



Published in final edited form as:

Sociol Methodol. 2009 July 2; 39(1): 327–355.

Using Instrumental Variable (IV) Tests to Evaluate Model Specification in Latent Variable Structural Equation Models*

James B. Kirby and

Agency for Healthcare Research and Quality

Kenneth A. Bollen

University of North Carolina, Chapel Hill

Abstract

Structural Equation Modeling with latent variables (SEM) is a powerful tool for social and behavioral scientists, combining many of the strengths of psychometrics and econometrics into a single framework. The most common estimator for SEM is the full-information maximum likelihood estimator (ML), but there is continuing interest in limited information estimators because of their distributional robustness and their greater resistance to structural specification errors. However, the literature discussing model fit for limited information estimators for latent variable models is sparse compared to that for full information estimators. We address this shortcoming by providing several specification tests based on the 2SLS estimator for latent variable structural equation models developed by Bollen (1996). We explain how these tests can be used to not only *identify* a misspecified model, but to help diagnose the source of misspecification within a model. We present and discuss results from a Monte Carlo experiment designed to evaluate the finite sample properties of these tests. Our findings suggest that the 2SLS tests successfully identify most misspecified models, even those with modest misspecification, and that they provide researchers with information that can help diagnose the source of misspecification.

Introduction

Structural Equation Modeling with latent variables (SEM) is a powerful tool for social scientists, allowing researchers to simultaneously estimate relationships between latent variables and observed indicators, and structural relationships between latent variables. The SEM approach thus combines much of the analytic strengths of the psychometric tradition, with its emphasis on measurement, with those of the econometric tradition, with its emphasis on modeling multiequation relations between observed variables. This powerful combination has made SEM an increasingly popular methodological approach in a variety of disciplines.

By far the most popular estimator for SEM is the maximum likelihood estimator (ML). This is a “full-information” estimator in that it arrives at all parameter estimates simultaneously by using information from the entire system. If the model under investigation is correctly specified and the observed variables do not have excessive kurtosis, the ML estimator is consistent, asymptotically unbiased, and asymptotically efficient (Browne 1984; Jöreskog

*The views expressed in this paper are those of the authors, and no official endorsement by the Agency for Healthcare Research and Quality, or the Department of Health and Human Services is intended or should be inferred. The authors would like to thank the editor and reviewers for their valuable comments. Bollen gratefully acknowledges support from NSF SES 0617276, NIDA 1-R01-DA13148-01 and DA013148-05A2.

and Sörbom 1996).¹ An added advantage of the ML estimator is that it provides a variety of statistics that help analysts evaluate how well the model under investigation “fits” the available data. While these are highly desirable properties, the ML estimator and other full information estimators do have drawbacks. One major drawback is that when any part of a model is misspecified, as is almost always the case, bias can spread to parts of the model that are not misspecified. For this reason, research on alternative, limited information estimators continues.

One of the most recent limited information estimators is Bollen's (1996) two stage least squares estimator for latent variable SEM (2SLS). This estimator is consistent and has the same asymptotic properties as the ML estimator, but it is efficient only among limited information estimators, not full information estimators. A recent simulation study suggests that the 2SLS estimator performed somewhat better than the ML estimator under modest misspecification and that the efficiency advantage of the ML estimator was modest at best, even in perfectly specified models (Bollen, Kirby, Curran, Paxton, and Chen 2007). A drawback of Bollen's 2SLS estimator for latent variable models is that, unlike the ML estimator, little has been written on how to assess model fit. Bollen (1996, Pp. 117-118) briefly discusses possible overidentification tests for equations, but does not provide a complete discussion on alternative tests and their relative performance.² In this paper, we describe how to make use of several well known econometric tests of the appropriateness of instrumental variables (IVs) so as to enable tests of model fit for each overidentified equation in a latent variable model. We explain how to use these tests not only to *identify* a misspecified model, but to help diagnose the source of misspecification within a model. This is something that goodness of fit tests from the ML and other full information estimators do not provide, although Lagrangian Multiplier tests (“modification index”) suggest parameter restrictions in a model that might be lifted. Finally, we conduct a large Monte Carlo experiment to evaluate the finite sample properties of these IVs tests. The results apply not only to Bollen's 2SLS estimator for SEMs, but also to more traditional 2SLS applications found in Sociology, Economics, and other disciplines where overidentification tests of IVs are applicable.

Technical Background

The SEM Framework

Using a form of the LISREL notation (Jöreskog 1973) modified to include intercepts, we can represent a structural equation model with the following equations:

$$\begin{aligned}\eta &= \alpha_{\eta} + \mathbf{B}\eta + \mathbf{\Gamma}\xi + \zeta \\ y &= \alpha_y + \Lambda_y\eta + \varepsilon \\ x &= \alpha_x + \Lambda_x\xi + \delta\end{aligned}\tag{1}$$

The first equation in (1) represents the latent variable model, in which η is a vector of endogenous latent variables, ξ is a vector of exogenous latent variables, and ζ is a vector of disturbances; \mathbf{B} is a matrix of coefficients measuring relationships among endogenous latent variables and $\mathbf{\Gamma}$ is matrix of coefficients relating the exogenous latent variables to the endogenous latent variables. The second two equations in (1) represent the measurement

¹The asymptotic efficiency that Browne (1984, page 68) proves is among the class of Generalized Least Squares estimators. He also shows that the usual ML estimator from SEMs can be written as a GLS estimator.

²Bollen and Maydeu-Oliveres (2007) describe a Polychoric Instrumental Variable estimator for ordinal observed variables and an extension that enables overall fit tests for these models. Their method could be applied to continuous variables as well though we do not consider that extension here.

model, in which \mathbf{y} and \mathbf{x} are vectors of observed variables affected by $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, respectively, with measurement error terms contained in the vectors $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$; $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ are matrices containing the coefficients relating the latent variables to the observed indicator variables (i.e. “factor loadings”). The standard assumptions of SEM are that all the disturbance terms have an expected value of zero, that they are uncorrelated with $\boldsymbol{\xi}$ and to each other, and that $(\mathbf{I} - \mathbf{B})$ is nonsingular (Bollen 1989).

The Maximum Likelihood Estimator

Most full information estimators iteratively search for parameter estimates that minimize the difference between the sample covariance matrix of observed variables and the covariance matrix implied by the model structure with respect to a distance formula or “fitting function”. The fitting function used by the ML estimator is:

$$F_{ML} = \ln \left| \boldsymbol{\Sigma}(\boldsymbol{\theta}) \right| + \text{tr} \left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \mathbf{S} \right] - \ln |\mathbf{S}| - p + [\bar{\mathbf{z}} - \boldsymbol{\mu}(\boldsymbol{\theta})]' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) [\bar{\mathbf{z}} - \boldsymbol{\mu}(\boldsymbol{\theta})]$$

where $\boldsymbol{\theta}$ is a vector containing the parameters of the model to be estimated, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the model-implied covariance matrix, \mathbf{S} is the covariance matrix observed in the sample, $\boldsymbol{\mu}(\boldsymbol{\theta})$ is the vector of model implied means, $\bar{\mathbf{z}}$ is the vector of sample means of the observed variables, and p is the number of observed variables.

Under correct model specification and no excess kurtosis in the observed variables, the ML estimator is consistent, asymptotically unbiased, asymptotically efficient, and asymptotically normal (Browne 1984; Jöreskog 1973). An added advantage is that a test statistic to evaluate model fit is formed by multiplying the value of F_{ML} at its minimum by $N-1$, where N is the sample size. Under the same set of assumptions made for the SEM, this statistic asymptotically follows a central chi square distribution. The null hypothesis for this test is that $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, the model-implied covariance matrix, equals $\boldsymbol{\Sigma}$, the population covariance matrix. A significant test statistic, therefore, indicates that a model *does not* fit the data well. Though there are many alternative measures of model fit using the ML estimator, most are based on the minimum value of the fitting function.

The 2SLS Estimator

The 2SLS estimator is best thought of as a family of estimators, not all of which are the same. Originally, a 2SLS estimator was developed in econometrics by Theil (1953) and Basmann (1957) for simultaneous equation models without latent variables. Madansky (1964), Häggglund (1983), and Jöreskog (1983) propose 2SLS estimators that provide the starting values in the Jöreskog and Sörbom LISREL software (Jöreskog and Sörbom 1996). But these estimators assume no correlated measurement errors and, in the case of Jöreskog's and Sörbom's work, it involves estimating the covariance matrix of the factors and applying another estimator for the latent variable model. These 2SLS estimators differ 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, Häggglund, and Madansky.

Bollen's 2SLS estimator for latent variable models differs from these others in that it permits correlated errors, does not require estimation of the covariance matrix of the factors, and it provides asymptotic standard errors for all coefficients (Bollen 1996; Bollen 2001). Our analysis focuses on this version of the 2SLS estimator. Unlike the ML estimator, the 2SLS estimator is not a “full-information” estimator; model parameters are estimated for each equation separately. To apply the 2SLS estimator, each latent variable must be scaled by

choosing one of its indicators and setting its coefficient to one and intercept to zero. If the scaling indicators for all of the endogenous and exogenous latent variables are placed in vectors \mathbf{y}_1 and \mathbf{x}_1 , respectively, we represent the scaling of the latent variables with the following expressions:

$$\begin{aligned} y_1 &= \boldsymbol{\eta} + \boldsymbol{\varepsilon}_1 \\ x_1 &= \boldsymbol{\xi} + \boldsymbol{\delta}_1 \end{aligned}$$

These equations can be rewritten as:

$$\begin{aligned} \boldsymbol{\eta} &= y_1 - \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\xi} &= x_1 - \boldsymbol{\delta}_1 \end{aligned} \quad (2)$$

If the expressions for $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ in (2) are substituted into (1), the entire SEM can be rewritten as:

$$\begin{aligned} y_1 &= \boldsymbol{\alpha}_\eta + \mathbf{B}y_1 + \boldsymbol{\Gamma}x_1 + \boldsymbol{\varepsilon}_1 - \mathbf{B}\boldsymbol{\varepsilon}_1 - \boldsymbol{\Gamma}\boldsymbol{\delta}_1 + \boldsymbol{\zeta} \\ y_2 &= \boldsymbol{\alpha}_{y_2} + \boldsymbol{\Lambda}_{y_2}y_1 - \boldsymbol{\Lambda}_{y_2}\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 \\ x_2 &= \boldsymbol{\alpha}_x + \boldsymbol{\Lambda}_{x_2}x_1 - \boldsymbol{\Lambda}_{x_2}\boldsymbol{\delta}_1 + \boldsymbol{\delta}_2 \end{aligned} \quad (3)$$

where \mathbf{y}_2 and \mathbf{x}_2 are vectors containing the non-scaling observed variables. Notice that this transformation has removed all latent variables from the system, while retaining all of the model parameters to be estimated. The system represented in (3) could be estimated using OLS except that the composite disturbance terms are generally not uncorrelated with the observed variables on the right hand side (RHS) of the equations. These composite error terms are $\boldsymbol{\varepsilon}_1 - \mathbf{B}\boldsymbol{\varepsilon}_1 - \boldsymbol{\Gamma}\boldsymbol{\delta}_1 + \boldsymbol{\zeta}$ in the structural model, and $\boldsymbol{\Lambda}_{y_2}\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2$ and $\boldsymbol{\Lambda}_{x_2}\boldsymbol{\delta}_1 + \boldsymbol{\delta}_2$ in the measurement model. The transformed equations represented in (3) can, however, be estimated with a 2SLS estimator. For a more detailed discussion of the 2SLS estimator for SEM, please see (Bollen 1996; Bollen 2001)

For identification, the 2SLS estimator requires one or more instrumental variables (IVs) for each endogenous observed variable on the right hand side of each equation. Instrumental variables are observed variables that are predictive of the endogenous RHS variables but that are uncorrelated with the disturbance term in the equation. We assume that the covariance matrix of the IVs is nonsingular, that the covariance matrix of the IVs with the endogenous RHS variables is nonzero and without “weak” IVs, that the covariance of the composite disturbance with the IVs is zero, and that the composite disturbance term is homoscedastic and nonautocorrelated (Bollen, 1996, page 114). Though these are the same requirements as the 2SLS estimator used in econometric contexts, there is an important conceptual difference. In econometric applications, IVs are generally found outside of the model being estimated; analysts usually search for variables that meet the criterion for being appropriate IVs. The resulting IVs need not have substantive or theoretical bearing on the model being tested (though in more recent economic applications they often do). In contrast, the IVs used for Bollen’s 2SLS estimator in the SEM context are always variables contained within the model, and they are *determined unambiguously* by the structure of the model itself. The fact that the IVs are determined by the model structure is central to our paper. By testing the appropriateness of IVs in the SEM context, we are not simply testing whether our selection

of IVs meets the statistical criterion necessary for consistent and asymptotically unbiased parameter estimation, we are testing the specification of the model under investigation.

Specification tests for the 2SLS Estimator

The top line of (3) appears similar to a simultaneous equation model from econometrics in that, other than the disturbances, the model is composed of observed variables. The latent variables have been eliminated through substitution and the models with latent variables have been transformed into models of observed variables. There are some differences from the typical simultaneous equation systems such as the complexity of the composite disturbances and the added equations for \mathbf{y}_2 and \mathbf{x}_2 . However, the similarities are sufficient that we can take advantage of the instrumental variable (IV) tests that are available in econometrics.

These statistics test the assumption that all IVs used to achieve identification in an equation are uncorrelated with the disturbance terms. The IV tests require that there are more IVs available than RHS endogenous variables (i.e. the equation is overidentified). The importance of these IV tests to model specification comes from a point made earlier. The IVs for each equation are *model-implied* IVs. That is, if the model is valid, then the IVs chosen for an equation will be uncorrelated with the disturbance. Failing an IV test suggests that at least one of the alleged IVs correlates with the disturbance which in turn means that the model that implied these IVs is structurally misspecified. The structural misspecification can be located in that equation or could be due to a misspecification in another equation that changes the list of eligible IVs for the equation of interest. We will provide examples of this when we present our simulation model.

The same basic idea underlies all of the tests; if the assumption that all IVs are uncorrelated with the disturbance term is valid, then they should not explain any of the variance in the 2SLS residuals. The (uncentered) R^2 from an OLS regression of the 2SLS residuals on the IVs should therefore be zero.

An early test of this sort was developed by Sargan (Sargan 1958). The original form of Sargan test statistic is given by

$$S_{\chi^2} = \frac{\widehat{\mathbf{u}}' \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \widehat{\mathbf{u}}}{\widehat{\mathbf{u}}' \widehat{\mathbf{u}} / N} = \frac{\widehat{\mathbf{u}}' \mathbf{P}_z \widehat{\mathbf{u}}}{\widehat{\mathbf{u}}' \widehat{\mathbf{u}} / N} \quad (4)$$

where $\widehat{\mathbf{u}}$ contains the residuals from the 2SLS estimation, \mathbf{Z} contains the IVs, and N is the sample size. This test can be computed easily as NR^2_u , where R^2_u is the uncentered R^2 from the OLS regression of the 2SLS residuals on the proposed instruments. It follows an asymptotic chi-squared distribution with degrees of freedom equal to the number of excess instruments. The null hypothesis is that the IVs are uncorrelated with the disturbance term and, therefore, a significant test statistic indicates a misspecified model.

Several similar tests have been developed in econometric research. Independently of Sargan, Basman (Basman 1960) developed a test statistic based on the same idea:

$$B_F = \frac{\widehat{\mathbf{u}}' \mathbf{P}_z \widehat{\mathbf{u}} / (L - K)}{\widehat{\mathbf{u}}' \mathbf{M}_z \widehat{\mathbf{u}} / (N - L)} \quad (5)$$

where $\mathbf{M}_z = \mathbf{I} - \mathbf{P}_z$, L is the total number of instruments, and K is the number of RHS endogenous variables. This statistic is asymptotically distributed as an F statistic with $L-K$ and $N-L$ degrees of freedom. Note that Basman's test can be computed from the same auxiliary regression as Sargan's test because $\hat{\mathbf{u}}'\mathbf{M}_z\hat{\mathbf{u}} = \hat{\mathbf{u}}'\hat{\mathbf{u}} - \hat{\mathbf{u}}'\mathbf{P}_z\hat{\mathbf{u}}$. The two tests differ only in the way they estimate the 2SLS residual variance in the denominators. Sargan's test estimates the residual variance with an OLS on the 2SLS residuals with the full set of overidentifying restrictions imposed, while Basman's test uses an estimate of the residual variance from a regression without the overidentifying restrictions imposed (Davidson and MacKinnon 1993). Asymptotically, the tests are equivalent (Baum, Schaffer, and Stillman 2003).

Basman's F test is frequently reported as an asymptotic chi-square statistic without the numerator degrees of freedom or

$$B_{\chi^2} = \frac{\hat{\mathbf{u}}'\mathbf{P}_z\hat{\mathbf{u}}}{\hat{\mathbf{u}}'\mathbf{M}_z\hat{\mathbf{u}}/(N-L)} \quad (6)$$

and there is a pseudo F-distribution form of Sargan's Chi square tests constructed by dividing the numerator $\hat{\mathbf{u}}'\mathbf{P}_z\hat{\mathbf{u}}$ by $L-K$ and replacing N in the denominator with $N-K$:

$$S_F = \frac{\hat{\mathbf{u}}'\mathbf{P}_z\hat{\mathbf{u}}/(L-K)}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(N-K)} \quad (7)$$

Finally, Sargan's asymptotic chi-squared statistic has a version with a small sample correction, replacing the estimated residual variance $\hat{\mathbf{u}}'\hat{\mathbf{u}}/N$ with $\hat{\mathbf{u}}'\hat{\mathbf{u}}/(N-K)$:

$$S_{\chi^2_c} = \frac{\hat{\mathbf{u}}'\mathbf{P}_z\hat{\mathbf{u}}}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(N-K)} \quad (8)$$

All of these tests rest on the same assumptions as the 2SLS estimator (see the previous section) and assume sufficiently large samples. However, there is little empirical evidence on their finite sample performance.

It is tempting to compare the functionality of the 2SLS tests of overidentification to that of the Lagrangian Multiplier (LM) tests from the ML estimator, but the two approaches differ substantially. The LM tests give an estimate of the change in the Chi-squared goodness of fit statistic if a particular restriction is lifted (e.g., a path is freed). While this is potentially useful information, previous research indicates that model respecification using the tests can result in serious errors (MacCallum 1986). Thus, additional diagnostic tools based on a different approach could be valuable. The 2SLS tests, in contrast to the LM tests, identify equations with IVs that are correlated with the disturbance term. If the model was correctly specified, then the IVs would all be suitable. So the failure of the test leads a researcher to consider respecifications of the model that would render one or more of the assumed IVs to be disqualified. In the next section, we discuss the results from a Monte Carlo experiment investigating the properties of all five tests discussed: 1) Sargan's pseudo-F (S_F), 2) Sargan's chi-squared (S_{χ^2}), 3) Sargan's chi-squared with small-sample correction ($S_{\chi^2_c}$), 4) Basman's

F (B_F), and 5) Basman's chi-squared (B_{χ^2}). This experiment examines the performance of the tests under several different misspecifications in three different models with seven different sample sizes ranging from 50 to 1000.

Simulation Study

Models and parameterization

To maximize external validity, we designed three “prototypical” models for our simulation study based on a systematic review of studies using SEM in several key social science journals. The models approximate the size of many SEMs and the structural misspecifications we use vary in magnitude. All models contain three latent variables in a causal chain. Model 1 has between three and four observed indicators per latent variable, Model 2 has five or six indicators per latent variables, and Model 3 has three or four indicators per latent variable plus four exogenous observed variables. Population parameters were chosen to result in a range of effect sizes (communalities and R^2 values ranging from 49% to 72%) and misspecifications were designed to give a wide range of bias in parameters (relative bias ranging from 0 to $\pm 37\%$). As the ML estimator is by far the most frequently used in SEM, we also chose parameter estimates to provide a range of power to detect misspecification with the Chi-square goodness-of-fit test (from 0.07 to 1.0). Figures 1 through 3 show path diagrams that summarize the three models and population parameter values.

We estimate each of the three models with four different specifications. For all models, Specification 1 is *properly* specified in that the models estimated are the same as those with which we generated the data. In other words, Specification 1 corresponds to the “*true*” model in the population while all other specifications are incorrect to some degree. In Model 1, Specification 2 omits the path linking latent variable F2 to observed indicator V7; Specification 3 additionally drops the path linking latent variable F3 with observed indicator V6; Specification 4 additionally drops the factor loading linking latent variable F1 with indicator V4. The specifications for Model 2 are analogous to those of Model 1. Specification 2 for Model 2 drops the factor loading linking latent variable F2 with observed indicator V11; Specification 3 additionally drops the factor loading linking F3 to V10; Specification 4 drops the factor loading linking F1 to V6.

The specifications for Model 3 are different than those for Models 1 and 2 in that the specifications are not all “nested”; selected factor loadings are dropped first, then paths linking exogenous observed variables to latent variables are dropped, then both factor loadings and paths from observed variables to latent variables are dropped. Specifically, Specification 2 for Model 3 omits all three factors loadings as in the fourth specification of Model 1. Specification 3 omits paths linking V10 to F2, V10 to F3, V12 to F2, and V12 to F3. Finally, Specification 4 omits all paths that were omitted in Specification 2 and Specification 3. All of the omitted paths constituting misspecification are indicated by dashed lines in Figures 1 through 3. For a more detailed description of the design of this simulation, see (Paxton, Bollen, Curran, Chen, and Kirby 2001).

Instrumental Variables

Table 1 shows the instrumental variable (IV) sets implied by each model and specification. This illustrates how misspecification affects the choice of IVs for the 2SLS estimator, and how the 2SLS estimator can isolate bias due to misspecification. The IV sets that are not uncorrelated with the equation disturbance terms in the population appear in bold.

For all models, column 1 shows the IVs implied by Specification 1, or the “*true*” model, so there are no incorrect IVs. In Model 1, Specification 2, dropping the path from F2 to V7

results in a single equation with incorrect IVs. As a result of dropping the path, the structure of the model being tested implies that V5 should be included as an IV in the V7 equation. However, in the “true” model (Specification 1), V5 is correlated with the disturbance term in the V7 equation because it contains ϵ_5 and, therefore, violates the assumption of being uncorrelated. In Specification 3 of Model 1, the path from F3 to V6 is omitted in addition to the omitted path in Specification 2. This results in an additional two equations where at least some of the model-implied IVs are incorrect; V8 is included as an IV for the V6 equation, and V6 is included as an IV in the F3 equation. Again, these IVs are inappropriate because they are not uncorrelated with the corresponding equation's disturbance term. All the specifications for Model 2 mirror those in Model 1 and, therefore, we will not discuss them.

In Model 3, Specification 2, three factor loadings are wrongly restricted to zero and these omissions result in four misspecified equations (V4, V6, V7, and F3). This is analogous to Specification 4 of Model 1. In Specification 3 of Model 3, four paths from exogenous observed variables to latent variables are dropped and this results in two equations with incorrect IVs (F2 and F3). Finally, Specification 4 combines Specification 2 and 3 by omitting both the factor loadings and the paths from exogenous variables to latent variables resulting in five misspecified equations (V4, V6, V7, F2, and F3).

The different specifications for Model 3 illustrate two interesting characteristics of the 2SLS estimator. First, in Specification 3, the omitted paths do not change the IVs implied by the model but the omissions nonetheless result in two misspecified equations (F2 and F3). This is because when the RHS exogenous variables are omitted from the second stage of the 2SLS estimation, their variance is subsumed in the disturbance term and, therefore, the disturbance term is correlated with the IVs implied by the structure of model. This is an example of how misspecification can occur not only because inappropriate variables are included as IVs, but because important exogenous variables are omitted from the second stage of the 2SLS estimation. This study will show that both types of specification error are detected using the proposed 2SLS specification tests.

A second important characteristic illustrated by Model 3 can be seen in the F1 equation. Note that this equation has the same number of instruments as there are RHS endogenous variables. In cases like this where an equation is exactly identified, no test of misspecification can be done; the Sargan and Basman tests are not defined. An analogous situation exists with measures of fit from the ML estimator; when a model is perfectly identified as is the case, for example, if a simple regression were estimated with the ML estimator, the observed covariance matrix would be exactly reproduced by the model, and no test of fit is possible.

In a different simulation study, the 2SLS estimator performed somewhat better when a reduced set of IVs were used (Bollen et al. 2007). This is consistent with econometric research showing that it is sometimes better to drop IVs that are only weakly related to the RHS endogenous variables (Staiger and Stock 1997). When using the 2SLS estimator for SEM, even though the model structure may imply that several IVs are correlated with the disturbance term, they may not be strongly related to the endogenous RHS variable. But, while using a reduced IV set may be desirable when using the 2SLS estimator for point estimates, it is not when using the 2SLS estimator for evaluating the specification of a model.³ A model's structure determines what variables should be included as IVs in each equation *unambiguously* and, therefore, all IVs implied by the model should be used when

³Although detecting structural misspecifications might be better accomplished with the full set of IVs, it is possible that the 2SLS estimator will produce estimates more robust to structural misspecifications with a reduced set of IVs. This follows because if the reduced set includes IVs that are proper in the correct model, the 2SLS estimator will be robust (see Bollen, 2001).

evaluating model specification. Using a reduced IV set generally makes the 2SLS estimator less sensitive to misspecification and while this is good when making point estimates, it is not good when evaluating model specification.

Data Generation and Model Estimation

We use the simulation features in Version 5 of EQS (Bentler 1995) to generate the data for each model at seven sample sizes: 50, 75, 100, 200, 400, 800, and 1000. At the smallest sample size ($N=50$), we generated 650 samples, and for all other sample sizes we generated 550 samples. This was done because previous studies based on the same data focused on the ML estimator, which frequently has convergence problems at small sample sizes. These previous studies removed samples with convergence problems (Bollen et al. 2007; Chen, Bollen, Paxton, Curran, and Kirby 2001; Chen, Curran, Bollen, Kirby, and Paxton 2008; Curran, Bollen, Paxton, Kirby, and Chen 2002), but we include all samples generated because the 2SLS estimator is not iterative and all models have formula-based solutions. Our outcomes are the five tests of overidentifying restrictions outlined in the previous section: Sargan's χ^2 (S_{χ^2}), Sargan's F (S_F), Sargan's χ^2 with small sample correction ($S_{\chi^2_c}$), Basman's F statistic (B_F), and Basman's χ^2 statistic (B_{χ^2}).

Results

Table 2 provides the total proportion of equations that are rejected (“rejection rate”) for both properly specified equations and misspecified equations (i.e. those with at least one IV that is correlated with the disturbance term) for each test statistic by sample size. The significance level used to calculate rejection rates is $\alpha=0.05$. Table 2 thus provides a description of how often properly specified equations are wrongly identified as being misspecified⁴, and how often misspecified equations are correctly identified as such.

At small sample sizes, there is considerable variance across the different tests in the proportion of correctly specified equations that are rejected and a general tendency to reject too frequently for all but Sargan's F, which is too low at small sample sizes. Sargan's χ^2 comes the closest to the nominal rejection rate of 0.05 across the sample sizes. Interestingly, it performs better than its counterpart with the small sample correction, which has a rejection rate of 7.28% at $N=50$. Basman's F-test rejects over twice as many properly specified equations as the nominal. All the statistics, however, converge fairly quickly toward the nominal rate and by sample size of 400 all rejection rates for properly specified equations are close to 5%.

Table 2 also shows the proportion of misspecified equations rejected in Specifications 2, 3, and 4. This can be thought of as the average power of the tests to detect misspecified equations across all of the models. At the smallest sample size, $N=50$, none of the tests reject more than 20% of misspecified equations. Basman's F-test rejects the highest percentage (18.91%), but it also rejected the highest percentage of properly specified equations (12.77%). Not surprisingly, as sample size increases, so does the percent of misspecified equations that are rejected and by sample size 1000, all tests reject over 75% of the misspecified equations.

A strength of the specification tests based on the 2SLS estimator is that they identify the equations within a model that have inappropriate IVs (i.e. IVs that are not uncorrelated with the disturbance) and, thus, can be helpful in locating and diagnosing structural

⁴Though incorrectly rejecting a properly specified model would usually be considered a Type II error, because the null hypothesis here is that the model is correct, it is actually a Type I error. Because the null hypotheses for all the tests of overidentifying restrictions in the 2SLS estimator are unconventional in this respect, we do not use the Type I/Type II terminology.

misspecification within a model. This is illustrated by the results shown in Tables 3 through 5. Each table shows the proportion of equations rejected by Sargan's χ^2 tests for each equation in each model by sample size. We focus on a single test for space considerations, but all of the five test statistics give a similar pattern of results. Each equation is identified by its dependent variable (column 1).

In Table 3, it is clear that all the properly specified equations, across all specifications, are rejected at a rate approaching 0.05 with increasing sample size. The power to identify equations with incorrect overidentifying restrictions within a model becomes clear when examining the results for Specifications 2, 3, and 4. In Specification 2, the equation with V7 as the dependent variable is misspecified (see Table 1) and the results in Table 3 clearly show that it is rejected at a much higher rate than the other equations. At sample size 1000, the proportion of equations rejected for the V7 equation in Model 1, Specification 2 is 70%. In Specification 3, there are three misspecified equations (the equations with dependent variables V6, V7, and F3) and the test statistics for these equations clearly stand out. About 79% of the V6 equations, 70% of the V7 equations, and 82% of the F3 equations are identified as having inappropriate IVs. Finally, Specification 4 has four misspecified equations and the rejection rate of each clearly stands out in Table 3. The misspecifications in Specification 3 are also present in Specification 4, so the equations for V6, V7, and F3 have the same rejection rates. The V4 equation is rejected in 83% of the samples.

An important lesson from these findings is that when an equation fails an IV test and is flagged as having inappropriate IVs, it does not always imply that the equation itself contains structural misspecification. A correlation between an equation's disturbance term and its IVs can sometimes be induced by structural misspecification in other parts of a model. An example of this can be seen in the F3 equation in Model 1, Specification 3. There is no structural misspecification in the F3 equation but, rather, there is an omitted path in the V6 equation that in turn implies an incorrect IV for the F3 equation. Thus, when analysts use the 2SLS tests of overidentifying restrictions to assess model specification, they should investigate all model restrictions that imply the set of IVs in question, not just those in the equation that failed the test.

A pattern of results very similar to that of Model 1 appears for Model 2 (Table 4), but the rejection rates are generally higher; the tests seem to have more power in Model 2 than in Model 1. Between 82% and 87% of equations with inappropriate IVs are identified as such by the 2SLS specification tests in Model 2. Model 2 has more total degrees of freedom than Model 1, as does each equation being tested, and this apparently translates into increased power in detecting misspecification.

In Model 3, as in all the Models, Specification 1 matches the population model; it is properly specified. Thus, in the first panel of Table 5, no equation stands out; all converge to something close to the nominal rate of 0.05 by sample size 1000. In Model 3, Specification 2, we omit the three factor loadings as in Model 1, Specification 4. The rejection rates for Specification 2 of Model 3 shown in the second panel of Table 5 look similar to those seen for Specification 4 of Model 1, but with a wider range. At the largest sample size, the rejection rates for the misspecified equations range from 63% for the F3 equation to 93% for the V4 equation. At the smallest sample size, between 10% and 17% of misspecified equations were rejected. Note that the only difference between Model 1, Specification 4 and Model 3 Specification 2 is that the latter has four exogenous variables affecting the latent variables. Though the exogenous variables add information to the system, it is not clear how this affects the power to detect misspecification. For the V6, V7, and F3 equations, the rejection rates are actually lower in Model 3 than in Model 1, but the rejection rate for the V4 equation is much higher in Model 3 than in Model 1.

In the third specification of Model 3, four paths from the exogenous variables to the latent variables are dropped (V10 to F2, V10 to F3, V12 to F2, and V12 to F3). These omissions cause the F2 and F3 equations to be misspecified. Even though the instrument sets for each equation are identical to those in the properly specified model (Specification 1), the omissions result in misspecification that is easily detected by the 2SLS tests. The third panel of Table 5 shows that by sample size 400, nearly 100% of both misspecified equations are identified as such, and by sample size 1000, all misspecified equations are identified.

Finally, Specification 4 of Model 3 omits both factor loadings and paths from exogenous variables to latent variables resulting in five misspecified equations. The results for Specification 4 in Table 5 show that, at the largest sample size, equation rejection rates range from 67% to 100%. These results look almost like a superimposition of the results for Specification 2 and Specification 3, but there is one interesting deviation. The F3 equation is misspecified in both Specification 2 and Specification 3. In Specification 4, however, only the omitted paths from the exogenous variables have an effect. More specifically, in Specification 2, V6 is incorrectly included as an IV in the 2SLS estimator, and this results in a rejection rate of 63% at the largest sample size. In specification 3, the IVs are the same, but the paths from the exogenous variables are omitted, and this causes the IVs to be correlated with the equation's disturbance. This gives a rejection rate of 100% at the largest sample size. The two misspecifications are not additive in that the omission of exogenous variables dominates the 2SLS test of misspecification.

Discussion

In this paper, we propose a group of related specification tests derived from the limited information 2SLS estimator for latent variable structural equation models (Bollen 1996; Bollen 2001). We find that the proposed tests identify most misspecified models, even when the misspecification is relatively minor (one cross-loading omitted), and that well-specified models are wrongly identified as being misspecified at the expected rate. Our simulation suggests that the 2SLS tests are especially sensitive to misspecifications involving the omission of structural paths between latent variables (as opposed to factor loadings). In addition to identifying misspecified models, the tests we propose have the advantage of identifying the equations within a model where IVs are not uncorrelated with disturbance terms. Analysts can use this information to help diagnose structural misspecification by investigating the model restrictions that imply the IVs in question. This makes the tests, and the 2SLS least squares estimator, valuable to analysts using structural equation modeling techniques.

Our simulation results are relevant not only to research involving the estimation of SEMs, but also to research that uses 2SLS estimation on observed-variable models. The Basman and Sargan statistics investigated in this paper have been used to test overidentifying restrictions in 2SLS estimation since the early 1960s, yet we have not found other large-scale simulation experiment that investigates their small sample properties. Though the population models used in our simulation experiment were designed based on research using SEMs, the experimental results provide a general gauge of the performance of the tests of overidentification for researchers using 2SLS estimation in more traditional models.

As with all simulation studies, our findings should be interpreted carefully. Generalizability is always open to question; our results may depend on the sort of models we chose for our experiment, as well as the parameters we chose. We did, however, go to great lengths to make our models realistic with respect to these attributes. This includes conducting a systematic review of the literature, and carefully choosing models and parameter values that are typical of SEM applications.

One potential addition to this line of research is the possibility of an omnibus misspecification test based on Hausman's general test (Hausman 1978) and comparing the ML and 2SLS estimates of coefficients. We found that the Hausman test performed poorly in our experiment. This could be because Hausman's test is not very powerful and a sample size of 1000 was not adequate to bring out its asymptotic properties. There are alternative calculation formulae for the Hausman test in simultaneous equation models (Mroz 1987), and these may provide a test with better properties. If so, this would provide a 2SLS-based omnibus specification test for model evaluation.

The Confirmatory Tetrad Analysis test for SEM is another overall fit statistic that could be used in conjunction with the 2SLS estimator (Bollen 1990; Bollen and Ting 1993). This provides a single measure of the fit of the whole model and could complement the overidentifying equation tests that were the subject of our analysis. In addition, it would be possible to adapt the overall chi square tests for the PIV estimator for categorical variables described in Bollen and Maydeu-Olivares (Bollen and Maydeu-Olivares 2007) to continuous variables as another fit measure for a model where all coefficients are estimated with 2SLS. However, both of these tests would be for the model as a whole rather than providing a separate test for each overidentified equation.

The specification tests we propose and evaluate in this study complement the ML estimator and its associated measures of fit. While most full information estimators evaluate model fit by measuring how well the model reproduces the covariance matrix (i.e. how close the model-implied covariance matrix is to the sample covariance matrix), the 2SLS specification tests we propose evaluate models equation-by-equation by testing whether the restrictions implied by the structure of the model are consistent with the data. This makes the set of specification tests a valuable addition to the wide variety of model evaluation tools in structural equation modeling. Moreover, the tests we propose can help analysts diagnose the source of misspecification in the model. Finally, the 2SLS estimator and the specification tests are readily available in most statistical packages, are much less computationally intensive than the ML estimator, and have no problems with convergence or improper solutions. There is thus little reason not to perform 2SLS estimation and compute the specification tests when estimating structural equation models.

References

- Basmann RL. A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation. *Econometrica*. 1957; 25:77–83.
- Basmann RL. On finite sample distributions on generalized classical linear identifiability test statistics. *Journal of the American Statistical Association*. 1960; 55:650–659.
- Baum, Christopher F.; Schaffer, Mark E.; Stillman, Steven. Working Paper 545. Boston College Department of Economics; 2003. Instrumental variables and GMM: Estimation and testing.
- Bentler, PM. EQS: Structural equations program manual, version 5.0. Los Angeles, CA: BMDP Statistical Software; 1995.
- Bollen, Kenneth A. Structural equations with latent variables. New York: Wiley; 1989.
- Bollen, Kenneth A. A Comment on Model Evaluation and Modification. *Multivariate Behavioral Research*. 1990; 25:181–185.
- Bollen, Kenneth A. An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*. 1996; 61:109–121.
- Bollen, Kenneth A. Two-stage least squares and latent variable models: simultaneous estimation and robustness to misspecifications. In: Cudeck, R.; Du Toit, S.; Sorbom, D., editors. *Structural Equation Modeling: Present and Future*. Lincolnwood, IL: Scientific Software; 2001.

- Bollen, Kenneth A.; Kirby, James B.; Curran, Patrick J.; Paxton, Pamela M.; Chen, Feinian. Latent Variable Models Under Misspecification: Two-Stage Least Squares (2SLS) and Maximum Likelihood (ML) Estimators. *Sociological Methods and Research*. 2007; 36:48–86.
- Bollen, Kenneth A.; Maydeu-Olivares, A. A polychoric instrumental variable (PIV) estimator for structural equation models with categorical variables. *Psychometrika*. 2007; 72:309–326.
- Bollen, Kenneth A.; Ting, KF. Confirmatory Tetrad Analysis. *Sociological Methodology* 1993, Vol 23. 1993; 23:147–175.
- Browne MW. Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*. 1984; 37:62–83. [PubMed: 6733054]
- Chen, Feinian; Bollen, Kenneth A.; Paxton, Pamela M.; Curran, Patrick J.; Kirby, James B. Improper solutions in structural equation models - Causes, consequences, and strategies. *Sociological Methods & Research*. 2001; 29:468–508.
- Chen, Feinian; Curran, Patrick J.; Bollen, Kenneth A.; Kirby, James B.; Paxton, Pamela M. An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*. 2008; 36:462–494. [PubMed: 19756246]
- Curran, Patrick J.; Bollen, Kenneth A.; Paxton, Pamela M.; Kirby, James B.; Chen, Feinian. The noncentral chi-square distribution in misspecified structural equation models: Finite sample results from a Monte Carlo simulation. *Multivariate Behavioral Research*. 2002; 37:1–36.
- Davidson, R.; MacKinnon, JG. Estimation and inference in econometrics. New York: Oxford University Press; 1993.
- Häggglund, Goesta. Factor Analysis by Instrumental Methods: A Monte Carlo Study of Some Estimation Procedures. University of Upsala; Sweden: Department of Statistics; Uppsala, Sweden: 1983.
- Hausman JA. Specification tests in econometrics. *Econometrica*. 1978; 46:1251–1271.
- Jöreskog, Karl G. A General Method for Estimating a Linear Structural Equation System. In: Goldberger, AS.; Duncan, OD., editors. *Structural Equation Models in the Social Science*. New York: Academic Press; 1973. p. 85-112.
- Jöreskog, Karl G. Factor Analysis as an Error-in-variables Model in *Principles of Modern Psychological Measurement*. Wainer, H.; Messick, S., editors. Hillsdale, NH: Lawrence Erlbaum Associates; 1983.
- Jöreskog, Karl G.; Sörbom, Dag. LISREL 8: User's Reference Guide. Chicago: Scientific Software International; 1996.
- Lance CE, Cornwell JM, Mulaik SA. Limited Information Parameter Estimates for Latent or Mixed Manifest and Latent Variable Models. *Multivariate Behavioral Research*. 1988; 23:155–167.
- MacCallum, Robert. Specification Searches in Covariance Structure Modeling. *Psychological Bulletin*. 1986; 100:107.
- Madansky, Albert. Instrumental Variables in Factor Analysis. *Psychometrika*. 1964; 29:105–113.
- Mroz, Thomas A. The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*. 1987; 55:765–799.
- Paxton, Pamela M.; Bollen, Kenneth A.; Curran, Patrick J.; Chen, Feinian; Kirby, James B. Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*. 2001; 8:287–312.
- Sargan J. The estimation of economic relationships using instrumental variables. *Econometrica*. 1958; 26:393–415.
- Douglas, Staiger; Stock, James H. Instrumental variable regression with weak instruments. *Econometrica*. 1997; 65:557–586.
- Theil, H. Estimation and Simultaneous Correlation in Complete Equation Systems. Central Planning Bureau; The Hague, Netherlands: 1953.

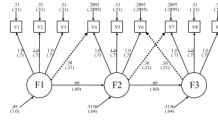


Figure 1.

Path diagram representation of Model 1

Note: numbers shown are unstandardized parameter values with standardized values in parenthesis; solid and dashed lines represent the population model structure, and dashed lines represent omitted parameters under model misspecification.

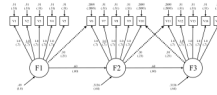


Figure 2.

Path diagram representation of Model 2

Note: numbers shown are unstandardized parameter values with standardized values in parenthesis; solid and dashed lines represent the population model structure, and dashed lines represent omitted parameters under model misspecification.

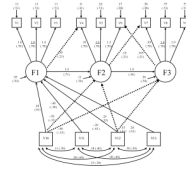


Figure 3.

Path diagram representation of Model 3.

Note: numbers shown are unstandardized parameter values with standardized values in parenthesis; solid and dashed lines represent the population model structure, and dashed lines represent omitted parameters under model misspecification.

Table 1

Instruments used in each 2SLS equation by model and specification

Dependent Variable	Instruments Implied by the Model Being Tested			
	Specification 1 (Properly Specified)	Specification 2	Specification 3	Specification 4
<u>Model 1</u>				
V1	V3-V9	V3-V9	V3-V9	V3-V9
V3	V1,V4-V9	V1,V4-V9	V1,V4-V9	V1,V4-V9
V4	V1,V3,V6-V9	V1,V3,V6-V9	V1,V3,V6-V9	V1-V3,V6-V9
V6	V1-V4,V7,V9	V1-V4,V7,V9	V1-V4,V7-V9	V1-V4,V7-V9
V7	V1-V4, V6, V9	V1-V6,V9	V1-V6,V9	V1-V6,V9
V9	V1-V7	V1-V7	V1-V7	V1-V7
F2	V1,V3	V1,V3	V1,V3	V1,V3
F3	V1-V4	V1-V4	V1-V4,V6	V1-V4,V6
<u>Model 2</u>				
V1	V3-V15	V3-V15	V3-V15	V3-V15
V3	V1,V4-V15	V1,V4-V15	V1,V4-V15	V1,V4-V15
V4	V1,V3,V5-V15	V1,V3,V5-V15	V1,V3,V5-V15	V1,V3,V5-V15
V5	V1,V3,V4,V6-V15	V1,V3,V4,V6-V15	V1,V3,V4,V6-V15	V1,V3,V4,V6-V15
V6	V1,V3-V5,V8-V15	V1,V3-V5,V8-V15	V1,V3-V5,V8-V15	V1-V5,V8-V15
V8	V1-V6,V9-V15	V1-V6,V9-V15	V1-V6,V9-V15	V1-V6,V9-V15
V9	V1-V6,V8,V10-V15	V1-V6,V8,V10-V15	V1-V6,V8,V10-V15	V1-V6,V8,V10-V15
V10	V1-V6,V8,V9,V11, V13-V15	V1-V6,V8,V9,V11,V13-V15	V1-V6,V8,V9,V11-V15	V1-V6,V8,V9,V11-V15
V11	V1-V6,V8-V10,V13-V15	V1-V10,V13-V15	V1-V10,V13-V15	V1-V10,V13-V15
V13	V1-V11,V14,V15	V1-V11,V14,V15	V1-V11,V14,V15	V1-V11,V14,V15
V14	V1-V11,V13,V15	V1-V11,V13,V15	V1-V11,V13,V15	V1-V11,V13,V15
V15	V1-V11,V13,V14	V1-V11,V13,V14	V1-V11,V13,V14	V1-V11,V13,V14
F2	V1,V3,V4,V5	V1,V3,V4,V5	V1,V3,V4,V5	V1,V3,V4,V5
F3	V1-V6,V8,V9	V1-V6,V8,V9	V1-V6,V8-V10	V1-V6,V8-V10
<u>Model 3</u>				
V1	V3-V9-V13	V3-V9-V13	V3-V9-V13	V3-V9-V13
V3	V1,V4-V9-V13	V1,V4-V9-V13	V1,V4-V9-V13	V1,V4-V9-V13

Dependent Variable	Instruments Implied by the Model Being Tested			
	Specification 1 (Properly Specified)	Specification 2	Specification 3	Specification 4
V4	V1,V3,V6-V9-V13	V1-V3,V6-V13	V1,V3,V6-V9-V13	V1-V3,V6-V13
V6	V1-V4,V7,V9-V13	V1-V4,V7-V13	V1-V4,V7,V9-V13	V1-V4,V7-V13
V7	V1-V4,V6,V9-V13	V1-V6,V9-V13	V1-V4,V6,V9-V13	V1-V6,V9-V13
V9	V1-V7, V10-V13	V1-V7, V10-V13	V1-V7, V10-V13	V1-V7, V10-V13
F2	V1,V3, V10-V13	V1,V3, V10-V13	V1,V3, V10-V13*	V1,V3, V10-V13*
F3	V1-V4, V10-V13	V1-V4,V6,V10-V13	V1-V4, V10-V13*	V1-V4,V6,V10-V13
F1	No endogenous RHS vars (OLS-exactly identified)	No endogenous RHS vars (OLS-exactly identified)	No endogenous RHS vars (OLS-exactly identified)	No endogenous RHS vars (OLS-exactly identified)

Note: Instruments sets that include incorrect IVs are in **boldface**

* IVs are the same as those implied by the "true" model, but the equation has omitted explanatory variables which induces endogeneity

Table 2
Proportion of equations identified as having inappropriate IVs by five test statistics, by sample size

	Sample Size						
	50	75	100	200	400	800	1000
Sargan's pseudo F (S_F)							
Percentage of properly specified equations incorrectly identified as having misspecification	2.84%	4.94%	4.98%	5.96%	4.68%	4.97%	4.92%
Percentage of misspecified equations correctly identified as having misspecification	6.80%	11.18%	14.47%	24.86%	40.51%	68.11%	75.10%
Sargan's Chi-squared (S_{χ^2})							
Percentage of properly specified equations incorrectly identified as having misspecification	5.51%	6.63%	6.37%	6.83%	5.02%	5.11%	5.06%
Percentage of misspecified equations correctly identified as having misspecification	10.72%	14.26%	16.98%	26.20%	41.39%	68.43%	75.47%
Sargan's Chi-squared with small sample correction ($S_{\chi^2_c}$)							
Percentage of properly specified equations incorrectly identified as having misspecification	7.28%	7.57%	7.12%	7.23%	5.22%	5.14%	5.09%
Percentage of misspecified equations correctly identified as having misspecification	13.07%	16.01%	17.96%	26.79%	41.64%	68.55%	75.63%
Basmann's F (B_F)							
Percentage of properly specified equations incorrectly identified as having misspecification	12.77%	10.41%	9.51%	8.28%	5.62%	5.33%	5.20%
Percentage of misspecified equations correctly identified as having misspecification	18.91%	20.05%	20.52%	28.20%	42.63%	68.96%	75.88%
Basmann's Chi-squared (B_{χ^2})							
Percentage of properly specified equations incorrectly identified as having misspecification	8.34%	8.04%	7.52%	7.37%	5.26%	5.15%	5.11%
Percentage of misspecified equations correctly identified as having misspecification	14.29%	16.48%	18.22%	27.00%	41.79%	68.62%	75.66%

Table 3
Proportion of equations failing Sargan's Chi-squared test (S_{χ^2}) by equation and sample size, Model 1

Equation Dependent Variable	Sample Size						
	50	75	100	200	400	800	1000
Specification 1							
V1	0.08	0.06	0.06	0.07	0.06	0.06	0.04
V3	0.07	0.07	0.07	0.06	0.06	0.06	0.04
V4	0.06	0.05	0.07	0.05	0.04	0.05	0.04
V6	0.06	0.06	0.07	0.04	0.04	0.03	0.05
V7	0.05	0.06	0.09	0.05	0.06	0.04	0.04
V9	0.07	0.07	0.07	0.06	0.05	0.04	0.06
F2	0.04	0.04	0.05	0.03	0.04	0.04	0.05
F3	0.07	0.03	0.04	0.06	0.07	0.05	0.04
Specification 2							
V1	0.08	0.06	0.06	0.07	0.06	0.06	0.04
V3	0.07	0.07	0.07	0.06	0.06	0.06	0.04
V4	0.06	0.05	0.07	0.05	0.04	0.05	0.04
V6	0.06	0.06	0.07	0.04	0.04	0.03	0.05
V7*	0.14	0.15	0.16	0.17	0.35	0.63	0.70
V9	0.07	0.07	0.07	0.06	0.05	0.04	0.06
F2	0.04	0.04	0.05	0.03	0.04	0.04	0.05
F3	0.07	0.03	0.04	0.06	0.07	0.05	0.04
Specification 3							
V1	0.08	0.06	0.06	0.07	0.06	0.06	0.04
V3	0.07	0.07	0.07	0.06	0.06	0.06	0.04
V4	0.06	0.05	0.07	0.05	0.04	0.05	0.04
V6*	0.12	0.13	0.15	0.22	0.38	0.72	0.79
V7*	0.14	0.15	0.16	0.17	0.35	0.63	0.70

Equation Dependent Variable	Sample Size						
	50	75	100	200	400	800	1000
V9	0.07	0.07	0.07	0.06	0.05	0.04	0.06
F2	0.04	0.04	0.05	0.03	0.04	0.04	0.05
F3*	0.10	0.11	0.12	0.23	0.46	0.78	0.82
Specification 4							
V1	0.08	0.06	0.06	0.07	0.06	0.06	0.04
V3	0.07	0.07	0.07	0.06	0.06	0.06	0.04
V4*	0.14	0.13	0.17	0.19	0.44	0.70	0.83
V6*	0.12	0.13	0.15	0.22	0.38	0.72	0.79
V7*	0.14	0.15	0.16	0.17	0.35	0.63	0.70
V9	0.07	0.07	0.07	0.06	0.05	0.04	0.06
F2	0.04	0.04	0.05	0.03	0.04	0.04	0.05
F3*	0.10	0.11	0.12	0.23	0.46	0.78	0.82

* Equation is misspecified

Table 4
 Proportion of equations failing Sargan's Chi-squared test (S_{χ^2}) by equation and sample size, Model 2

Equation Dependent Variable	Sample size						
	50	75	100	200	400	800	1000
Specification 1							
V1	0.07	0.06	0.09	0.05	0.06	0.04	0.05
V3	0.07	0.07	0.07	0.08	0.05	0.06	0.04
V4	0.06	0.07	0.06	0.09	0.06	0.05	0.05
V5	0.08	0.05	0.04	0.08	0.06	0.06	0.04
V6	0.07	0.07	0.06	0.07	0.05	0.05	0.07
V8	0.08	0.08	0.06	0.08	0.04	0.06	0.07
V9	0.07	0.08	0.06	0.07	0.05	0.05	0.04
V10	0.07	0.06	0.08	0.07	0.05	0.06	0.06
V11	0.07	0.08	0.08	0.06	0.04	0.07	0.05
V13	0.06	0.06	0.08	0.07	0.06	0.04	0.04
V14	0.08	0.08	0.08	0.06	0.05	0.05	0.06
V15	0.08	0.07	0.06	0.04	0.06	0.06	0.05
F2	0.05	0.05	0.04	0.05	0.05	0.06	0.07
F3	0.05	0.04	0.04	0.05	0.04	0.06	0.05
Specification 2							
V1	0.07	0.06	0.09	0.05	0.06	0.04	0.05
V3	0.07	0.07	0.07	0.08	0.05	0.06	0.04
V4	0.06	0.07	0.06	0.09	0.06	0.05	0.05
V5	0.08	0.05	0.04	0.08	0.06	0.06	0.04
V6	0.07	0.07	0.06	0.07	0.05	0.05	0.07
V8	0.08	0.08	0.06	0.08	0.04	0.06	0.07
V9	0.07	0.08	0.06	0.07	0.05	0.05	0.04
V10	0.07	0.06	0.08	0.07	0.05	0.06	0.06
V11*	0.14	0.17	0.18	0.21	0.36	0.72	0.82
V13	0.06	0.06	0.08	0.07	0.06	0.04	0.04

Equation Dependent Variable	Sample size						
	50	75	100	200	400	800	1000
V14	0.08	0.08	0.08	0.06	0.05	0.05	0.06
V15	0.08	0.07	0.06	0.04	0.06	0.06	0.05
F2	0.05	0.05	0.04	0.05	0.05	0.06	0.07
F3	0.05	0.04	0.04	0.05	0.04	0.06	0.05
Specification 3							
V1	0.07	0.06	0.09	0.05	0.06	0.04	0.05
V3	0.07	0.07	0.07	0.08	0.05	0.06	0.04
V4	0.06	0.07	0.06	0.09	0.06	0.05	0.05
V5	0.08	0.05	0.04	0.08	0.06	0.06	0.04
V6	0.07	0.07	0.06	0.07	0.05	0.05	0.07
V8	0.08	0.08	0.06	0.08	0.04	0.06	0.07
V9	0.07	0.08	0.06	0.07	0.05	0.05	0.04
V10*	0.14	0.13	0.17	0.22	0.38	0.71	0.85
V11*	0.14	0.17	0.18	0.21	0.36	0.72	0.82
V13	0.06	0.06	0.08	0.07	0.06	0.04	0.04
V14	0.08	0.08	0.08	0.06	0.05	0.05	0.06
V15	0.08	0.07	0.06	0.04	0.06	0.06	0.05
F2	0.05	0.05	0.04	0.05	0.05	0.06	0.07
F3*	0.08	0.08	0.14	0.21	0.41	0.77	0.87
Specification 4							
V1	0.07	0.06	0.09	0.05	0.06	0.04	0.05
V3	0.07	0.07	0.07	0.08	0.05	0.06	0.04
V4	0.06	0.07	0.06	0.09	0.06	0.05	0.05
V5	0.08	0.05	0.04	0.08	0.06	0.06	0.04
V6*	0.15	0.13	0.14	0.23	0.38	0.71	0.85
V8	0.08	0.08	0.06	0.08	0.04	0.06	0.07
V9	0.07	0.08	0.06	0.07	0.05	0.05	0.04

Equation	Dependent Variable	Sample size							
		50	75	100	200	400	800	1000	
V10*		0.14	0.13	0.17	0.22	0.38	0.71	0.85	
V11*		0.14	0.17	0.18	0.21	0.36	0.72	0.82	
V13		0.06	0.06	0.08	0.07	0.06	0.04	0.04	
V14		0.08	0.08	0.08	0.06	0.05	0.05	0.06	
V15		0.08	0.07	0.06	0.04	0.06	0.06	0.05	
F2		0.05	0.05	0.04	0.05	0.05	0.06	0.07	
F3*		0.08	0.08	0.14	0.21	0.41	0.77	0.87	

* Equation is misspecified

Table 5
 Proportion of equations failing Sargan's Chi-squared test (S_{χ^2}) by equation and sample size, Model 3

Equation Dependent Variable	Sample size						
	50	75	100	200	400	800	1000
Specification 1							
V1	0.06	0.07	0.07	0.06	0.04	0.05	0.05
V3	0.08	0.08	0.05	0.05	0.06	0.04	0.07
V4	0.07	0.07	0.06	0.06	0.04	0.05	0.06
V6	0.08	0.08	0.07	0.08	0.05	0.06	0.05
V7	0.06	0.09	0.08	0.07	0.06	0.06	0.05
V9	0.08	0.07	0.07	0.08	0.06	0.06	0.04
F2	0.06	0.05	0.05	0.05	0.05	0.05	0.05
F3	0.07	0.05	0.04	0.07	0.05	0.05	0.03
Specification 2							
V1	0.06	0.07	0.07	0.06	0.04	0.05	0.05
V3	0.08	0.08	0.05	0.05	0.06	0.04	0.07
V4*	0.17	0.17	0.16	0.25	0.48	0.84	0.93
V6*	0.15	0.16	0.14	0.22	0.31	0.66	0.72
V7*	0.14	0.16	0.13	0.22	0.27	0.58	0.67
V9	0.08	0.07	0.07	0.08	0.06	0.06	0.04
F2	0.06	0.05	0.05	0.05	0.05	0.05	0.05
F3*	0.10	0.10	0.08	0.15	0.29	0.54	0.63
Specification 3							
V1	0.06	0.07	0.07	0.06	0.04	0.05	0.05
V3	0.08	0.08	0.05	0.05	0.06	0.04	0.07
V4	0.07	0.07	0.06	0.06	0.04	0.05	0.06
V6	0.08	0.08	0.07	0.08	0.05	0.06	0.05
V7	0.06	0.09	0.08	0.07	0.06	0.06	0.05

Equation Dependent Variable	Sample size						
	50	75	100	200	400	800	1000
V9	0.08	0.07	0.07	0.08	0.06	0.06	0.04
F2*	0.29	0.53	0.65	0.96	1.00	1.00	1.00
F3*	0.17	0.26	0.28	0.64	0.95	1.00	1.00
Specification 4							
V1	0.06	0.07	0.07	0.06	0.04	0.05	0.05
V3	0.08	0.08	0.05	0.05	0.06	0.04	0.07
V4*	0.17	0.17	0.16	0.25	0.48	0.84	0.93
V6*	0.15	0.16	0.14	0.22	0.31	0.66	0.72
V7*	0.14	0.16	0.13	0.22	0.27	0.58	0.67
V9	0.08	0.07	0.07	0.08	0.06	0.06	0.04
F2*	0.29	0.53	0.65	0.96	1.00	1.00	1.00
F3*	0.17	0.26	0.28	0.64	0.95	1.00	1.00

* Equation is misspecified