PSYCHOMETRIKA—VOL. 61, NO. 1, 109–121
MARCH 1996

AN ALTERNATIVE TWO STAGE LEAST SQUARES (2SLS) ESTIMATOR FOR LATENT VARIABLE EQUATIONS

KENNETH A. BOLLEN

SOCIOLOGY DEPARTMENT
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

The Maximum-likelihood estimator dominates the estimation of general structural equation models. Noniterative, equation-by-equation estimators for factor analysis have received some attention, but little has been done on such estimators for latent variable equations. I propose an alternative 2SLS estimator of the parameters in LISREL type models and contrast it with the existing ones. The new 2SLS estimator allows observed and latent variables to originate from nonnormal distributions, is consistent, has a known asymptotic covariance matrix, and is estimable with standard statistical software. Diagnostics for evaluating instrumental variables are described. An empirical example illustrates the estimator.

Key words: structural equation models, covariance structure models, LISREL, Two Stage Least Squares, 2SLS, latent variables, factor analysis, noniterative estimators, instrumental variables.

Full-information estimators such as maximum likelihood (ML), generalized least square (GLS), and weighted least squares (WLS) dominate the applications of structural equation modeling (SEM). Despite this, researchers continue to show interest in limited-information, noniterative estimators. Several reasons support this interest. First, in general the ML, GLS, and WLS estimators are more computationally-intensive procedures than are the noniterative estimators. Second, SEM software packages like LISREL (Jöreskog & Sörbom, 1993) use limited-information, noniterative estimators to provide starting values that can reduce the number of iterations and that can lessen the chances of nonconvergence for the iterative estimators.

Aside from these computational aspects, consistent noniterative estimators have value in their own right. ML, GLS, and WLS incorporate information from throughout the system in developing estimates. This can improve estimator efficiency. A drawback is that specification error in one part of the system can spread bias to other parts of the model. Limited-information estimators tend to isolate the bias to fewer parts of the system.

Nearly all the work on noniterative estimators in the SEM tradition has concentrated on the measurement model or factor analysis (Hägglund, 1982; Jöreskog, 1983; Madansky, 1964). An exception is Jöreskog and Sörbom (1993) who use a two-stage least squares (2SLS)/instrumental variable (IV) procedure to estimate the coefficients of the latent variable ("structural") model as well. Their estimates require that the measurement model be estimated first and they do not give the distributional properties, standard errors, or significance tests for the coefficients of the latent variable model.

This paper has several purposes. First, I propose an alternative 2SLS estimator of

I gratefully acknowledge support for this research from the Sociology Program of the National Science Foundation (SES-9121564) and the Center for Advanced Study in the Behavioral Sciences, Stanford, California. This paper was presented at the Interdisciplinary Consortium for Statistical Applications at Indiana University at Bloomington (March 2, 1994) and at the RMD Conference on Causal Modeling at Purdue University, West Lafayette, Indiana (March 3–5, 1994).

Requests for reprints should be sent to Kenneth A. Bollen, CB 3210 Hamilton, Department of Sociology, University of North Carolina, Chapel Hill, NC 27599-3210 or bollen@unc.edu (internet).

0033-3123/96/0300-94194$00.75/0

the coefficients in the equations of the latent variable model in "LISREL" models. I also briefly present the estimator for the coefficients of the measurement model and a method to estimate the other parameters in the full LISREL model. Second, I provide the standard errors and significance tests for the coefficient estimators. These are asymptotically valid even when observed variables come from nonnormal distributions. Third, I present an overidentification test that researchers can apply to each overidentified equation. Finally, I illustrate the methods with an empirical example.

The next section briefly reviews the literature on limited-information, noniterative estimators. A section on the notation, model, and the development of the estimator along with its properties follows. The section after presents methods to estimate the rest of the coefficients and variances and covariances in the LISREL model. Next is a section that provides diagnostics to evaluate the instrumental variables (IVs) of the 2SLS procedure. An empirical example then illustrates the methodologies. The last section has the conclusions.

Literature Review

Two groups of literature are relevant. One is the econometric literature on instrumental variable (IV) estimators, including Two-Stage Least Squares (2SLS). The other is work on IV/2SLS estimators in the general structural equation model, sometimes called the LISREL model. IV estimators have a long history in econometrics (see, e.g., Bowden & Turkington, 1984, pp. 16–20). An early application of IV treated measurement error in the explanatory variable of a regression equation (Reirsøl, 1941). The key idea is that random measurement error in an explanatory variable creates a correlation between that explanatory variable and the disturbance term of the regression equation. An IV is another variable that has no direct impact on the dependent variable, but has a correlation with the explanatory variable and no correlation with the disturbance term. In econometrics these IVs are nearly always exogenous or lagged endogenous observed variables which are called predetermined variables. The IV estimator uses the instrument variables to find a consistent estimator of the coefficient for the original explanatory variable. The 2SLS estimator is an IV estimator that forms the optimal combination of IVs when more than one IV is available.

Econometric texts that discuss the IV estimator for the errors in variable problem typically focus on bivariate or multiple regression where one or more explanatory variables are measured with error and a single indicator is available for each explanatory variable (e.g., Johnston, 1984, pp. 428–35). Extensions are usually limited to a single equation model where an explanatory variable has multiple indicators or to specific examples of two or three equation models (e.g., Bowden & Turkington, 1984, pp. 3–7, 58–62; Aigner, Hsiao, Kapteyn, & Wansbeek, 1984). I am unaware of any attempts in the econometric literature to apply IV/2SLS to a general SEM such as the LISREL model that I describe in the next section. These SEM models allow multiple indicators of all latent variables, factor structures where a single observed variable may be influenced by multiple latent variables, reciprocal relationships between latent variables, and a variety of other relationships not treated as part of a single model in the econometric literature. In addition, in econometric applications at least some of the observed variables are treated as exogenous variables. And the econometric model commonly is written in a reduced-form where each endogenous variable is a function of the predetermined (exogenous and lagged endogenous) observed variables, coefficients, and disturbances. Many of the models that arise in SEM cannot be written in such a reduced form since they have no exogenous or lagged endogenous observed variables. Rather a reduced form that has *latent* exogenous variables is more typical.

Despite these differences, the econometric literature on IV/2SLS has greatly influenced the work of those dealing with such models. Madansky (1964) had perhaps the first contribution in the factor analysis literature on the use of IV. He suggested an IV estimator for the factor loadings in factor analysis models. Hägglund (1982) and Jöreskog (1983) built on this work and proposed IV and two-stage least squares (2SLS) estimators for factor analysis models with uncorrelated errors of measurement. Studies such as those of Jennrich (1987) and Cudeck (1991) have looked at computational aspects as well as selection of the scaling variables for factors using the IV method. Bollen (1989, p. 415) gave a simple way to extend these techniques to models with correlated errors. Satorra and Bentler (1991) provide a test statistic for factor analysis models based on the IV estimator. The Monte Carlo evidence to date shows that the IV and 2SLS estimators perform well in factor analysis models (Brown, 1990; Hägglund, 1983; Lukashov, 1994).

The work on limited-information, noniterative estimators for the latent variable model (or "structural model") in general SEM is rare. One potential source of confusion in reviewing these efforts is that each is a 2SLS estimator, even though they are not the same. The problem is that applying the principles that lie behind the 2SLS estimator to equations with latent variables does not lead to a single method of estimation. So it is not possible to say that only one of these estimators is the "true" 2SLS estimator. Which 2SLS estimator I am referring to should be clear from the context of my discussion.

Jöreskog and Sörbom's (1993) 2SLS estimator for equations from the latent variable model has two components: First, they use the Madansky-Hägglund-Jöreskog IV and 2SLS estimators for the factor loadings of the measurement model along with formulas from Hägglund (1982) to estimate the covariance matrices of the latent variables. Second, they use this estimated covariance matrix of the latent variables and apply a version of 2SLS to estimate the coefficient estimates for the latent variable model. In essence they treat the covariance matrix of the latent variables as if it were a covariance matrix of observed variables and then apply the usual 2SLS formula written in terms of the covariance matrix of observed variables. Since their major interest in this estimator was to estimate starting values for the iterative estimators in LISREL, Jöreskog and Sörbom (1993) do not discuss the properties of these 2SLS/IV estimators for the latent variable model. To avoid confusing this 2SLS estimator with the others, I refer to their estimator as the JS-2SLS/IV estimator.

Lance, Cornwell, & Mulaik (1988) propose a noniterative estimator in the same spirit as the JS-2SLS/IV estimator. The difference is that they estimate the factor analysis part of the SEM with ML or the other full-information estimators. From this they estimate the correlation matrix of the latent variables and any perfectly measured observed variables. Then like Jöreskog and Sörbom, they treat this matrix as if it were a covariance matrix of observed variables and apply 2SLS/IV (or ordinary least squares for recursive models) formulas for covariance matrices.[1] I refer to this as the LCM-2SLS/IV estimator.

The JS-2SLS/IV and LCM-2SLS/IV estimators share the advantage of providing starting values for iterative procedures. They also give an equation-by-equation estimator that might help in identifying and isolating specification error in the latent variable model. However, they both require estimation of the measurement model (or factor analysis model) prior to estimating the latent variable model. The LCM-2SLS/IV uses the system-wide estimators such as ML, so that the reduction in computational resources is less than the JS-2SLS/IV estimator. As the next section demonstrates, it is

[1] See Bentler (1982) for another noniterative estimator for factor analysis.

possible to develop a 2SLS estimator that does not require estimation of the factor analysis model. Like the JS-2SLS/IV estimator, the distributional assumptions for and the properties of the LCM-2SLS/IV estimator are not provided. The next section describes the underlying assumptions and properties for the 2SLS estimator that I recommend. Unless I indicate otherwise, the 2SLS estimator that I refer to henceforth is the one developed in this paper.

Model and Estimator

I begin with the widely used "LISREL" notation (Jöreskog & Sörbom, 1993) for structural equation models (SEMs). The latent variable model is:

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{1}$$

where $\boldsymbol{\eta}$ is an $m \times 1$ vector of latent endogenous random variables, $\mathbf{B}$ is a $m \times m$ matrix of coefficients that give the impact of the $\boldsymbol{\eta}$'s on each other, $\boldsymbol{\xi}$ is an $n \times 1$ vector of latent exogenous variables, $\boldsymbol{\Gamma}$ is the $m \times n$ coefficient matrix giving $\boldsymbol{\xi}$'s impact on $\boldsymbol{\eta}$, $\boldsymbol{\alpha}$ is an $m \times 1$ vector of intercept terms, and $\boldsymbol{\zeta}$ is an $m \times 1$ vector of random disturbances with the $E(\boldsymbol{\zeta}) = \mathbf{0}$ and Cov $(\boldsymbol{\xi}, \boldsymbol{\zeta}') = \mathbf{0}$.

The goal is to estimate $\boldsymbol{\alpha}$, $\mathbf{B}$, and $\boldsymbol{\Gamma}$ from the latent variable model, but to do so it is important to know the measurement model as well. Two equations summarize this aspect of the SEM:

$$\mathbf{x} = \boldsymbol{\tau}_{\mathbf{x}} + \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta} \tag{2}$$

$$\mathbf{y} = \boldsymbol{\tau}_{\mathbf{y}} + \boldsymbol{\Lambda}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{3}$$

where $\mathbf{x}$ is a $q \times 1$ vector of observed indicators of $\boldsymbol{\xi}$, $\boldsymbol{\Lambda}_x$ is a $q \times n$ matrix of "factor loadings" (regression coefficients) giving the impact of $\boldsymbol{\xi}$ on $\mathbf{x}$, $\boldsymbol{\tau}_{\mathbf{x}}$ is $q \times 1$ vector of intercept terms, and $\boldsymbol{\delta}$ is a $q \times 1$ vector of measurement errors with $E(\boldsymbol{\delta}) = \mathbf{0}$ and Cov $(\boldsymbol{\xi}, \boldsymbol{\delta}') = \mathbf{0}$. Similarly in (3) $\mathbf{y}$ is a $p \times 1$ vector of indicators of $\boldsymbol{\eta}$, $\boldsymbol{\Lambda}_{\mathbf{y}}$ is the $p \times m$ matrix of factor loadings, $\boldsymbol{\tau}_{\mathbf{y}}$ is a $p \times 1$ vector of intercept terms, and $\boldsymbol{\varepsilon}$ is a $p \times 1$ vector of errors with $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and Cov $(\boldsymbol{\eta}, \boldsymbol{\varepsilon}') = \mathbf{0}$. Additional assumptions are that $\boldsymbol{\varepsilon}$, $\boldsymbol{\delta}$, and $\boldsymbol{\zeta}$ are mutually uncorrelated.

As mentioned in the literature review, Madansky (1964), Hägglund (1982), and Jöreskog (1983) propose instrumental variable (IV)/two-stage least squares (2SLS) estimators for factor analysis models such as those in equation (2) [or (3)]. However, they ignore the intercept terms ($\boldsymbol{\tau}_{\mathbf{x}}$ or $\boldsymbol{\tau}_{\mathbf{y}}$) and work in deviation scores. My focus is on the latent variable model in (1). To identify the latent variable model, a researcher must scale it. A straightforward scaling option is to choose one indicator per latent variable for which its factor loading is set to one and its intercept set to zero (see Bollen, 1989, pp. 350–352). I assume that the scaling variable is only influenced by a single latent variable and an error term.

After scaling all latent variables, sort the $\mathbf{y}$ and $\mathbf{x}$ vectors so that the indicators that scale the latent variables come first. Then create partitioned vectors for $\mathbf{y}$ and $\mathbf{x}$:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix},$$

where $\mathbf{y}_1$ is the $m \times 1$ vector of y's that scale $\boldsymbol{\eta}$, $\mathbf{y}_2$ consists of the $(p - m) \times 1$ vector of remaining y variables, $\mathbf{x}_1$ is the $n \times 1$ vector of x's that scale $\boldsymbol{\xi}$, and $\mathbf{x}_2$ is the $(q - n) \times 1$ vector of remaining x variables.

This means that

$$\mathbf{y}_1 = \boldsymbol{\eta} + \boldsymbol{\varepsilon}_1, \quad \text{or} \tag{4}$$

$$\boldsymbol{\eta} = \mathbf{y}_1 - \boldsymbol{\varepsilon}_1, \tag{5}$$

and

$$\mathbf{x}_1 = \boldsymbol{\xi} + \boldsymbol{\delta}_1, \quad \text{or} \tag{6}$$

$$\boldsymbol{\xi} = \mathbf{x}_1 - \boldsymbol{\delta}_1, \tag{7}$$

where $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\delta}_1$ contain the errors that correspond to $\mathbf{y}_1$ and $\mathbf{x}_1$. Substituting (5) and (7) into (1) leads to:

$$\mathbf{y}_1 = \boldsymbol{\alpha} + \mathbf{B}\mathbf{y}_1 + \boldsymbol{\Gamma}\mathbf{x}_1 + \mathbf{u}, \tag{8}$$

where $\mathbf{u} = \boldsymbol{\varepsilon}_1 - \mathbf{B}\boldsymbol{\varepsilon}_1 - \boldsymbol{\Gamma}\boldsymbol{\delta}_1 + \boldsymbol{\zeta}$. Note that these manipulations recast the latent variable model into a simultaneous equation model where all variables are observed except for the composite disturbance term. Equation (8) appears to match the classical econometric simultaneous equation model (e.g., Bollen, 1989, pp. 80–81), but an important difference is that we cannot assume that $\mathbf{u}$ and $\mathbf{x}_1$ are uncorrelated. Recall that $\mathbf{u}$ contains $\boldsymbol{\delta}_1$ and that $\boldsymbol{\delta}_1$ correlates with those x's that are measured with error. That is, in general $\mathbf{x}_1$ is not a vector of exogenous (predetermined) variables, though some x's in $\mathbf{x}_1$ that are error-free could be exogenous (predetermined).

At this point it will help to consider a single equation from (8). I represent the i-th equation from y_1 as:

$$y_i = \alpha_i + \mathbf{B}_i\mathbf{y}_1 + \boldsymbol{\Gamma}_i\mathbf{x}_1 + u_i, \tag{9}$$

where y_i is the i-th y from $\mathbf{y}_1$, α_i is the corresponding intercept, $\mathbf{B}_i$ is the i-th row from $\mathbf{B}$, $\boldsymbol{\Gamma}_i$ is the i-th row from $\boldsymbol{\Gamma}$, and u_i is the i-th element from $\mathbf{u}$.

Define $\mathbf{A}_i$ to be a column vector that contains α_i and all of the nonzero elements of $\mathbf{B}_i$ and $\boldsymbol{\Gamma}_i$ strung together in a column. Let N equal the number of cases and $\mathbf{Z}_i$ be an N row matrix that contains 1's in the first column and the N rows of elements from $\mathbf{y}_1$ and $\mathbf{x}_1$ that have nonzero coefficients associated with them in the remaining columns. The $N \times 1$ vector $\mathbf{y}_i$ contains the N values of y_i in the sample and $\mathbf{u}_i$ is an $N \times 1$ vector of the values of u_i. Then we can rewrite (9) as:

$$\mathbf{y}_i = \mathbf{Z}_i\mathbf{A}_i + \mathbf{u}_i. \tag{10}$$

Ordinary least squares (OLS) is inappropriate for (10) because in most cases at least some of the variables in $\mathbf{Z}_i$ will be correlated with $\mathbf{u}_i$. The only exception would be if $\mathbf{Z}_i$ consisted of perfectly measured x's $(x_j = \xi_j)$ or of perfectly measured x's and y's $(y_k = \eta_k)$ where the y's also were uncorrelated with ζ_i. Other than these situations, OLS is not a consistent estimator of (10). However, the two-stage least squares (2SLS) estimator provides an alternative consistent estimator of $\mathbf{A}_i$.

The 2SLS estimator requires instrumental variables (IVs) for $\mathbf{Z}_i$. The IVs must be: (a) correlated with $\mathbf{Z}_i$, (b) uncorrelated with $\mathbf{u}_i$, and (c) sufficient in number so that there are at least as many IVs as the number of explanatory variables on the right-hand side of the equation. Generally, the pool of potential IVs comes from those y's and x's *not* included in $\mathbf{Z}_i$ (excluding, of course, y_i). The exceptions are any variables in $\mathbf{Z}_i$ that are uncorrelated with u_i, since such variables can serve as IVs. Exogenous (predetermined) x's would be an example of IVs that might appear on the right-hand side of (9).

The sample correlations between the potential IVs and $\mathbf{Z}_i$ provide a check of condition (a). Condition (b) is more difficult to establish but the full model structure is essential in evaluating it. Recall that u_i equals $(\varepsilon_i - \mathbf{B}_i\boldsymbol{\varepsilon}_1 - \boldsymbol{\Gamma}_i\boldsymbol{\delta}_1 + \zeta_i)$. IVs must be uncorrelated with each component in the composite. The $\mathbf{B}_i\boldsymbol{\varepsilon}_1$ term rules out using y's that scale the latent variables and that have a nonzero impact on y_i. In addition because of $\mathbf{B}_i\boldsymbol{\varepsilon}_1$ we cannot use y's as IVs that have correlated errors of measurement with those y's in $\mathbf{y}_1$ that appear in the y_i equation.

The $\boldsymbol{\Gamma}_i\boldsymbol{\delta}_1$ term rules out x's as IVs that scale the latent ξ's and that have a nonzero direct impact on y_i. In addition any x's with correlated errors of measurement with such x's that appear in the y_i equation cannot be IVs. Finally, the ζ_i in the $\mathbf{u}_i$ eliminates any y's as IV that correlate with ζ_i. This means, for example, that y's that are indicators of η's that appear after η_i in a "causal chain" would be ineligible to be IVs.

Let v_j be a possible IV that correlates with $\mathbf{Z}_i$. Another way to describe the check of condition (b) is to form the Cov (v_j, u_i). Substitute in the equation for u_i and substitute in the reduced form equation for v_j, if v_j is an endogenous variable. Determine the Cov (v_j, u_i). If it is zero, then v_j satisfies condition (b) for an IV.

Condition (c) ensures that there are a sufficient number of IVs to allow estimation of the model. Without further restrictions imposed, this counting rule is a necessary condition of identifying parameters in (9) and in using the 2SLS estimator.

I will illustrate the selection of IVs in the examples. For now assume that we collect all eligible IVs for $\mathbf{Z}_i$ and a column of 1's in an N row matrix $\mathbf{V}_i$. Then the first stage of 2SLS is to regress $\mathbf{Z}_i$ on $\mathbf{V}_i$ where (11) provides the coefficient estimator:

$$(\mathbf{V}_i'\mathbf{V}_i)^{-1}\mathbf{V}_i'\mathbf{Z}_i. \tag{11}$$

Form $\hat{\mathbf{Z}}_i$ as:

$$\hat{\mathbf{Z}}_i = \mathbf{V}_i(\mathbf{V}_i'\mathbf{V}_i)^{-1}\mathbf{V}_i'\mathbf{Z}_i. \tag{12}$$

The second stage is the OLS regression of $\mathbf{y}_i$ on $\hat{\mathbf{Z}}_i$ so that

$$\hat{\mathbf{A}} = (\hat{\mathbf{Z}}_i'\hat{\mathbf{Z}}_i)^{-1}\hat{\mathbf{Z}}_i'\mathbf{y}_i. \tag{13}$$

Equation (13) clarifies the importance and the necessity of having enough IVs as described in condition (c) for IVs. If there are fewer IVs than variables with nonzero coefficients, then $\hat{\mathbf{Z}}_i$ will be less than full rank and $(\hat{\mathbf{Z}}_i'\hat{\mathbf{Z}}_i)^{-1}$ is nonexistent.

Reviewing the assumptions underlying this 2SLS estimator will expedite the discussion of its properties. Assume that:

$$\text{plim}\left(\frac{1}{N}\mathbf{V}_i'\mathbf{Z}_i\right) = \boldsymbol{\Sigma}_{\mathbf{VZ}_i}, \tag{14}$$

$$\text{plim}\left(\frac{1}{N}\mathbf{V}_i'\mathbf{V}_i\right) = \boldsymbol{\Sigma}_{\mathbf{V}_i\mathbf{V}_i}, \quad \text{and} \tag{15}$$

$$\text{plim}\left(\frac{1}{N}\mathbf{V}_i'\mathbf{u}_i\right) = \mathbf{0}. \tag{16}$$

Here plim refers to the probability limits as N goes to infinity of the term in parentheses. The right hand side matrices of (14) to (16) are finite, $\boldsymbol{\Sigma}_{\mathbf{V}_i\mathbf{V}_i}$ is nonsingular, and $\boldsymbol{\Sigma}_{\mathbf{V}_i\mathbf{Z}_i}$ is nonzero. Assume that $E[\mathbf{u}_i\mathbf{u}_i'] = \sigma_{u_i}^2\mathbf{I}$ and $E(\mathbf{u}_i) = \mathbf{0}$.

Note that $\hat{\mathbf{A}}_i$ equals:

$$\begin{aligned}\hat{\mathbf{A}}_i &= (\hat{\mathbf{Z}}_i'\hat{\mathbf{Z}}_i)^{-1}\hat{\mathbf{Z}}_i'\mathbf{y}_i \\ &= (\hat{\mathbf{Z}}_i'\hat{\mathbf{Z}}_i)^{-1}\hat{\mathbf{Z}}_i'(\hat{\mathbf{Z}}_i\mathbf{A}_i + \mathbf{u}_i) \\ &= \mathbf{A}_i + (\hat{\mathbf{Z}}_i'\hat{\mathbf{Z}}_i)^{-1}(\hat{\mathbf{Z}}_i'\mathbf{u}_i). \qquad (17)\end{aligned}$$

The plim ($\hat{\mathbf{A}}_i$) is

$$\text{plim } (\hat{\mathbf{A}}_i) = \mathbf{A}_i + \text{plim}\left[\left(\frac{1}{N}\hat{\mathbf{Z}}_i'\hat{\mathbf{Z}}_i\right)^{-1}\frac{1}{N}\hat{\mathbf{Z}}_i'\mathbf{u}_i\right] = \mathbf{A}_i \qquad (18)$$

Thus as is well known, the 2SLS estimator is consistent and the consistency holds for this application.

In practice analysts want to estimate the covariance matrix of $\hat{\mathbf{A}}_i$ and its distribution so that statistical inference and hypothesis testing are plausible. Few finite sample results are known for the 2SLS estimator for general econometric models, so there is not much to draw on for this new application to latent variable models. However, some asymptotic properties apply. To derive these consider (17) again. From (17)

$$\hat{\mathbf{A}}_i = \mathbf{A}_i + (\hat{\mathbf{Z}}_i'\hat{\mathbf{Z}}_i)^{-1}\hat{\mathbf{Z}}_i'\mathbf{u}_i,$$

so

$$\begin{aligned}\hat{\mathbf{A}}_i - \mathbf{A}_i &= (\hat{\mathbf{Z}}_i'\hat{\mathbf{Z}}_i)^{-1}\hat{\mathbf{Z}}_i'\mathbf{u}_i \\ &= \left(\frac{1}{N}\hat{\mathbf{Z}}_i'\hat{\mathbf{Z}}_i\right)^{-1}\frac{1}{N}\hat{\mathbf{Z}}_i'\mathbf{u}_i. \qquad (19)\end{aligned}$$

Assume that:

$$\frac{1}{\sqrt{N}}\hat{\mathbf{Z}}_i'\mathbf{u}_i \sim AN(\mathbf{0},\ \sigma^2_{u_i}\ \boldsymbol{\Sigma}_{\hat{\mathbf{Z}}_i\hat{\mathbf{Z}}_i}) \qquad (20)$$

$$\text{plim}\left(\frac{1}{N}\hat{\mathbf{Z}}_i'\hat{\mathbf{Z}}_i\right)^{-1} = \boldsymbol{\Sigma}^{-1}_{\hat{\mathbf{Z}}_i\hat{\mathbf{Z}}_i} \qquad (21)$$

where $AN(\ \ )$ refers to an asymptotic normal distribution. Appeal to a variant of the central limit theorem justifies (20) under fairly general conditions. The previous assumptions (14) and (15) justify (21). These assumptions lead to:

$$\sqrt{N}\ (\hat{\mathbf{A}}_i - \mathbf{A}_i) \xrightarrow{\text{D}} N(\mathbf{0},\ \sigma^2_{u_i}\boldsymbol{\Sigma}^{-1}_{\hat{\mathbf{Z}}_i\hat{\mathbf{Z}}_i}), \qquad (22)$$

so that the asymptotic distribution of $\hat{\mathbf{A}}_i$ is normal with a covariance matrix of $\sigma^2_{u_i}\boldsymbol{\Sigma}^{-1}_{\hat{\mathbf{Z}}_i\hat{\mathbf{Z}}_i}$.

The estimate of the asymptotic covariance matrix is:

$$\text{acov } (\hat{\mathbf{A}}_i) = \hat{\sigma}^2_{u_i}(\hat{\mathbf{Z}}_i'\hat{\mathbf{Z}}_i)^{-1}, \qquad (23)$$

where

$$\hat{\sigma}^2_{u_i} = (\mathbf{y}_i - \mathbf{Z}_i\hat{\mathbf{A}}_i)'(\mathbf{y}_i - \mathbf{Z}_i\hat{\mathbf{A}}_i)/N \qquad (24)$$

and acov signifies the sample estimate of the asymptotic covariance. These estimates allow significance tests for the coefficients.

It bears emphasis that I have not assumed that the observed variables (x's, y's, or Z's) are normally distributed. So the 2SLS estimator is applicable even for some observed variables that come from nonnormal distributions. In particular, note that the consistency of the estimator shown in (18) does not depend on a normality assumption. The main distributional assumption is that the *limiting* distribution of $1/\sqrt{N}\, \hat{\mathbf{Z}}_i'\mathbf{u}_i$ is normal and this assumption helps to establish the asymptotic covariance matrix of $\hat{\mathbf{A}}_i$. In general, this seems a reasonable assumption, but these are large sample results.

Applying the above procedure to all equations in the latent variable model, gives coefficient estimates and asymptotic standard errors for $\boldsymbol{\alpha}$, $\mathbf{B}$, and $\boldsymbol{\Gamma}$ from (1).

Estimating Other Parameters

Estimates of the intercepts and factor loadings of the measurement model follow an analogous procedure to that of the latent variable model. More specifically, consider the measurement model for $\mathbf{x}$ in (2). Substitute equation (7) for $\boldsymbol{\xi}$ into (2) which leads to:

$$\mathbf{x} = \boldsymbol{\tau}_{\mathbf{x}} + \boldsymbol{\Lambda}_{\mathbf{x}}\mathbf{x}_1 - \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\delta}_1 + \boldsymbol{\delta} \qquad (25)$$

Since $\mathbf{x}_1$, the scaling variables, come first in $\mathbf{x}$, the first n rows of $\boldsymbol{\tau}_x$ are zero and the first n rows and n columns of $\boldsymbol{\Lambda}_x$ form an identity matrix.

Choosing one of the x's from the $\mathbf{x}_2$ vector of nonscaling x's leads to:

$$x_i = \tau_{x_i} + \boldsymbol{\Lambda}_{xi}\mathbf{x}_1 + d_i \qquad (26)$$

where d_i equals $(-\boldsymbol{\Lambda}_{xi}\boldsymbol{\delta}_1 + \delta_i)$, τ_{x_i} is the intercept for the x_i equation, $\boldsymbol{\Lambda}_{xi}$ is the i-th row of $\boldsymbol{\Lambda}_x$, δ_i is the error of measurement for x_i. Define $\mathbf{C}_i$ to be a column vector that contains τ_{x_i} and all of the nonzero factor loadings in $\boldsymbol{\Lambda}_{xi}$ put together in a column. N is the number of observations and let $\mathbf{W}_i$ be an N row matrix that contains 1's in the first column and the N rows of elements from $\mathbf{x}_1$ that have nonzero factor loadings associated with them in the remaining columns. The $\mathbf{x}_i$ vector is $N \times 1$ and contains the N values of x_i in the sample and $\mathbf{d}_i$ is an $N \times 1$ vector of the values of d_i.

These definitions are akin to those for the latent variable model. For the measurement model for $\mathbf{x}$, they lead to

$$\mathbf{x}_i = \mathbf{W}_i\mathbf{C}_i + \mathbf{d}_i. \qquad (27)$$

Equation (27) is analogous to (10). In fact, the 2SLS estimator applies to this equation as it did to (10). The major difference is that a researcher must select IVs that are correlated with $\mathbf{W}_i$ and uncorrelated with $\mathbf{d}_i$. The procedure is so similar to that described for the latent variable model that I will not repeat it here. In addition, the same steps apply to estimating the equations for the measurement model for $\mathbf{y}$ (see (3)).

This 2SLS estimator for the *measurement model* is closely related to Hägglund's (1982) FABIN3 estimator. The major differences are that Hägglund's FABIN3 does not allow correlated errors of measurement and FABIN3 does not estimate intercepts. The correlated errors affect the selection of IVs so that some variables that would be eligible IVs with uncorrelated errors are no longer so with correlated errors (see Bollen, 1989, p. 415).

The results up to this point provide the means to estimate magnitudes and asymptotic standard errors for intercepts and coefficients for all the equations in the latent variable and in the measurement model. In most applications the magnitude and statistical significance of the factor loadings and regressions coefficients are of primary concern. However, if interest also lies in the variances and covariances of the latent exogenous variables ($\boldsymbol{\xi}$), of the equation disturbances ($\boldsymbol{\zeta}$), or of the errors of measurement (ε, δ), then to estimate them requires further steps. A simple option is to input the 2SLS estimates of the intercepts, coefficients, and factor loadings as fixed values and to estimate the remaining variances and covariances of latent exogenous variables, disturbances, and errors with the maximum likelihood, generalized least squares, or other fitting functions in a SEM program. This would provide consistent estimators of all the parameters of the full structural equation model. This latter property follows from the consistency of the intercept, coefficients, and factor loadings from the 2SLS estimator and the consistency of the estimates from the ML, GLS, WLS, and other common fitting functions.[2]

Evaluating Instrumental Variables (IVs)

Several diagnostics help to evaluate the appropriateness of IVs for a given equation. The simplest condition is that the 2SLS estimator requires at least as many IVs as there are endogenous variables on the right-hand side of the equation to be estimated. This was discussed in the previous section.

A simple diagnostic for the "quality" of the IVs is the R^2 resulting from the regression of each right-hand side endogenous variable on the IVs for that equation. A low R^2 (e.g., <0.1) in most cases will lead to a poor 2SLS estimator. This could be due to a poor scaling indicator that is weakly related to the latent variable or it could be due to inadequate IVs. One check for the former is to try another scaling indicator and to determine if a similar low R^2 results. If not, then a poor scaling indicator is the likely problem and the researcher can estimate the model with the new scaling indicator. If the R^2 remains low regardless of the selection of the scaling indicator, I generally would not recommend the use of the 2SLS estimator. Indeed, full-information fitting functions such as ML also may perform badly under these conditions.

Tests of the overidentifying restrictions in an equation are available when the number of IVs exceeds the number of right-hand side endogenous variables. Such tests are helpful in locating variables that are not proper IVs. This is important since incorrectly treating a variable as an IV can lead to an inconsistent coefficient estimator in the equations where it is improperly used. These test statistics were developed for classical econometric models but they suit the latent variable models of this paper. Davidson and MacKinnon (1993, p. 236) describe a simple test that I illustrate using (10) from the latent variable model. Consider the 2SLS residuals,

$$\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{Z}_i\hat{\mathbf{A}}_i. \tag{28}$$

Perform an OLS regression of these residuals on $\mathbf{V}_i$, the IVs for the equation. Form NR^2 where N is the sample size and R^2 is the squared multiple correlation from this equation. This test statistic has an asymptotic chi-square distribution with degrees of freedom equal to the difference in the column dimension of $\mathbf{V}_i$ minus the column dimension of $\mathbf{Z}_i$. A significant test statistic suggests that at least one of the IVs is correlated with the disturbance term, contrary to the required condition for IVs. Other econometric overidentification tests could be adapted to these models as well (see, e.g., Basmann, 1960).

These diagnostic tools apply to each overidentified equation in the full SEM. This contrasts with the full-information estimators like ML or GLS that provide a chi-square test of an overidentified model as a whole. These equation by equation checks might be more useful in localizing the source of specification error than is an overall test, though

[2] See Browne (1984) for consistency property of the ML, GLS, WLS, and other fitting functions that are standard in estimating SEMs.

diagnostics such as the Lagrangian Multiplier tests ("modification index") play a similar role in the more traditional SEM approach.

I will illustrate the diagnostic tests in the next section with the empirical example.

Industrialization and Political Democracy Example

The example looks at the impact of industrialization on political democracy for 75 developing countries between 1960 and 1965. Figure 1 is the path diagram. The model and data are from Bollen (1989). The latent variable model represents political democracy (η_1) in 1960 as dependent on industrialization (ξ_1) in 1960. Political democracy (η_2) in 1965 is influenced by industrialization (ξ_1) in 1960 and the prior level of political democracy (η_1) in 1960. The equations for the latent variable model are:

$$\eta_1 = \alpha_1 + \gamma_{11}\xi_1 + \zeta_1 \tag{29}$$

$$\eta_2 = \alpha_2 + \beta_{21}\eta_1 + \gamma_{21}\xi_1 + \zeta_2. \tag{30}$$

Substituting $(y_1 - \varepsilon_1)$ in for η_1, $(y_5 - \varepsilon_5)$ in for η_2, and $(x_1 - \delta_1)$ in for ξ_1 and rearranging terms leads to:

$$y_1 = \alpha_1 + \gamma_{11}x_1 - \gamma_{11}\delta_1 + \varepsilon_1 + \zeta_1 \tag{31}$$

$$y_5 = \alpha_2 + \beta_{21}y_1 + \gamma_{21}x_1 - \beta_{21}\varepsilon_1 - \gamma_{21}\delta_1 + \varepsilon_5 + \zeta_2. \tag{32}$$

The variables on the right-hand side in each equation are correlated with their respective composite disturbance terms. Following the rules for selection of IVs in the prior section, the IV for x_1 in equation (31) are x_2 and x_3. The IVs for y_1 and x_1 in (32) are y_2, y_3, y_4, x_2 and x_3.

Table 1 reports the 2SLS estimates and standard errors for the latent variable model and contrasts them with the corresponding ML and GLS estimates from LISREL 8 (Jöreskog & Sörbom, 1993). The estimates and standard errors are similar. Interestingly, the values from the two system-wide estimators—ML and GLS—are not consistently closer to each other than either is to the 2SLS estimates. So the differences do not seem to be totally due to system wide estimators versus an equation by equation one. Rather a plausible explanation for the differences are the sampling fluctuations that accompany a small sample size and a moderately sized model.

The diagnostics for the IVs in each equation for the 2SLS estimator did not indicate any problems. The R^2 for x_1 regressed on the IVs (x_2 and x_3) in the first equation was 0.81, a high value. The R^2's for y_1 and x_1 regressed on the IVs (y_2, y_3, y_4, x_2, and x_3) in the second equation were 0.61 and 0.82, respectively. The overidentification tests of the IVs also were quite favorable. In the first equation the chi-square test statistic was 0.50 with 1 degree of freedom ($p = .48$) and the second equation had a test statistic of 0.80 with 3 df ($p = .85$). Thus the evidence supports the choice of IVs.[3]

As a further check of the overidentification test, I misspecified equation (30) so that industrialization (ξ_1) was omitted. Under this incorrect specification, x_1 is added to the list of IVs for the revised equation that has only the earlier value of political democracy as an explanatory variable. The new overidentification test statistic for this equation is 10.93 with 5 df ($p = .05$) so that the test statistic is statistically significant indicating a problem with one or more of the IVs. It is interesting to note that the overall chi-square test statistic for this misspecified model is 43.5 with 36 df ($p = .18$) with the ML estimator. Reliance only on this overall chi-square test would not reveal the specification problem.

[3] The Basmann (1960) overidentification test results were virtually identical.

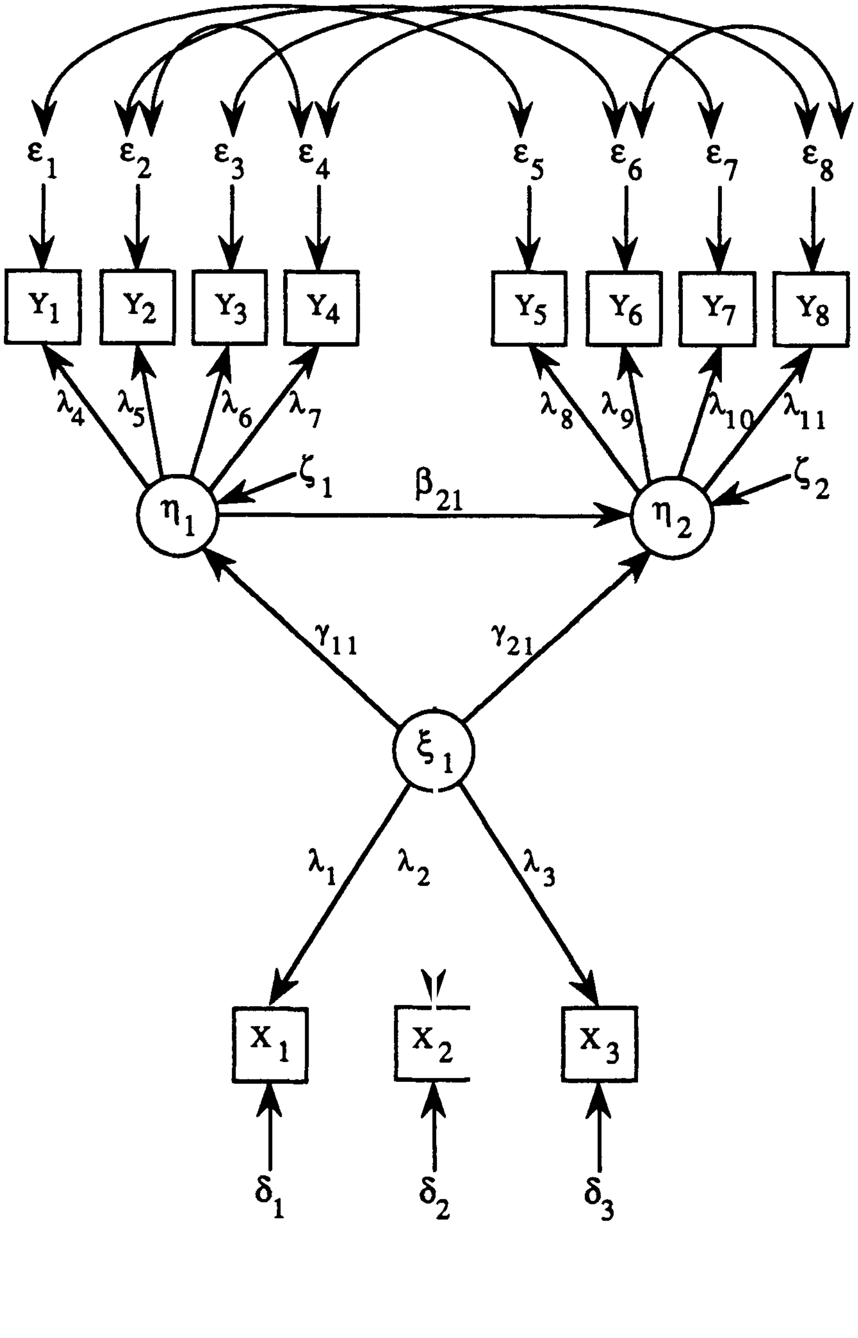
ε1
ε2
ε3
ε4
ε5
ε6
ε7
ε8
Y1
Y2
Y3
Y4
Y5
Y6
Y7
Y8
λ4
λ5
λ6
λ7
λ8
λ9
λ10
λ11
ζ1
ζ2
η1
β21
η2
γ11
γ21
ξ1
λ1
λ2
λ3
X1
X2
X3
δ1
δ2
δ3

FIGURE 1.

TABLE 1.

Table 1. Maximum Likelihood (ML), Generalized Least Squares (GLS)and Two-Stage Least Squares (2SLS) Estimates and (standard errors) for Industrialization (ξ_1) and Political Democracy (η_1,η_2) Panel Data (N=75)

	Dependent Variable					
	1960 Political democracy η_1			1965 Political democracy η_2		
Explanatory Variable	ML	GLS	2SLS	ML	GLS	2SLS
η_1	–	–	–	0.87 (0.08)	0.81 (0.10)	0.72 (0.10)
ξ_1	1.48 (0.40)	1.76 (0.49)	1.26 (0.43)	0.57 (0.22)	0.67 (0.28)	1.12 (0.32)
Intercept	-2.03 (2.05)	−3.41 (2.48)	-0.91 (2.20)	−2.33 (1.13)	−2.66 (1.36)	-4.50 (1.45)

Conclusions

System-wide ("Full-Information") estimators dominate structural equation modeling. However, starting with the factor analysis or measurement model component of the general SEM, researchers have begun to be interested in noniterative estimators. There is a small but growing literature on 2SLS/IV estimators of factor analysis models. The literature on applying such estimators to the latent variable model is even sparser with Jöreskog and Sörbom (1993) and Lance, et al. (1988) the only ones I have located. Though they both label their procedures as 2SLS estimators they differ from each other as well as from the 2SLS estimator that I presented. One important difference is that their procedures require estimation of the measurement model prior to estimating the latent variable model and my approach does not. In addition, I described the distributional assumptions underlying the proposed estimator and its properties. From these researchers can estimate standard errors and perform significance tests. Another positive feature is that the estimator does not require that the observed variables come from a multinormal distribution. Given the recent discouraging performance of the WLS estimator for SEM (Muthén & Kaplan, 1992), the 2SLS estimator bears investigation as an estimator for observed variables that come from nonnormal distributions.

A drawback is that like the other SEM estimators (e.g., ML) the properties established are asymptotically valid. Its finite sample properties cannot be known without much more experience. The simulation work on 2SLS estimators in factor analysis that I cited and the literature on 2SLS in simultaneous equation models from econometrics seems relevant here. For instance, Kennedy (1985, p. 134) summarizes the evidence on 2SLS in simultaneous equation models as follows: "Monte Carlo studies have shown it [2SLS] to have small sample properties superior on most criteria to all other estimators.

They have also shown it [2SLS] to be quite robust." Hägglund's (1983, p. 33) Monte Carlo experiment with exploratory factor analysis concludes that 2SLS works better than ML methods for small samples and that ML does not seem to outperform 2SLS in large samples. Lukashov's (1994) simulation experiment for the 2SLS estimator in confirmatory factor analysis also found good performance. These are encouraging results for the 2SLS estimator for latent variables.

The 2SLS estimator does not require specialized software to implement it. Researchers can use any software programs that have 2SLS procedures to estimate the latent variable models described in the paper. Indeed, strictly speaking all that would be required is an OLS regression package as long as the proper adjustments were made to the estimates of the standard errors.

References

Aigner, Dennis J., Cheng Hsiao, Arie Kapteyn, & Tom Warsbeek. (1984). Latent variable models in econometrics. In Zvi Griliches & Michael D. Intriligator (Eds.), *Handbook of Econometrics* (pp. 1321–1393). Amsterdam: North-Holland.

Basmann, R. L. (1960). On finite sample distributions of generalized classical linear identifiability test statistics. *Econometrica, 45*, 939–952.

Bentler, Peter M. (1982). Confirmatory factor analysis via noniterative estimation: A fast, inexpensive method. *Journal of Marketing Research, 19*, 417–424.

Bollen, Kenneth A. (1989). *Structural equations with latent variables*. NY: Wiley.

Bowden, Roger J., & Turkington, Darrell A. (1984). *Instrumental variables*. Cambridge: Cambridge University Press.

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, M. W. (1984). Asymptotic distribution free methods in analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62–83.

Cudeck, Robert. (1991). Noniterative factor analysis estimators, with algorithms for subset and instrumental variable selection. *Journal of Educational Statistics, 16*, 35–52.

Davidson, Russell, & MacKinnon, James G. (1993). *Estimation and inference in econometrics*. New York: Oxford University.

Hägglund, G. (1982). Factor analysis by instrumental variables. *Psychometrika, 47*, 209–2.

Hägglund, G. (1983). *Factor analysis by instrumental methods: A Monte Carlo study of some estimation procedures* (Report No. 80-2). Uppsala, Sweden: University of Uppsala, Department of Statistics.

Jennrich, R. I. (1987). Tableau algorithms for factor analysis by instrumental variable method. *Psychometrika, 52*, 469–476.

Johnston, J. (1984). *Econometric methods*. NY: McGraw-Hill.

Jöreskog, K. G. (1983). Factor analysis as an error-in-variables model. In H. Wainer & S. Messick (Eds.) *Principles of modern psychological measurement* (pp. 185–196). Hillsdale, NJ: Lawrence Erlbaum Associates.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8*. Mooresville, IN; Scientific Software.

Kennedy, P. (1985). *A Guide to econometrics*. Cambridge, MA: MIT Press.

Lance, C. E., Cornwell, J. M., & Mulaik, S. A. (1988). Limited information parameter estimates for latent or mixed manifest and latent variable models. *Multivariate Behavioral Research, 23*, 155–167.

Lukashov, Andrey. (1994). A Monte Carlo study of the IV estimator in factor analysis. Unpublished Master's thesis. Chapel Hill, NC: University of North Carolina, Department of Sociology.

Madansky, A. (1964). Instrumental variables in factor analysis. *Psychometrika, 29*, 105–113.

Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology, 45*, 19–30.

Reiersøl, O. (1941). Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica, 9*, 1–24.

Satorra, Albert, & Bentler, Peter M. (1991). Goodness-of-fit test under IV estimation: Asymptotic robustness of a NT test statistic. In Ramón Gutiérrez & Mariano J. Valderrana (Eds.), *Applied stochastic models and data analysis* (pp. 555–566). Singapore: World Scientific.

Manuscript received 3/17/94
Final version received 1/20/95