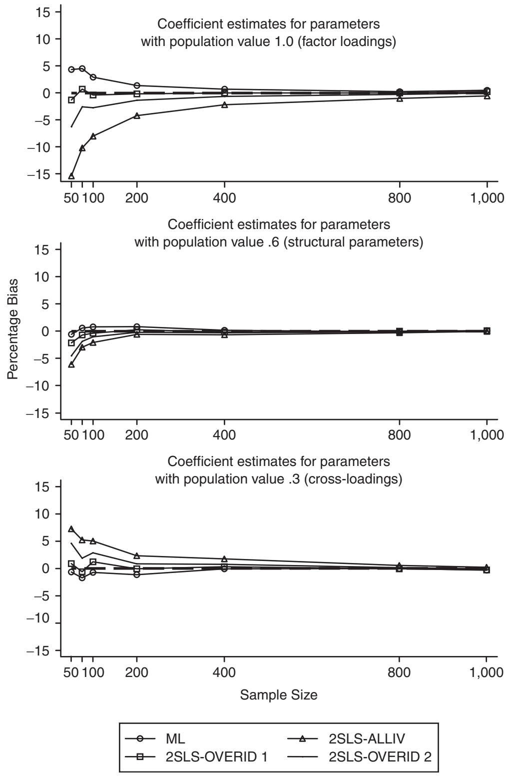
Our major prediction for the correct model was that all estimators would have negligible bias in large samples (Research Issue 1). Figure 2 plots the relative bias for each estimator, and percentage of relative bias is calculated as

$$\%\ \text{Bias} = \left(\frac{\hat{\theta} - \theta}{\theta}\right) \times 100,$$

where θ is the coefficient parameter and $\hat{\theta}$ is the estimate of the parameter.[9] We consider absolute values of *% Bias* exceeding 10 percent to reflect meaningful bias.

We calculate the mean of the *% Bias* for each of the three parameter values in the model: .3, .6, and 1.0.[10] Figure 2 shows these *% Bias* plots by sample size for Specification 1, the correctly specified model. As expected,

Figure 2
Percentage of Bias in Four Estimators by Sample Size and Parameter Value in Specification 1, the Correctly Specified Model



Note: ML = maximum likelihood; 2SLS = two-stage least squares.

the bias in all four estimators is negligible for sample sizes of 400 or more. With the exception of the 2SLS-ALLIV, even in the smallest sample sizes the estimators exhibit bias of at most 6 percent. The greatest *% Bias*es are for the 2SLS-ALLIV estimator that is as negative as −15 percent for the 1.0 loadings and as positive as 7 percent for the .3 cross-loading parameters. The magnitude of the *% Bias*es is not too different for the 2SLS-OVERID1 and the ML estimators for the .3 and .6 parameter values. However, the bias for the ML is larger than that for the 2SLS-OVERID1 for the parameters that are 1.0. Compared to both ML and 2SLS-OVERID1, the 2SLS-OVERID2 estimator has slightly greater *% Bias* at sample sizes of 100 or less. Overall, the 2SLS-OVERID1 has the least bias across sample sizes and parameter values, followed closely by the ML and then the 2SLS-OVERID2 estimator. The 2SLS-ALLIV has the most bias when the model is correct. However, these biases are hardly evident in any estimator when there are several hundred cases or more for the correct specification (Specification 1).

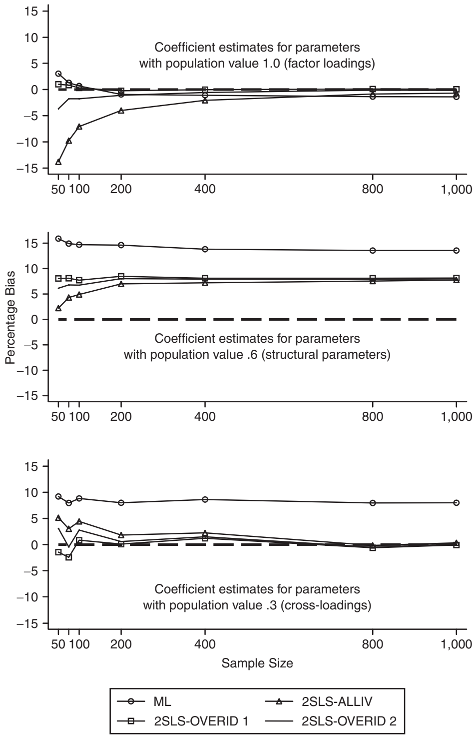
We next turn to the case where the model is misspecified. As outlined above, our misspecified model (Specification 3) omits two paths, the path from η_2 to variable 7 (Y7) and the path from η_3 to variable 6 (Y6), both of which have population values of .3. The rest of the model is correct. Figure 3 plots the *% Bias* for the four estimators across the different sample sizes and parameter values. Figure 3 includes only the coefficients from equations that are correctly specified. We do this to determine the degree to which specification error in one equation has consequences for the estimates of parameters from equations that are correctly specified. We expect that the ML estimator is more likely to spread the bias due to misspecification because it is a full information estimator. This expectation is met for the ML estimator for the .3 cross-loading and .6 latent variables' coefficient values, and it is a bias that persists at the larger sample sizes.

However, the ML estimator for the 1.0 factor loadings is more robust than it is for the .3 and .6 coefficients. The 2SLS-ALLIV estimator exhibits less bias than the ML for the .3 and .6 coefficients but generally has greater (negative) bias for the 1.0 coefficient. Finally, the 2SLS-OVERID1 and 2SLS-OVERID2 estimators are quite robust for the .3 and 1.0 coefficients. These same two estimators exhibit specification bias for the .6 coefficient, but the bias is roughly half of that for the ML. A more detailed examination of the two .6 coefficients revealed that the 2SLS estimators are robust for the path from η_1 to η_2. All the bias in the 2SLS estimators are in the estimates of the path from η_2 to η_3. Overall, these results in the misspecified model suggest that the 2SLS-OVERID1 has the best performance, closely followed by the 2SLS-OVERID2 estimator. The 2SLS-ALLIV comes next except for the 1.0 factor loadings in the smaller sample sizes.[11]

We also examined the *% Bias* of coefficient estimates from the equations in which the specification errors occurred. All estimators were biased in these situations where a path is omitted. However, the magnitude of bias was greatest for the ML estimator compared to the 2SLS estimators.

If we consider all four estimators across correct and incorrect specifications, it appears that the 2SLS-OVERID1 estimator exhibits the least bias.

Figure 3
Percentage of Bias in Four Estimators by Sample Size and Parameter Value in Specification 3, in Which Two Paths Are Omitted



Note: ML = maximum likelihood; 2SLS = two-stage least squares.

The 2SLS-OVERID2 estimator has slightly larger bias than ML at smaller samples sizes for the correct model specification. However, its greater robustness than ML under the misspecified model would lead us to put 2SLS-OVERID2 second in its resistance to bias. The relative ranking of 2SLS-ALLIV and ML is complicated. On one hand, the 2SLS-ALLIV estimator has greater robustness than the ML estimator for the .3 and .6 coefficients in the misspecified models. On the other hand, in the correct specification at the smaller sample sizes, the ML has less bias than the 2SLS-ALLIV estimator. At the larger sample sizes, the greater robustness to misspecification would give the edge to the 2SLS-ALLIV compared to ML.

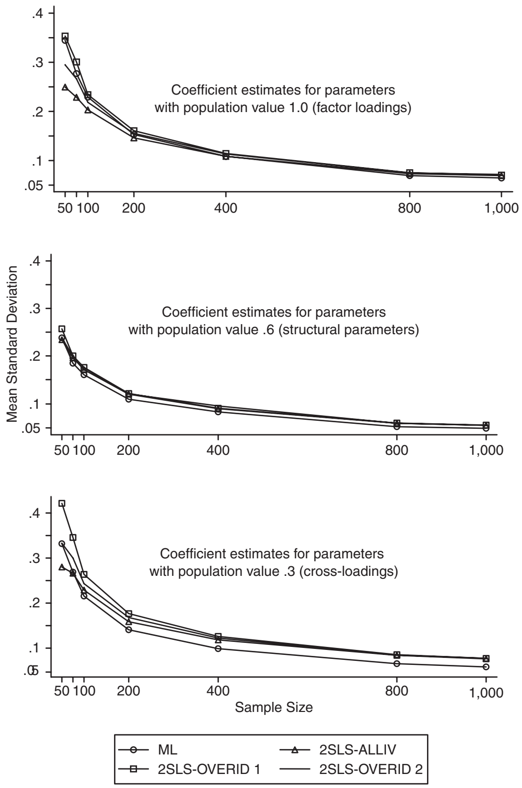
Variability of Estimators

In addition to bias, it is important to know the variability of an estimator (see preceding Research Issues 3 and 4). In this section, we examine the mean of the standard deviations for coefficient estimators across the replications in the simulation. Since the standard deviations are influenced by the magnitude of the parameter, we calculate the mean standard deviations separately by the magnitude of the coefficients. Mean standard deviations are constructed by calculating the standard deviation for each parameter across the replications, then averaging the standard deviations for a given parameter value. Asymptotic theory predicts that with a correctly specified model the ML estimator should have smaller variances in large samples than do the 2SLS estimators (see Research Issue 3), though we do not know whether this would hold at smaller sample sizes or in misspecified models (Research Issue 4).

Figure 4 plots the mean standard deviations for each estimator and for each parameter across the different sample sizes for Specification 1, the correctly specified models. Based on the asymptotic efficiency of the ML estimator, we predicted that its mean standard deviations would be smaller than those of the 2SLS estimators in large samples. This pattern is most evident for the parameters with population values of .3 (the cross-loadings), where the ML estimator has a lower mean standard deviation than the other estimators at all but the smallest sample size. However, we did not expect to find how close in magnitude were the standard deviations of the ML and 2SLS estimators for the other parameters, especially in the larger samples ($\geq$ 400). Indeed, with the exception of the population value of .3, it is hard to distinguish the lines representing the standard deviations for the different estimators. The variability advantage of ML is even less evident at the smaller sample sizes. In fact, the 2SLS-ALLIV estimator has a smaller mean standard deviation than ML for the 1.0 coefficients at sample sizes of 400 or smaller. This same pattern is evident in the graphs of the mean standard deviations for Specification 3, the misspecified model, in Figure 5, so we do not repeat the description of the graphs.

In sum, we find that the mean standard deviations of all four estimators are quite close in the larger sample sizes in the correct and misspecified models. Exceptions are the parameters representing cross-loadings (population value of .3), where the ML estimator shows modestly smaller variance. In the smaller sample sizes, we see some slight differences with the 2SLS estimator sometimes having a smaller standard deviation. The number of instruments used in the 2SLS estimator is negatively related to the variance

Figure 4
Mean Standard Deviation of Estimates From Four Estimators by Sample Size for Parameter Estimates From Specification 1, the Correctly Specified Model

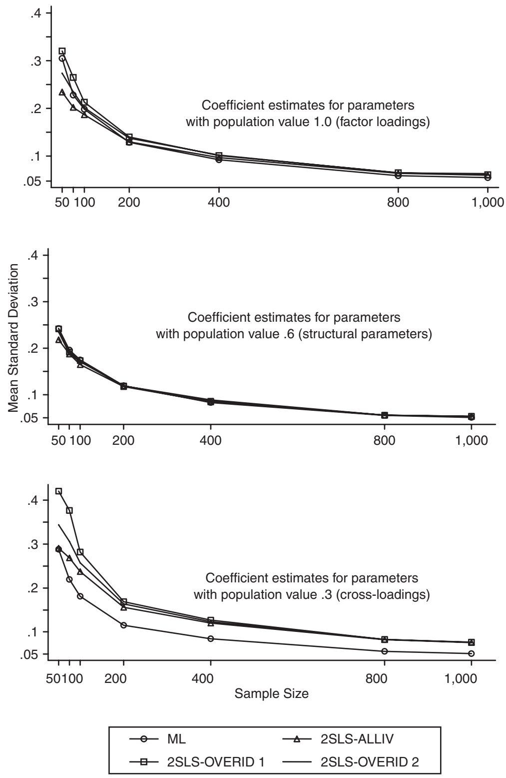


Note: ML = maximum likelihood; 2SLS = two-stage least squares.

of the estimates, as each instrument provides additional information. In general, we find little evidence of a great efficiency gain of the ML over the 2SLS estimators, at least under the conditions that we studied here.

In a supplemental analysis (results not shown), we also examined the accuracy of the standard errors by comparing them to the empirical standard deviations of the sampling distributions of each point estimate. We found that the standard errors from all estimators were close to their corresponding empirical standard deviations, with very little difference across

Figure 5
Mean Standard Deviations of Four Estimators by Sample Size for Parameter Estimates From Specification 3, in Which Two Factor Loadings Are Omitted



Note: ML = maximum likelihood; 2SLS = two-stage least squares.

estimators. There was a slight tendency for the mean of the standard errors to be too small for the ML in sample sizes less than 400. But this underestimation was generally less than 10 percent.

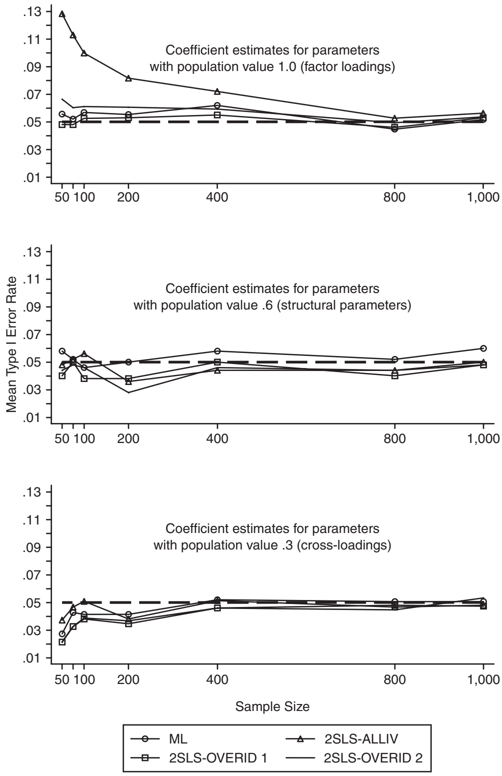
Type I Error Rate

We next examined the accuracy of significance testing of the estimators by testing the null hypothesis that the parameter is equal to the generating parameter (either .3, .6, or 1.0) using a Type I error of .05.[12] Ideally, we would expect the hypothesis to be rejected for 5 percent of the samples. The accuracy of the test depends on the unbiasedness of the estimator, the accuracy of the estimated asymptotic standard errors, and the accuracy of the asymptotic normality of the estimator. In this sense, the significance test provides a good check on whether the several components of the test combined result in accurate inferences.[13] We would expect our tests to be relatively accurate for all estimators in large samples and in correctly specified models (see Research Issue 5). Figure 6 supports this prediction for samples greater than 400. The least accurate test is the 2SLS-ALLIV estimator, which rejects too frequently for 400 cases or fewer when the parameter value is 1.0, whereas its Type I error is more accurate for the smaller coefficients of .6 and .3. For example, when the parameter value is 1.0, the Type I error rates for 2SLS-ALLIV are .128, .113, and .10 for sample sizes of 50, 75, and 100, respectively. At sample size 50 and a parameter value of .3, all estimators reject too infrequently, though this bias is less evident for all estimators at the parameter value of .6. Overall, for the 2SLS-OVERID1, 2SLS-OVERID 2, and the ML estimators, Type I errors are relatively accurate, with the possible exception of the smallest sample size with the smallest parameter value.

The situation changes substantially when the model is misspecified (Research Issue 6). For example, Figure 7 shows the rejection rates for the four estimators for tests done on parameter estimates that come from correct equations in the model that omits two cross-loadings. Most obvious in this figure is that the rejection rate for the ML-based tests is too high and grows as the sample increases in size for the parameter values of .6 and .3. Indeed, Type I error rates range from .216 ($N = 200$) to .814 ($N = 1{,}000$) for the parameter value of .60. This marked inflation is attributable to ML spreading specification error in one part of the system to other parameter estimates. The inflated Type I error rates for the ML in Figure 7 are thus due to the corresponding positive bias in the coefficient estimates. The 2SLS estimators exhibit too frequent rejection for the parameter value of .6, but the biases in the error rate are less than that for ML. The parameter value of .3 exhibits the greatest distinction between the Type I error rates for the ML and for the 2SLS estimators. The 2SLS estimators are far more accurate due to their robustness to specification errors. All of the estimators maintain accurate Type I error rates for the parameter value of 1.0 in samples of 200 or more.

In sum, in the correctly specified model and for $N \geq 400$, we find relatively accurate Type I error rates for all estimators. The 2SLS-ALLIV

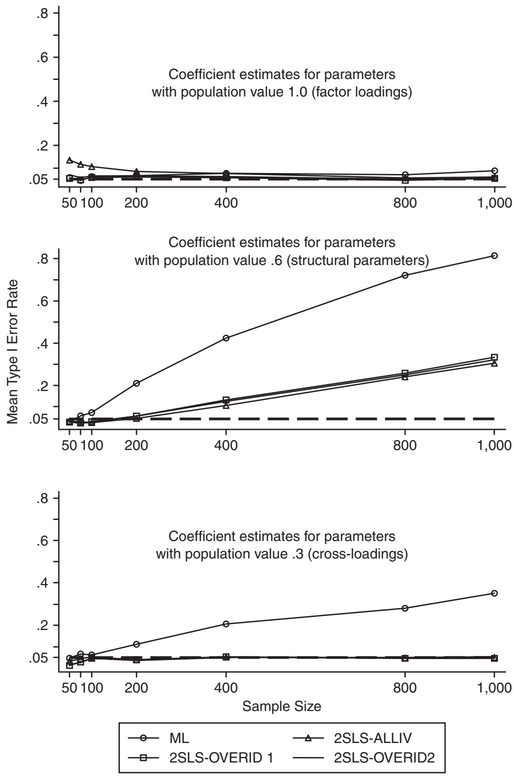
Figure 6
Type I Error Rate of Four Estimators by Sample Size for Parameter Estimates From Specification 1, the Correctly Specified Model



Note: ML = maximum likelihood; 2SLS = two-stage least squares.

rejected too frequently at smaller sample sizes for the parameter value of 1.0 but performed better for the other parameter values. The 2SLS-OVERID1, 2SLS-OVERID2, and ML estimators had fairly accurate Type I error rates with the exception of sample sizes 50 and 75 for the parameter value of .3. In the misspecified models, the ML exhibited the least accurate Type I errors for the parameter values of .6 and .3; the 2SLS estimators had less extreme bias for the .6 coefficient and negligible bias in Type I error for the .3 parameter value.

Figure 7
Type I Error Rate for Four Estimators by Sample Size for Parameter Estimates From Specification 3, in Which Two Factor Loadings Are Omitted



Note: ML = maximum likelihood; 2SLS = two-stage least squares.

Summary and Conclusions

Here, we compare the original research issues and expectations about the performance of the ML and 2SLS estimators to the simulation results:

1. *When the model is correctly specified, we expect the bias of all estimators to be negligible in the larger sample sizes, but we expect some degree of bias at the smaller sample sizes.* Our simulation

results were consistent with this expectation in that for samples of 400 or more, we found little bias in 2SLS-ALLIV, 2SLS-OVERID1, 2SLS-OVERID2, and ML. The greatest bias occurred at the smallest sample sizes. But even here, with the exception of the 2SLS-ALLIV estimator the biases were within ± 6 percent. Overall, the 2SLS-OVERID1 had the lowest mean bias.

2. *When the model is incorrectly specified, we expect the bias of the ML estimator to be greater than that of the 2SLS estimators and that this bias will still be present in larger sample sizes.* This prediction was also largely supported by the analysis of the population covariance matrix and by the simulation results. The analysis of the population covariance matrix showed that the ML estimator spread bias beyond the structurally misspecified equations, whereas the 2SLS estimators better isolated it to the misspecified equations. In addition, the magnitude of the bias was generally larger for the ML than for the 2SLS estimators. The simulation results for finite samples exhibited similar patterns. For the parameter coefficients of .6 and .3, the ML mean bias ranged from about 8 percent to 16 percent, and this bias was fairly stable across sample sizes. The 2SLS estimators had roughly 2 percent to 9 percent bias for the .6 coefficient. The 2SLS-OVERID1 and 2SLS-OVERID2 estimators had negligible bias for the .3 coefficient (less than 3 percent), while the 2SLS-ALLIV was only slightly more biased at the lowest sample size for the same coefficient. Interestingly, all but the 2SLS-ALLIV estimator at the smaller sample sizes had little bias for the 1.0 coefficient. Considering these bias results as a whole, the 2SLS-OVERID1 and 2SLS-OVERID2 estimators had the best performance.
3. *When the model is correctly specified, we expect the variance of the ML estimator to be smaller than that of the 2SLS estimators in the larger sample sizes. We have no prediction at the smaller sample sizes.* The efficiency advantage of the ML estimator did not clearly emerge. Rather, in the larger samples the standard deviations were very similar across estimators. The only separation clearly visible was for the standard deviations of the estimators for the .3 coefficient where the ML standard deviation was lower, but even here the difference was far from dramatic. We did not expect this result for several reasons. One is that the simulation conditions tended to favor the ML estimator. This is because the variables came from normal distributions as assumed for the ML estimator; in this first set of results, the model was correctly specified, and we did not use

any replication samples where the ML estimator did not converge or converged to an improper solution. These combine to favor the ML since the most extreme cases are removed from the sample and if included they would tend to increase the standard deviations. At the smaller sample sizes the standard deviations were fairly close, with the exception of the 2SLS-ALLIV estimator having the smallest at the lower sample sizes for the 1.0 coefficient. This would be an attractive property for the 2SLS-ALLIV, except for the greater bias that it has in small samples. Overall, these results do not strongly favor any of the estimators based on efficiency arguments.

4. *When the model is incorrectly specified, we have no prediction about the relative size of the variances of the estimators.* Although we made no prediction for these results, our simulation conclusions were similar to those described in the model with the correct specification, and we do not repeat them here.
5. *When the model is correctly specified, we predict that the Type I error rates will be accurate in the larger sample sizes.* In the samples with $N \geq 400$, we did find that the empirical Type I error rates were close to the .05 expected value, with the exception that the 2SLS-ALLIV, on average, rejected too frequently for the 1.0 coefficient value. At the smaller samples, the 2SLS-OVERID1, 2SLS-OVERID2, and ML generally had good accuracy, with the exception that the error rate was too low for the .3 coefficient at the smallest sample sizes. Overall, the simulation results for a correctly specified model support the use of 2SLS-OVERID1, 2SLS-OVERID2, and ML for accuracy of Type I error rates in testing coefficient estimates.
6. *When the model is incorrectly specified, we expect that the Type I error rates will be inaccurate for those coefficients that are not robust to the specification error.* The simulation results were consistent with the predictions of this hypothesis. When examining the bias of the estimators, we found that all of them were relatively insensitive to the misspecifications for the 1.0 coefficient. The results for the Type I error rate for the 1.0 coefficient are similar to those just described. However, in the section on bias we found that the ML was the most sensitive to misspecification for the .6 and .3 coefficient values, and this sensitivity to bias is manifested in the inaccurate Type I error rates for those coefficients when the ML is used. The 2SLS estimators had some bias for the .6 coefficient, and due to this, they also have inaccurate Type I errors, although they are not as biased as the ML's. Finally, the 2SLS estimator's Type I

errors are very accurate for the .3 coefficient, and this is largely due to their robustness to the specification errors. Overall, these results favor the 2SLS-OVERID1 and 2SLS-OVERID2 estimators for their relative accuracy of Type I error across the most conditions.

To our knowledge, this simulation is the first to compare the ML and 2SLS estimators in latent variable SEMs. These results have several implications. First, they illustrate that the tendency of 2SLS estimators to have greater robustness to structural misspecification than the ML estimator found in simultaneous equation models carries over to latent variable models.[14] Unless the researcher has tremendous confidence in the correctness of his or her model, it would be prudent to consider a 2SLS estimator as a complement or substitute for the ML estimator that might better isolate the impact of these structural misspecifications.

Second, the asymptotic efficiency advantage claimed for ML estimators was far less than expected. The efficiency advantage could be greater in other types of models, but for the correct and incorrect ones we considered and for the different coefficient values, the advantage was not that evident. This was even more surprising because the simulation conditions favored the ML in that variables came from normal distributions and we trimmed the ML estimator by eliminating sample replications that did not converge or that had improper solutions.

Third, we also found that the degree of overidentification for the 2SLS estimator matters. In general, not using all possible IVs worked better than including a large number of IVs with respect to minimizing bias in small samples, though using all available instruments lowered the variance of the 2SLS estimator. Specifically, the 2SLS-OVERID1 and 2SLS-OVERID2 outperformed the 2SLS-ALLIV estimator. This was mostly true in small samples since the 2SLS estimators had fewer differences in larger samples.

These results suggest that researchers should consider using 2SLS when they suspect that their models have omitted paths or other incorrect structures. Although none of the estimators prevent bias in the equation that is misspecified, the 2SLS estimators exhibited greater resistance to spreading the bias to other correctly specified equations. Furthermore, among the 2SLS estimators we examined, we would recommend the 2SLS-OVERID1 and 2SLS-OVERID2 estimators because of their lack of bias across a range of conditions. As we stated earlier, the 2SLS estimator is robust when the nature of the specification error does not change the IVs used (Bollen 2001). The 2SLS-OVERID1 and 2SLS-OVERID2 use fewer IVs and hence are less likely to include ineligible IVs under misspecification. At the same time, the empirical standard deviations of these estimators in moderately sized samples were competitive with the other estimators. The 2SLS-ALLIV exhibited greater bias at the smaller sample sizes, despite its greater efficiency with a small N.

When an equation is overidentified (i.e., more IVs than the minimum required), it is possible to test whether all IVs are uncorrelated with the equation disturbance, as is required for proper IVs. Two tests are based on the idea that if the overidentifying restrictions for a particular equation are correct, then the corresponding IVs should not explain any additional variance in the dependent variable when added to the right-hand side of the equation. Based on this, Anderson and Rubin (1949) developed a χ^2 test statistic and Basman (1960) developed an F test of the null hypothesis that the residual sum of squares for a regression estimated without the variables to be excluded is the same as the residual sum of squares for the regression estimated with the variables to be excluded. A third test statistic evaluates the appropriateness of the overidentifying restrictions by performing an OLS regression of the IVs on the residuals from the equation, forming NR^2, where N is the sample size and R^2 is the squared multiple correlation. This statistic follows a χ^2 distribution with degrees of freedom equal to the number of excess IVs and tests the null hypothesis of whether all IVs are uncorrelated with the disturbance term (Bollen 1996b; Davidson and MacKinnon 1993:236). These equation-by-equation tests contrast with the overall χ^2 test of overidentification routinely used with the ML estimator, and these former tests might be helpful in localizing the sources of specification error.[15]

We close with a cautionary note. As with all simulation results, our results might be dependent on the models, distributions, and other specific conditions considered. It is possible that different experimental conditions could alter our findings. Our experimental conditions treat a broader range of conditions than other simulations of simultaneous equations that compare these estimators, but we would encourage researchers to develop other models with which to explore these estimators.

Notes

1. See Bollen (2002) for a definition of latent variables.

2. Our simulation looks at variables that are incorrectly omitted from some equations in the model but appear in at least one equation. Other possible omitted variables are ones that are part of the true model but that do not appear anywhere in the system.

3. A Wald test is a straightforward method for this. Bollen (2001) provides the asymptotic covariance matrix for all coefficients in the model that would enable simultaneous tests of coefficients in a single equation or in multiple equations. Also note that an overidentification test of whether all instrumental variables (IVs) are uncorrelated with the equation disturbances is available for each overidentified equation (e.g., see Bassman 1960).

4. More formally, we can write the assumptions about the IVs for the two-stage least squares (2SLS) estimator as $\text{plim}\left(\frac{1}{N}\boldsymbol{V}_j'\boldsymbol{Z}_j\right) = \boldsymbol{\Sigma}_{\boldsymbol{VZ}j}$, $\text{plim}\left(\frac{1}{N}\boldsymbol{V}_j'\boldsymbol{V}_j\right) = \boldsymbol{\Sigma}_{\boldsymbol{VV}j}$, and $\text{plim}\left(\frac{1}{N}\boldsymbol{V}_j'\boldsymbol{u}_j\right) = \mathbf{0}$, where plim refers to the probability limit as N goes to infinity. The first assumption indicates that the covariance matrix of the IVs and the variables in $\boldsymbol{Z}_j$ must exist. Furthermore, this must be a nonzero association. Similarly, the covariance matrix of the IVs must exist and be nonsingular. A key requirement of the IVs in $\boldsymbol{V}_j$ is that they are uncorrelated with the disturbance of the equation. Finally, we assume that $E(\boldsymbol{u}_j) = 0$ and $E(\boldsymbol{u}_j\boldsymbol{u}_j\prime) = \sigma^2_{u_j}\mathrm{I}$. We can modify the homoscedasticity assumption (see Bollen 1996a), but we do not consider this further.

5. A minor exception is that the original Jöreskog–Hägglund–Madansky 2SLS estimator did not include intercepts, whereas Bollen's version of two-stage least squares (2SLS) does. However, this has no impact on the Hägglund (1983) and Brown (1990) simulations. Note also that Hägglund's FABIN3 estimator is the one that is equivalent to Bollen's (1996b) 2SLS under the conditions described in the text.

6. In addition, those studies that are done do not examine Bollen's (1996b) type of 2SLS estimator in the general case.

7. The combination of normality, omitting nonconverged solutions under maximum likelihood (ML), and including some correct specifications should privilege ML, making our comparison to the newer 2SLS a conservative comparison.

8. Although this is similar in structure to a simplex model with latent variables, it is not intended as one. If it were, we would have introduced correlated errors for indicators that were repeated over time. More generally, it represents a causal chain among the latent variables that leads the latent variable model to be overidentified, and each latent variable has relatively few indicators, but enough to overidentify the measurement model.

9. Here and throughout our analysis, we report the unstandardized coefficients.

10. To save space, we do not report each individual parameter but group them by their magnitude. In practice, the position of a parameter in a model can have an effect. Table 2 gives more specific information on the individual parameters and their biases under each structural misspecification.

11. The pattern of results is the same if the two additional misspecified models are considered (omitting either one or three cross-loadings from the correct model).

12. The null hypothesis for the significance test is not that the parameter is zero, but that it is equal to the true, generating parameter. This preserves the assumption that the null hypothesis is true, and we can then estimate the Type I probability.

13. It is possible, but unlikely, that inaccuracies in these components would counterbalance each other and lead to an accurate test. But even in this case, we still would have information on the accuracy of tests of significance when using the different estimators.

14. It might be thought that the misspecifications in the model using ML are less problematic in that their existence would be detected by the likelihood ratio χ^2 test routinely reported by structural equation modeling (SEM) software. However, there are at least two problems with this perspective. One is that statistical power of the χ^2 test is not always sufficient, particularly

in small samples, to detect structural misspecification. The mean *p* values of the χ^2 tests reported in Table 1 illustrate this possibility. Furthermore, even when the χ^2 test is significant, researchers often turn to alternative fit indices to assess model fit. Even in our misspecified models, it was common to find large fit indices (e.g., Tucker–Lewis, comparative fit index) that could lead a researcher to downplay a significant χ^2. Given current practice using ML, it is not evident that researchers will be aware of the misspecifications in their models.

15. The Lagrangian multiplier ("modification index") test (LM test) and the expected parameter change (EPC) statistics are alternative ways to localize specification errors when the full information estimators are used. It would be interesting to combine the equation-by-equation overidentification tests of 2SLS with the LM test and EPC of the full information estimators as diagnostics to locate structural misspecifications, though we are unaware of any attempts to do so. Two of the authors are investigating the finite sample behavior of the overidentification tests for equations that we mentioned in the text.

References

Anderson, T. W. and Herman Rubin. 1949. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *Annals of Mathematical Statistics* 20:46-63.

Basmann, R. L. 1957. "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation." *Econometrica* 25:77-83.

———. 1960. "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics." *Journal of the American Statistical Association* 55:650-9.

Bentler, P. M. 1995. *EQS Program Manual*. Los Angeles: Multivariate Software.

Bollen, Kenneth A. 1989. *Structural Equations With Latent Variables.* New York: John Wiley.

———. 1995. "Structural Equation Models That Are Nonlinear in Latent Variables: A Least-Squares Estimator." Pp. 225-51 in *Sociological Methodology*, edited by P. V. Marsden. Cambridge, MA: Blackwell.

———. 1996a. "A Limited-Information Estimator for LISREL Models With and Without Heteroscedastic Errors." Pp. 227-41 in *Advanced Structural Equation Modeling Techniques*, edited by G. Marcoulides and R. Schumacker. Mahwah, NJ: Lawrence Erlbaum.

———. 1996b. "An Alternative Two Stage Least Squares (2SLS) Estimator for Latent Variable Equations." *Psychometrika* 61:109-21.

———. 2001. "Two-Stage Least Squares and Latent Variable Models: Simultaneous Estimation and Robustness to Misspecifications." Pp. 119-38 in *Structural Equation Modeling: Present and Future*, edited by R. Cudeck, S. Du Toit, and D. Sorbom. Lincolnwood, IL: Scientific Software.

———. 2002. "Latent Variables in Psychology and the Social Sciences." *Annual Review of Psychology* 53:605-34.

Bollen, Kenneth A. and Daniel J. Bauer. 2004. "Automating the Selection of Model-Implied Instrumental Variables." *Sociological Methods & Research* 32:425-52.

Bollen, Kenneth A. and Jeremy C. Biesanz. 2002. "A Note on a Two-Stage Least Squares Estimator for Higher-Order Factor Analyses." *Sociological Methods & Research* 30:568-79.

Bollen, Kenneth A. and Pamela Paxton. 1998. "Interactions of Latent Variables in Structural Equation Models." *Structural Equation Modeling* 5:267-93.

Bollen, Kenneth A. and Robert Stine. 1990. "Direct and Indirect Effects: Classical and Bootstrap Estimates of Variability." *Sociological Methodology* 20:115-40.

Bollen, Kenneth A. and Robert A. Stine. 1993. "Bootstrapping Goodness-of-Fit Measures in Structural Equation Models." Pp. 111-35 in *Testing Structural Equation Models*, edited by Kenneth A. Bollen and J. Scott Long. Newbury Park, CA: Sage.

Boomsma, Anne. 1982. "The Robustness of LISREL Against Small Sample Sizes in Factor Analysis Models." Pp. 149-73 in *Systems Under Indirect Observation, Part I*, edited by Karl G. Joreskog and H. Wold. Amsterdam: North-Holland.

Boomsma, Anne and J. J. Hoogland. 2001. "The Robustness of LISREL Modeling Revisited." Pp. 139-68 in *Structural Equation Models: Present and Future. A Festschrift in Honor of Karl Jöreskog*, edited by R. Cudeck, S. du Toit, and D. Sörbom. Chicago: Scientific Software International.

Bowden, Roger J. and Darrell A. Turkington. 1984. *Instrumental Variables.* New York: Cambridge.

Brown, R. L. 1990. "The Robustness of 2SLS Estimation of a Non-Normally Distributed Confirmatory Factor Analysis Model." *Multivariate Behavioral Research* 25:455-66.

Browne, Michael W. 1984. "Asymptotically Distribution-Free Methods for the Analysis of Covariance Structures." *British Journal of Mathematical and Statistical Psychology* 37:62-83.

Buse, A. 1992. "The Bias of Instrumental Variable Estimators (in Notes and Comments)." *Econometrica* 60:173-80.

Cragg, John G. 1967. "On the Relative Small-Sample Properties of Several Structural-Equation Estimators." *Econometrica* 35:89-110.

———. 1968. "Some Effects of Incorrect Specification on the Small-Sample Properties of Several Simultaneous-Equation Estimators." *International Economic Review* 9:63-86.

Curran, Patrick J. 1994. "The Robustness of Confirmatory Factor Analysis to Model Misspecification and Violations of Normality." Ph.D. dissertation, Department of Psychology, Arizona State University, Tempe, AZ.

Davidson, Russell and James G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. New York: Oxford University.

Donatos, George S. and George C. Michailidis. 1996. "A Monte Carlo Study of Some Limited and Full Information Simultaneous Equation Estimators With Normal and Nonnormal Autocorrelated Disturbances." *Journal of Statistical Planning and Inference* 50:273-82.

Gerbing, David W. and James C. Anderson. 1985. "The Effects of Sampling Error and Model Characteristics on Parameter-Estimation for Maximum-Likelihood Confirmatory Factor-Analysis." *Multivariate Behavior Research* 20:255-71.

Hägglund, Gösta. 1982. "Factor Analysis by Instrumental Variables Methods." *Psychometrika* 47:209-22.

———. 1983. "Factor Analysis by Instrumental Methods: A Monte Carlo Study of Some Estimation Procedures." Ph.D. dissertation, Department of Statistics, University of Uppsala, Sweden.

Hatanaka, Michio. 1973. "On the Existence and Approximations Formulae for the Moments of K-Class Estimators." *Economic Studies Quarterly* 24:1-15.

Hausman, Jerry A. 1983. "Specification and Estimation of Simultaneous Equation Models." Pp. 392-448 in *Handbook of Econometrics*, vol. 1, edited by Z. Griliches and M. D. Intriligator. New York: North-Holland.

Johnston, John. 1972. *Econometric Methods.* New York: McGraw-Hill.

Jöreskog, Karl G. 1973. “A General Method for Estimating a Linear Structural Equation System.” Pp. 85-112 in *Structural Equation Models in the Social Sciences*, edited by A. S. Goldberger and O. D. Duncan. New York: Seminar Press.

———. 1983. “Factor Analysis as an Error-in-Variables Model. Pp. 185-96 in *Principles of Modern Psychological Measurement*, edited by H. Wainer and S. Messick. Hillsdale, NJ: Lawrence Erlbaum.

Jöreskog, Karl G. and Dag Sörbom. 1993. *LISREL 8.* Mooresville, IN: Scientific Software.

Kaplan, David. 1988. “The Impact of Specification Error on the Estimation, Testing, and Improvement of Structural Equation Models.” *Multivariate Behavioral Research* 23:69-86.

Kinal, Terrence W. 1980. “The Existence of Moments of K-Class Estimator.” *Econometrica* 48:241-9.

Lance, Charles E., John M. Cornwell, and Stanley A. Mulaik. 1988. “Limited Information Parameter Estimates for Latent or Mixed Manifest and Latent Variable Models.” *Multivariate Behavioral Research* 23:171-87.

Madansky, Albert. 1964. “Instrumental Variables in Factor Analysis.” *Psychometrika* 29:105-13.

Mariano, R. S. 1972. “The Existence of Moments of the Ordinary Least Squares and Two-Stage Least Squares Estimators.” *Econometrica* 40:643-52.

Mariano, Roberto S. 1982. “Analytical Small-Sample Distribution Theory in Econometrics: The Simultaneous-Equations Case.” *International Economic Review* 23:503-33.

Micceri, Theodore. 1989. “The Unicorn, the Normal Curve, and Other Improbable Creatures.” *Psychological Bulletin* 105:156-66.

Mikhail, W. M. 1975. “A Comparative Monte Carlo Study of the Properties of Econometric Estimators.” *Journal of the American Statistical Association* 70:94-104.

Nagar, A. L. 1959. “The Bias and Moment Matrix of the General K-Class Estimators of the Parameters in Simultaneous Equations.” *Econometrica* 27:575-95.

Oczkowski, Edward. 2002. “Discriminating Between Measurement Scales Using Non-Nested Tests and 2SLS: Monte Carlo Evidence.” *Structural Equation Modeling* 9:103-25.

Paxton, Pamela M., Patrick Curran, Kenneth A. Bollen, J. Kirby, and Feinian Chen. 2001. “Monte Carlo Simulations in Structural Equation Models.” *Structural Equation Modeling* 8:287-312.

Phillips, Peter C. B. 1983. “Exact Small Sample Theory in the Simultaneous Equations Model.” Pp. 449-516 in *Handbook of Econometrics*, vol. 1, edited by Z. Griliches and M. D. Intriligator. New York: North-Holland.

Richardson, David H. 1970. “The Asymptotic Unbiasedness of Two-Stage Least Squares.” *Econometrica* 38:772.

Sargan, J. D. (1970). “The Finite Sample Distribution of FIML Estimators.” Presented to the World Congress of the Econometric Society, Cambridge, UK.

Satorra, Albert. 1990. “Robustness Issues in Structural Equation Modeling: A Review of Recent Developments.” *Quality & Quantity* 24:367-86.

Satorra, Albert and Peter M. Bentler. 1988. “Scaling Corrections for Chi-Square Statistics in Covariance Structure Analysis.” *1988 Proceedings of the Business and Economic Statistics Section of the American Statistical Association*: 308-13.

Satorra, Albert and Willem E. Saris. 1985. “Power of the Likelihood Ratio Test in Covariance Structure Analysis.” *Psychometrika* 50:83-90.

Sawa, Takamitsu. 1969. “The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators.” *Journal of the American Statistical Association* 64:923-37.

Theil, H. 1961. *Economic Forecasts and Policy*. Amsterdam: North-Holland.
Theil, Henry. 1953. "Estimation and Simultaneous Correlation in Complete Equation Systems." Central Planning Bureau, The Hague, Netherlands. Unpublished manuscript.
———. 1958. *Economic Forecasts and Policy.* Amsterdam: North-Holland.

Kenneth A. Bollen is director of the Odum Institute for Research in Social Science and the H. R. Immerwahr Distinguished Professor of Sociology at The University of North Carolina at Chapel Hill. He is the Year 2000 recipient of the Lazarsfeld Award for Methodological Contributions in Sociology. The Institute for Scientific Information named him among the World's Most Cited Authors in the Social Sciences. He is coauthor of *Latent Curve Models: A Structural Equations Approach* (with P. Curran, 2006, Wiley) and author of *Structural Equation Models With Latent Variables* (1989, Wiley) and of more than 100 articles. Bollen's primary areas of statistical research are structural equation models and latent curve models.

James B. Kirby is a senior social scientist at the Agency for Healthcare Research and Quality. His research focuses on family and community influences on health, health behaviors, and access to health care in the United States and on quantitative methods. He has recently published articles in the *Journal of Health and Social Behavior*, *Medical Care*, and *Journal of Family Issues.*

Patrick J. Curran is a professor in the L. L. Thurstone Quantitative Laboratory in the Department of Psychology at The University of North Carolina at Chapel Hill. His primary area of research focuses on the development of new quantitative methods for longitudinal data analysis and the application of these methods to study drug and alcohol use in children and adolescents. He is coauthor of *Latent Curve Models: A Structural Equations Approach* (with K. Bollen, 2006, Wiley) and has recently published articles in *Psychological Methods*, *Multivariate Behavioral Research*, *Structural Equation Modeling*, *Journal of Abnormal Psychology*, and *Development and Psychopathology.*

Pamela R. Paxton is an associate professor of sociology at The Ohio State University. She has intersecting interests in political sociology, prosocial behavior, and methodology. She is coauthor of *Women, Politics, and Power: A Global Perspective* (with Melanie Hughes, 2007, Sage). Some of her previous research appears in the *American Sociological Review*, the *American Journal of Sociology*, and *Social Forces.* Her current research focuses on women's political participation over time.

Feinian Chen is an assistant professor of sociology at North Carolina State University. Her research interests include family, demography, aging, and quantitative methods. She has recently published articles in *Social Forces*, *Population and Development Review*, and *Journal of Comparative Family Studies.* She is currently investigating the issue of grandparents caring for grandchildren in China.