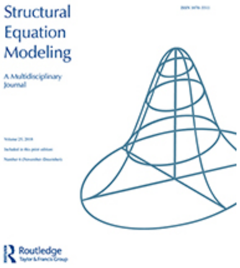

Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: http://www.tandfonline.com/loi/hsem20

Robustness Conditions for MIIV-2SLS When the Latent Variable or Measurement Model is Structurally Misspecified

Kenneth A. Bollen, Kathleen M. Gates & Zachary Fisher

Structural Equation Modeling: A Multidisciplinary Journal, 25: 848–859, 2018

ISSN: 1070-5511 print / 1532-8007 online
DOI: https://doi.org/10.1080/10705511.2018.1456341

Robustness Conditions for MIIV-2SLS When the Latent Variable or Measurement Model is Structurally Misspecified

Kenneth A. Bollen, Kathleen M. Gates, and Zachary Fisher
University of North Carolina at Chapel Hill

Most researchers acknowledge that virtually all structural equation models (SEMs) are approximations due to violating distributional assumptions and structural misspecifications. There is a large literature on the unmet distributional assumptions, but much less on structural misspecifications. In this paper, we examine the robustness to structural misspecification of the model implied instrumental variable, two-stage least square (MIIV-2SLS) estimator of SEMs. We introduce two types of robustness: robust-unchanged and robust-consistent. We develop new robustness analytic conditions for MIIV-2SLS and illustrate these with hypothetical models, simulated data, and an empirical example. Our conditions enable a researcher to know whether, for example, a structural misspecification in the latent variable model influences the MIIV-2SLS estimator for measurement model equations and vice versa. Similarly, we establish robustness conditions for correlated errors. The new robustness conditions provide guidance on the types of structural misspecifications that affect parameter estimates and they assist in diagnosing the source of detected problems with MIIVs.

Keywords: 2SLS, MIIV, model implied instrumental variables, robustness, structural misspecifications

> …it is implausible that any model that we uese is anything more than an approximation to reality. (Browne & Cudeck, 1993, p. 137)

INTRODUCTION

Most researchers readily acknowledge that virtually all structural equation models (SEMs) are approximations (e.g., Browne & Cudeck, 1993; MacCallum, Browne, & Cai, 2007; McDonald, 2010). Two primary sources of approximation to reality are failures of the distributional assumptions (e.g., normality) and structural misspecifications. Structural misspecifications (structural errors) take a variety of forms such as omitted variables, incorrect latent variable paths, neglected correlated errors, or even the wrong dimensionality of our measures. The consequences of structural misspecifications are more serious than those of distributional misspecifications. The widely used maximum likelihood (ML) estimator in SEM remains a consistent estimator if the only problem is distributional misspecifications, though significance tests can be biased (Browne, 1984). In contrast, structural misspecifications affect both the consistency (e.g., Yuan et al., 2003) and the significance tests of the ML estimator (e.g., Kolenikov & Bollen, 2012).

The SEM literature on the consequences of and the robustness to structural errors is sparse with nearly all existing work dedicated to the system wide ML estimator (e.g., Kaplan, 1989; Kaplan & Wenger, 1993; Yuan et al., 2003; Yuan, Kouros, & Kelley, 2008). The MIIV-2SLS (model implied instrumental variable, two-stage least squares) estimator for SEMs (Bollen, 1996) has received far less attention. Bollen (2001) provides two general conditions on MIIV-2SLS's robustness to structural misspecification. Several other studies (Bollen, Kirby, Curran, Paxton, & Chen, 2007; Bollen & Maydeu-Oliveres, 2007; Jin, Luo, & Yang-Wallentin, 2016; Nestler, 2013) provide simulation

Correspondence should be addressed to Kenneth Bollen, Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, CB 3270 Davie Hall, Chapel Hill, NC 27599. E-mail: bollen@unc.edu

evidence that the MIIV-2SLS estimator is more robust to structural misspecifications than the more widely used system wide (e.g., ML) estimators but do not give general *analytic* conditions for robustness. Our goal is to provide new, specific, conditions of when the MIIV-2SLS estimator will be robust to structural misspecifications.

In models with structural misspecifications, at least some of the MIIV-2SLS coefficient estimates are biased.[1] No estimator is immune to these problems for *all* of its equations. However, particular equations can be robust to specification errors located elsewhere in the model. It is important to know which MIIV-2SLS coefficient estimates are robust to structural misspecifications. For instance, suppose we have great uncertainty about the specification of the latent variable model, but we have more confidence in the measurement model specification. If we use MIIV-2SLS technique, will misspecification errors in the latent variable model contaminate estimates of the measurement model? Alternatively, suppose the latent variable model is well established, but the measurement model is not. Will errors in the measurement model bias the estimation of the latent variable model?

Bollen's (2001) general robustness conditions for MIIV-2SLS, which we describe later, are not designed to characterize possible cross contamination between the latent variable and measurement models. Indeed, we know of no study that addresses these robustness questions. The primary purpose of our paper is to develop new, more specific conditions for when the MIIV-2SLS estimator of equation coefficients is robust to structural misspecification. We also explain how these robustness conditions help in diagnosing the source of misspecifications. Our results are analytical, but we illustrate them with simulated and empirical data.

In the next section, we introduce the model notation and assumptions. An overview of the MIIV-2SLS approach follows. Then, we present a section on robustness conditions that first reviews Bollen's (2001) general robustness conditions and second presents new conditions for the consequences of misspecified latent variable models for the measurement model and vice versa. We follow these results with several simulation examples followed by an empirical example. Finally, we give the conclusions and implications.

MODEL

We use a slight modification of the LISREL notation (Jöreskog & Sörbom, 2001; Bollen, 2001) to represent the general SEM:

$$\begin{aligned} \boldsymbol{\eta} &= \boldsymbol{\alpha}_\eta + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \\ \mathbf{y} &= \boldsymbol{\alpha}_y + \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon} \\ \mathbf{x} &= \boldsymbol{\alpha}_x + \boldsymbol{\Lambda}_y\boldsymbol{\xi} + \boldsymbol{\delta} \end{aligned} \qquad (1)$$

where the first equation gives the latent variable model and the next two equations give the measurement model. In the first equation, $\boldsymbol{\eta}$ is the vector of latent endogenous variables, $\boldsymbol{\xi}$ is the vector of latent exogenous variables, and $\boldsymbol{\zeta}$ is the vector of equation errors or disturbances. The $\mathbf{B}$ matrix gives the expected effects of the latent endogenous variables on each other and the $\boldsymbol{\Gamma}$ matrix gives the expected effects of the latent exogenous variables ($\boldsymbol{\xi}$) on the latent endogenous variables ($\boldsymbol{\eta}$). The $\boldsymbol{\alpha}_\eta$ is the intercept term that gives the expected value of the dependent variable of an equation when all the right-hand side (RHS) variables are zero. ζ contains the errors (disturbances) and we assume that $E(\boldsymbol{\zeta}) = \mathbf{0}$ and $\boldsymbol{C}(\boldsymbol{\xi}, \boldsymbol{\zeta}) = \mathbf{0}$.

The last two equations give the measurement model, where $\mathbf{y}$ is the vector of measures of $\boldsymbol{\eta}$, $\mathbf{x}$ is the vector of measures of $\boldsymbol{\xi}$, $\boldsymbol{\alpha}_y$ and $\boldsymbol{\alpha}_x$ are the intercept terms, $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ are the factor loadings, and $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are the unique factors (errors) of the indicators. The unique factors have means of zero $[E(\boldsymbol{\varepsilon}) = \mathbf{0}, E(\boldsymbol{\delta}) = \mathbf{0}]$, are uncorrelated with their respective latent variables $[C(\boldsymbol{\varepsilon}, \boldsymbol{\eta}) = \mathbf{0}, C(\boldsymbol{\varepsilon}, \boldsymbol{\xi}) = \mathbf{0}, C(\boldsymbol{\delta}, \boldsymbol{\xi}) = \mathbf{0}]$, and the errors and unique factors are uncorrelated with each other $[C(\boldsymbol{\varepsilon}, \boldsymbol{\delta}) = \mathbf{0}, C(\boldsymbol{\varepsilon}, \boldsymbol{\zeta}) = \mathbf{0}, C(\boldsymbol{\delta}, \boldsymbol{\zeta}) = \mathbf{0}]$. Note that we do not make normality assumptions about the observed variables or errors when using MIIV-2SLS. The MIIV-2SLS is robust to nonnormality (see Bollen, 1996).

MODEL IMPLIED INSTRUMENTAL VARIABLE APPROACH

Robustness to structural misspecifications depends on the estimator. We are using the MIIV-2SLS estimator from Bollen (1996). Those familiar with it can skip this section. For others, we describe the MIIV-2SLS estimator so that the paper is self-contained. The first step in the MIIV-2SLS procedure is to transform the SEM in latent variables to an equivalent form in observed variables. We abbreviate this latent to observed variable transformation as L2O and briefly describe it in the next subsection.

L2O transformation

Each latent variable is scaled to one of its indicators by setting the scaling indicator's intercept to zero and its factor loading to one (Bollen, 1989). The vector $\mathbf{y}$ contains all indicators of $\boldsymbol{\eta}$. We partition $\mathbf{y}$ so that all scaling indicators come first and the nonscaling indicators come second resulting in

[1] One possible exception to this claim is if dropping an indicator is considered a structural misspecification. In that situation, the estimators of the remaining parameters need not be biased. But this is a less common meaning for structural misspecification.

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_s \\ \mathbf{y}_{ns} \end{bmatrix} \quad (2)$$

where $\mathbf{y}_s$ is the vector of scaling indicators and $\mathbf{y}_{ns}$ is the vector of nonscaling indicators. In a similar fashion, we partition the $\mathbf{x}$ vector to get

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_{ns} \end{bmatrix} \quad (3)$$

with analogous definitions.

The equations for just the scaling indicators are

$$\begin{aligned} \mathbf{y}_s &= \boldsymbol{\eta} + \boldsymbol{\varepsilon}_s \\ \mathbf{x}_s &= \boldsymbol{\xi} + \boldsymbol{\delta}_s \end{aligned} \quad (4)$$

We can reexpress these equations as

$$\begin{aligned} \boldsymbol{\eta} &= \mathbf{y}_s - \boldsymbol{\varepsilon}_s \\ \boldsymbol{\xi} &= \mathbf{x}_s - \boldsymbol{\delta}_s \end{aligned} \quad (5)$$

Substituting this equation into the equations of Equation (1), we get

$$\begin{aligned} \mathbf{y}_s &= \boldsymbol{\alpha}_\eta + \mathbf{B}\mathbf{y}_s + \boldsymbol{\Gamma}\mathbf{x}_s + \boldsymbol{\varepsilon}_s - \boldsymbol{B}\boldsymbol{\varepsilon}_s - \boldsymbol{\Gamma}\boldsymbol{\delta}_s + \boldsymbol{\zeta} \\ \mathbf{y}_{ns} &= \boldsymbol{\alpha}_{y,ns} + \boldsymbol{\Lambda}_{y,ns}\mathbf{y}_s - \boldsymbol{\Lambda}_{y,ns}\boldsymbol{\varepsilon}_s + \boldsymbol{\varepsilon}_{ns} \\ \mathbf{x}_{ns} &= \boldsymbol{\alpha}_{x,ns} + \boldsymbol{\Lambda}_{x,ns}\mathbf{x}_s - \boldsymbol{\Lambda}_{x,ns}\boldsymbol{\delta}_s + \boldsymbol{\delta}_{ns} \end{aligned} \quad (6)$$

where the "*s*" subscript refers to vectors (matrices) corresponding to the scaling indicators and "*ns*" signifies vector (matrices) of nonscaling indicators. Moving from Equation (1) to Equation (6) in observed variables is the L2O transformation.

In the observed variable form, it is typical to have the composite error correlate with variables on the RHS of the equation. For example, the composite error $(\boldsymbol{\varepsilon}_s - B\boldsymbol{\varepsilon}_s - \boldsymbol{\Gamma}\boldsymbol{\delta}_s + \boldsymbol{\zeta})$ in the first equation in Equation (6) generally correlates with some elements of $\mathbf{y}_s$ and $\mathbf{x}_s$ because the errors of these scaling indicators are part of the composite error. The correlation of the error with the RHS variables rules out ordinary least squares (OLS) because this biases the OLS estimates. Bollen (1996, 2001) shows that an instrumental variable estimator is applicable. Perhaps more importantly, Bollen (1996) shows that in general an identified SEM has sufficient instruments among the observed variables already in the model. These are the model implied instrumental variables (MIIVs) for the MIIV-2SLS estimator.

Estimators

The L2O transformation and the procedure for finding MIIVs enable a variety of instrumental variable estimators. Bollen (1996) developed the MIIV-2SLS while Bollen, Kolenikov, and Bauldry (2014) proposed the generalized method of moments (GMM) estimator [MIIV-GMM]. Other instrumental variable variants are possible. We focus on MIIV-2SLS because it is better known and because MIIV-2SLS is likely more robust than other multi-equation estimators.

Details on the MIIV-2SLS estimator are in several places (e.g., Bollen, 1996, 2001; Bollen et al., 2007). Here, we illustrate the method with the latent variable model in Equation (6). Consider the *j*th equation from this L2O version of the latent variable model:

$$y_j = \boldsymbol{\alpha}_j + \mathbf{B}_j y_s + \boldsymbol{\Gamma}_j x_s + u_j \quad (7)$$

where y_j is the *j*th dependent variable from $\mathbf{y}_s$, $\boldsymbol{\alpha}_j$ is that equation's intercept, $\mathbf{B}_j$ and $\boldsymbol{\Gamma}_j$ are the *j*th row in $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$, respectively, and u_j is the *j*th element of $\boldsymbol{u}$, where $\boldsymbol{u}$ is the composite error term.

To put this into a more convenient form with which to present the MIIV-2SLS estimator, let $\boldsymbol{y}_j$ and $\boldsymbol{u}_j$ be $N \times 1$ vectors of the values for the *j*th equation dependent variable and composite residual, respectively.[2] Define $\boldsymbol{A}_j$ as a column vector of coefficients that has $\boldsymbol{\alpha}_j$ and all the nonzero elements of $\mathbf{B}_j$ and $\boldsymbol{\Gamma}_j$ stacked in a column. Let $\boldsymbol{Z}_j$ be an N row matrix with 1s in the first column and the N row elements of $\boldsymbol{y}_s$ and $\boldsymbol{x}_s$ that have nonzero coefficients for the equation in the remaining columns. Then we can rewrite Equation (7) as

$$\boldsymbol{y}_j = \boldsymbol{Z}_j\boldsymbol{A}_j + \boldsymbol{u}_j \quad (8)$$

To estimate the coefficients in $\boldsymbol{A}_j$ collect all the MIIVs for the *j*th equation into an N row matrix $\boldsymbol{V}_j$ with a vector of 1s in the first column and the values of the MIIVs for the *j*th equation in the remaining columns where $j = 1, 2, 3, \ldots, J$ and J is the total number of equations in the model. The MIIV-2SLS estimator of coefficients is $\widehat{A}_j = (\widehat{\boldsymbol{Z}}_j\widehat{\boldsymbol{Z}}_j)^{-1}\widehat{\boldsymbol{Z}}'_j\boldsymbol{y}_j$, where $\widehat{\boldsymbol{Z}}_j = \boldsymbol{V}_j(\boldsymbol{V}'_j\boldsymbol{V}_j)^{-1}\boldsymbol{V}'_j\boldsymbol{Z}_j$. A similar series of steps would hold to find the MIIV-2SLS estimator of the factor loadings and intercepts from the measurement model.

As discussed elsewhere (e.g., Bollen, 1996), the MIIV-2SLS estimator is a consistent and asymptotic unbiased estimator of the coefficients for a correctly specified model. A concern of ours is when does the MIIV-2SLS remain a consistent estimator of the coefficients of an equation when there are misspecifications in other equations. Prior to turning to this issue, we have a brief discussion on MIIVs.

[2] Note that y_j below equation (7) refers to a single value while $\boldsymbol{y}_j$ in the footnoted sentence refers to all $N \times 1$ values of this variable.

MIIVs

Finding the MIIVs is an important part of the MIIV-2SLS procedure. For each equation in Equation (6), the MIIVs are those observed variables that would be uncorrelated with the composite disturbance of that equation *if the model of interest were valid*. In other words, the MIIV approach uses the model structure to determine those observed variables that are uncorrelated with the disturbances. Of course, the model needs not be correct and there are test statistics to determine the lack of correlation between the MIIVs and composite errors as implied by the model structure (Kirby & Bollen, 2009).

The MIIVs for a given equation should satisfy several conditions: (a) the MIIVs ($\boldsymbol{V}_j$) should be uncorrelated with the composite error of the equation ($\boldsymbol{u}_j$), (b) the MIIVs ($\boldsymbol{V}_j$) should correlate with the explanatory variables ($\boldsymbol{Z}_j$), and (c) the rank of $\boldsymbol{V}_j^{'}\boldsymbol{Z}_j$ equals the number of columns in $\boldsymbol{Z}_j$ (e.g., Bollen, 1996, p. 114). An identified SEM has observed variables that are uncorrelated with the composite error of each equation and satisfies the preceding conditions. See Bollen (1996) for more details. SAS and Stata procedures are available for finding MIIVs (Bollen & Bauer, 2004; Bauldry, 2014) and the R package MIIVsem (Fisher, Bollen, Gates & Rönkkö, 2017) can be used to find MIIVs and estimate the model parameters using MIIV-2SLS.

ROBUSTNESS CONDITIONS FOR MIIV-2SLS

General robustness to structural misspecification

We define two different meanings of "robust." Consider the coefficient A_{jk} where the subscript j indexes the J equations in the model with $j = 1, 2, 3, \ldots, J$ and k indexes the parameter in the jth equation where $k = 1, \ldots, K_j$, with K_j the total number of coefficients in the jth equation. To avoid repetition and more notation, we use the coefficient vector $\boldsymbol{A}_j$ in a general fashion so that it refers to the coefficients of an L2O equation from either the latent variable or the measurement model. The MIIV-2SLS estimator of A_{jk} is $\hat{A}_{jk}$ and we use $\hat{\mathbf{A}}_j$ without the k subscript to reference all coefficients in the jth equation as we did earlier. The symbol M_0 refers to a model with no structural misspecifications in any equation, M_1 references a second model with at least one structural misspecification, and M_2 denotes a third model with different structural misspecifications than M_1. The parameters of all three models can be nested or nonnested.

We also propose two different types of robustness to structural misspecifications. *Robust-unchanged* occurs when $\hat{A}_{jk}$ (or $\hat{\mathbf{A}}_j$) takes the same value for a given pair of models. Robust-unchanged permits structural misspecifications in both models, for instance M_1 and M_2, or just one, as when comparing M_0 and M_1. This is in harmony with the idea that models are approximations and as such we might be comparing two models each of which involves structural misspecifications or is approximate. The only requirement for robust-unchanged is that $\hat{A}_{jk}$ (or $\hat{\mathbf{A}}_j$) is the same value when comparing two models.

Robust-consistent occurs when $\hat{A}_{jk}$ (or $\hat{\mathbf{A}}_j$) remains a consistent estimator under at least two different models (M_1 and M_2, M_0 and M_1, or M_0 and M_2). The MIIV-2SLS estimator $\hat{A}_{jk}$ (or $\hat{\mathbf{A}}_j$) is a consistent estimator of A_{jk} (or $\mathbf{A}_j$) if $plim\left(\hat{\mathbf{A}}_{jk}\right) = \mathbf{A}_{jk}$]. The MIIV-2SLS estimator is a consistent estimator for the correct model M_0 for all coefficients and all equations (e.g., Bollen, 1996). Furthermore, the MIIV-2SLS estimator might be a consistent estimator of A_{jk} ($\mathbf{A}_j$) under some M_1 or M_2. The MIIV-2SLS estimators $\hat{A}_{jk}$ $\left(\hat{\mathbf{A}}_j\right)$ will be consistent in model M_1 if equation j is correctly specified and the MIIVs chosen using M_1 are the same or are a subset of the MIIVs obtained using M_0 (Bollen, 2001). One exception to the requirements of a correctly specified equation is when there are omitted variables that are uncorrelated with the included variables of the equation. Note also that robust-consistent implies robust-unchanged, but not vice versa. In other words, the MIIV-2SLS estimator can be robust-unchanged for an equation but not robust-consistent because both equations produce the same inconsistent estimator.

Distinguishing between robust-unchanged and robust-consistent is important for at least two reasons. First, these definitions of robustness allow us to consider approximate models that contain errors, a situation that is typical. Our equation can fall short of perfection and hence have MIIV-2SLS not be a consistent estimator, yet the estimator might still be useful or close to consistency. Without prejudging the validity of the model structure, we can say something about whether structural misspecifications elsewhere in the system affect our estimates of a specific coefficient ($\hat{A}_{jk}$) or all coefficients ($\hat{\mathbf{A}}_j$) in a particular equation. This holds even when the particular equation contains specification errors. Second, knowing that the MIIV-2SLS $\hat{A}_{jk}$ (or $\hat{\mathbf{A}}_j$) is robust-unchanged to a certain set of structural misspecifications means that if a test for structural errors [such as the Sargan's (1958) test] reveals a potential problem in the equation, then we can rule out these misspecifications as the source of the problem. In other words, if the MIIV-2SLS $\hat{A}_{jk}$ (or $\hat{\mathbf{A}}_j$) is robust-unchanged in M_1 and M_2, then we know that the structural misspecifications that distinguish M_2 from M_1 are not the source of misspecification. We need to look elsewhere for the structural errors.

The only robustness conditions for the MIIV-2SLS estimator of which we know are from Bollen (2001). He describes two conditions for the MIIV-2SLS to be robust-consistent to structural misspecifications: (1) the equation of interest is correctly specified and (2) the list of MIIVs for the equation of interest remains the same whether the other equations are correctly or incorrectly specified.

The first condition requires that there is not an omitted variable or other structural misspecification in the equation to estimate. For example, if an equation omits a variable that correlates with the included RHS variables, the remaining coefficients are likely to be adversely affected.[3] The second condition asks us to consider the list of MIIVs we would obtain for a given equation when the whole model is correctly specified and the list of MIIVs when it is not. If the list of MIIVs is identical, then the MIIV-2SLS estimator for a given equation is robust-consistent to the structural misspecifications. The robust-consistent status of one equation can differ from that of another in the same model and under the same structural misspecifications. For this reason, we cannot generalize about robustness of MIIV-2SLS for one equation based solely on the robustness of another equation.

These conditions are a valuable first step in that they can tell us whether to expect robustness for a given correctly specified equation. But they do not tell us when the MIIV-2SLS is robust-unchanged, a weaker condition of robustness that is more applicable to approximate models. These general robust-unchanged conditions do tell us about consistency when the MIIVs stay the same across the correct and structurally misspecified models. They do not, however, give us general conditions for when we can expect the MIIVs to be the same. We develop new conditions that generalize to larger parts of the model such as whether structural misspecifications in the latent variable model affect the MIIV-2SLS estimator for equations from the measurement model. Or under what conditions will our MIIV-2SLS estimator of the latent variable model parameters be robust-unchanged to structural misspecifications in the measurement model? These questions are relevant because they can help us find the general location of specification errors when the MIIVs correlate with the composite error for a particular equation. The conditions we develop are devised so that $\hat{A}_{jk}$ $\left(\hat{\mathbf{A}}_j\right)$ are robust-unchanged and are applicable to approximate as well as exactly valid equations. In those cases where consistency of the MIIV-2SLS holds, our same conditions are robust-consistent. It is to these issues that we turn.

Robustness of measurement model to misspecified latent variable model

The robustness question of this section is whether structural misspecifications in the latent variable model ($\boldsymbol{\eta} = \boldsymbol{\alpha}_\eta + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \zeta$) alter the MIIV-2SLS estimator of coefficients for equations from the measurement model. Answering this question is valuable in several ways. First, rather than focusing on a single equation, we would rather have conditions that apply to all equations in the measurement model. In other words, if we develop a general result for this issue, we do not need to check individual equations and can be confident in the robustness of all equations from the measurement model.

[3] An omitted variable that is uncorrelated with all included variables is a rare exception.

Answering this question is also important when we find a poorly fitting equation in the measurement model because we have overidentification tests for each overidentified equation (Bollen, 1996; Kirby & Bollen, 2009). These test whether all MIIVs are uncorrelated with the equation error (as they will be in the population if the model structure is correct). Kirby and Bollen (2009) found that the Sargan (1958) overidentification test works reasonably well across different sample sizes and thus is the test used herein. Rejection tells us that misspecification in the model is likely. If the Sargan's test rejects the null hypothesis that the MIIVs are uncorrelated with the composite error, then the question is where to look for the misspecification. If we can for the first time establish that structural misspecifications in the latent variable model do not influence the MIIVs for the measurement model, we know an unfavorable test statistic is due to structural misspecifications in the measurement model rather than in the latent variable model.

Our starting point is to describe the two general types of structural misspecifications in the latent variable model that we consider: (1) $\mathbf{B}$ or $\boldsymbol{\Gamma}$ have incorrect fixed to zero or freed coefficients and (2) covariance matrix of $\boldsymbol{\zeta}$ has incorrect fixed or free parameters. Condition (1) occurs when latent variables are omitted or needlessly included in one or more equations from the latent variable model. Condition (2) happens when equation errors (ζs) correlate but are fixed to zero or when covariances of errors or their variances are needlessly introduced.

Having described these two general types of structural misspecifications in the latent variable model, we turn to the question of whether these conditions affect the selection of MIIVs for the measurement model. Below, we repeat the L2O equations for the measurement model,

$$\begin{array}{l}\mathbf{y}_{ns} = \boldsymbol{\alpha}_{y,ns} + \boldsymbol{\Lambda}_{y,ns}\mathbf{y}_s - \boldsymbol{\Lambda}_{y,ns}\boldsymbol{\varepsilon}_s + \boldsymbol{\varepsilon}_{ns} \\ \mathbf{x}_{ns} = \boldsymbol{\alpha}_{x,ns} + \boldsymbol{\Lambda}_{x,ns}\mathbf{x}_s - \boldsymbol{\Lambda}_{x,ns}\boldsymbol{\delta}_s + \boldsymbol{\delta}_{ns}\end{array}. \tag{9}$$

The MIIVs for an equation in the measurement model are selected because they are uncorrelated with the composite disturbance of that equation. Condition (1) stipulates that the wrong latent variables are included or excluded from one or more of the latent variable equations. This would manifest itself in incorrect $\mathbf{B}$ or $\boldsymbol{\Gamma}$ matrices.

We consider whether changes in $\mathbf{B}$ or $\boldsymbol{\Gamma}$ affect the MIIVs for a given equation in the measurement model. In general, the MIIVs are those observed variables that are *not* directly or indirectly influenced by the composite errors of a given equation. We can see this by either deriving the direct and indirect effects of the error using algebraic methods such as in Bollen (1987) or by tracing effects of the errors in a path diagram. In addition, any observed variables that are directly or indirectly influenced by an error that correlates with the composite error are ineligible to be a MIIV for that equation (Bollen, 1996). In Equation (9), note only measurement errors are part of the composite errors

$(-\mathbf{\Lambda}_{y,ns}\boldsymbol{\varepsilon}_s + \boldsymbol{\varepsilon}_{ns}, -\mathbf{\Lambda}_{x,ns}\boldsymbol{\delta}_s + \boldsymbol{\delta}_{ns})$. Furthermore, measurement errors only have direct effects (no indirect effects) on their respective indicators. This means that indicators whose errors are included in the composite error cannot be MIIVs. The only other observed variables eliminated are those that have an error (unique factor) that correlates with any errors that are part of the composite error. The pattern of fixed and freed elements in either $\mathbf{B}$ or $\mathbf{\Gamma}$ has no effect on the MIIVs because these elements cannot change the direct effects of the composite errors or the direct effects of any measurement errors that correlate with the composite errors. Moreover, changes in the pattern of fixed and freed elements in the covariance matrix of $\boldsymbol{\zeta}$ cannot change the direct effects of the composite errors or the correlated measurement errors from the L2O form of the measurement model. Based on these results, we reach the following conclusion: *the MIIV-2SLS estimator of any measurement model coefficient is robust-unchanged to structural misspecifications of using the wrong pattern of fixed or freed elements in either* $\mathbf{B}$ or $\mathbf{\Gamma}$ *or using the wrong covariance matrix of* $\boldsymbol{\zeta}$ *in the latent variable model.*

We can test whether all MIIVs are uncorrelated with the error of an equation as long as we have more MIIVs than the required minimum (e.g., Kirby & Bollen, 2009). An implication of our preceding finding is that if we find a measurement equation that fails the test, the failure is due to the measurement model specification and not to the specification of the latent variable model. In other words, we should seek structural misspecifications in the measurement model and not in the latent variable model if we fail the overidentification test for a measurement equation. We now turn to a different question. Is the MIIV-2SLS estimator of the latent variable model robust to structural misspecifications in the measurement model?

Robustness of latent variable model to misspecified measurement model

In this section, we consider two general types of measurement model misspecification: (1) Covariance matrices of $\boldsymbol{\varepsilon}_{y,ns}$ or $\boldsymbol{\varepsilon}_{x,ns}$ have incorrect fixed or free parameters and (2) $\boldsymbol{\Lambda}_{y,ns}$ or $\boldsymbol{\Lambda}_{x,ns}$ have incorrect fixed or free parameters. Condition (1) happens when there are excluded correlated unique factors or incorrect constraints on variances of the unique factors of the nonscaling indicators. Condition (2) occurs when we fail to include direct effects from one latent variable to one or more indicators or needlessly include a factor loading that is zero in the factor loading matrix.

The L2O form of the latent variable model that we estimate is

$$\mathbf{y}_s = \boldsymbol{\alpha}_\eta + \mathbf{B}\mathbf{y}_s + \mathbf{\Gamma}\mathbf{x}_s + \boldsymbol{\varepsilon}_s - B\boldsymbol{\varepsilon}_s - \Gamma\boldsymbol{\delta}_s + \boldsymbol{\zeta}. \tag{10}$$

To determine the robustness of the MIIV-2SLS estimator of Equation (10), we look at each type of structural misspecification and determine whether the set of MIIVs for a latent variable equation is influenced by the misspecification.

As was true previously, any observed variable that is directly or indirectly influenced by any error from the composite error for an equation is ineligible to be a MIIV for that equation. Furthermore, any observed variable that is directly or indirectly influenced by an error that correlates with the composite error is also ineligible. MIIVs are altered as a result of structural misspecification when the set of observed variables which are directly or indirectly influenced by the composite error change as a result of this misspecification. We start with condition (1). The structural misspecification under condition (1) includes omitting correlated errors (uniquenesses) among the nonscaling indicators ($\boldsymbol{\varepsilon}_{y,ns}$ or $\boldsymbol{\varepsilon}_{x,ns}$). Examination of the composite error in equation (10) reveals the absence of any nonscaling unique factors. Hence, the failure to properly model correlated uniquenesses has no effect on the selection of the MIIVs for the latent variable model. In other words, *MIIV-2SLS estimates of the latent variable model parameters are robust-unchanged to the misspecification of covariances among the errors of nonscaling indicators.*

Condition (2) is more complicated to analyze. We first consider the situation where a factor loading to a nonscaling indicator is incorrectly assumed to be zero. That is, a latent variable influences a nonscaling indicator but the model incorrectly omits that path. Because the unique factor of the scaling indicator's only effect is its direct effect on the scaling indicator, these scaling indicator's unique factors have no consequences for the selection of MIIVs when there is an omitted factor loading to a nonscaling indicator. Therefore, we can ignore this part of the composite error.

The only remaining term in the composite error is $\boldsymbol{\zeta}$. This term is more complicated than the unique factors in that $\boldsymbol{\zeta}$ can have indirect effects on more than one observed variable in the model. For instance, consider the causal chain of $\xi i_1 \rightarrow \eta_1 \rightarrow \eta_2$ in Figure 1. The error term (ζ_1) for η_1 indirectly affects all of the indicators of η_1 and all of the indicators of η_2.

Leaving out the path labeled, A is a misspecification of the measurement model. Applying the L2O transformation, the η_1 equation is,

$$y_1 = \alpha_{\eta_1} + \gamma_{11}x_1 - \gamma_{11}\delta_1 + \varepsilon e_1 + \zeta_1 \tag{11}$$

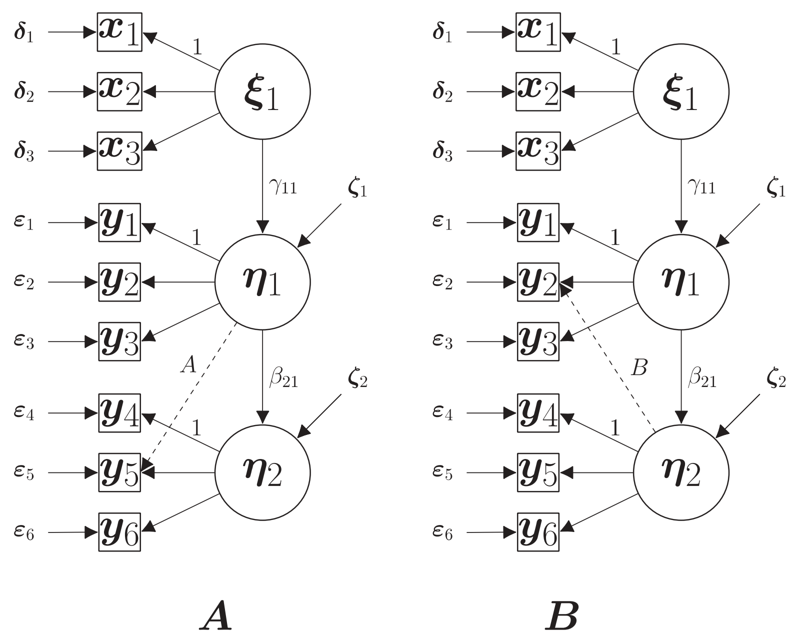

FIGURE 1 Illustration of misspecified measurement models with omitted cross-loadings. Solid and dashed lines comprise the data-generating model; dashed lines represent those omitted in the misspecification.

with MIIVs of x_1 and x_2. If we now use the correct model including the $\eta_1 \rightarrow y_5$ path, the L2O transformation is the same as are the MIIVs, x_1 and x_2 In other words, the MIIV-2SLS estimator of the latent variable equation for η_1 is robust to this omitted path in the measurement model. Furthermore, we could demonstrate that MIIV-2SLS estimation of the η_2 latent variable equation is also robust to this omitted path in the measurement model.

In contrast, now look at Figure 1b. The misspecified measurement model here omits the path labeled B. Interestingly, the L2O transformation and MIIV-2SLS estimation of the η_1 latent variable equation are robust to this omitted path, just like we found for the first misspecified equation. But a different story emerges for the η_2 latent variable equation. The L2O transformation leads to,

$$y_4 = \alpha_{\eta_1} + \beta_{21} y_1 - \beta_{21} \varepsilon_1 + \varepsilon_4 + \zeta_2 \quad (12)$$

with MIIVs x_1, x_2, y_2, and y_3. However, in the true measurement model, there is an indirect path from $\zeta_2 \rightarrow \eta_2 \rightarrow y_2$ and this leads y_2 not to be a MIIV. Therefore, this second type of misspecification leaves MIIV-2SLS to be robust for the first latent variable equation but not for the second latent variable equation.

Generalizing from this example, *if the omitted cross-loading when present creates a nonzero indirect path from the latent variable equation error* (ζ_j) *to one or more of the MIIVs from the misspecified model, then the MIIV-2SLS estimator is not robust-unchanged. Alternatively, if the omitted cross-loading when present does not create a nonzero indirect path from the latent variable equation error* (ζ_j) *to one or more of the MIIVs from the misspecified model, then the MIIV-2SLS is robust-unchanged.*

There are useful implications of these robustness conditions. If an equation from the latent variable model fails an overidentification test, then we know that this failure is not due to omitted correlated uniquenesses among the nonscaling indicators. However, if we look toward the measurement model, the failure could be due to particular omitted factor loadings. We can determine candidates for this omitted factor loading by considering those initial MIIVs that might be indirectly influenced by the equation error (the ζ for the equation).[4] Though this does not pinpoint the offending variable, it can narrow down the search.

Before closing this section, we briefly mention two other robust-unchanged conditions to add to the first two conditions of when a misspecified measurement model will not alter the MIIV-2SLS estimates for the latent variable model. The third one focuses on correlated errors for the scaling indicators for those latent variables that are part of the equation of interest: *The MIIV-2SLS estimator is robust-unchanged to the omission of correlated unique factors for the scaling indicators of latent variables that are included in the latent variable equation*. The reason is that these scaling indicators that are part of the L2O transformed version of the latent variable equation are excluded as MIIVs to start

[4] Of course, a misspecified equation in the latent variable model also is a possibility.

with and this is true whether their unique factors correlate or not. For example, suppose in a latent variable equation, latent variable one directly affects a second latent variable two. In the L2O transformation, the scaling indicators of these two latent variables and their respective unique factors will appear and hence rule both variables out as MIIVs. Whether the unique factors of these two scaling indicators correlate or not makes no difference because they are already not part of the MIIVs. This condition is particularly useful with longitudinal models where correlated errors are likely.

Finally, a fourth condition expresses the robustness of the MIIV-2SLS estimator of a latent variable equation when a researcher drops observed variables from an analysis. The MIIV-2SLS is robust-unchanged to: *the omission of any observed variable that is not part of the L2O transformed equation or is not among the MIIVs for that equation*. Though obvious when pointed out, this does provide another type of robust-unchanged condition involving dropping observed variables from an analysis and when the MIIV-2SLS estimates remain unaltered.

SIMULATION ILLUSTRATIONS

The preceding analytic conditions tell us when MIIV-2SLS is robust and are more comprehensive than any simulation design. However, simulated data are helpful in illustrating the robustness conditions when we know the true structure.

Robustness of measurement model to misspecified latent variable model

To illustrate the robustness of the measurement model to misspecification of the latent variable model, we generated 1000 observations using the correctly specified model (M_0) in Figure 2. We then estimate the MIIV-2SLS coefficients for the correct and incorrect latent variable models (M_1) using MIIVsem (Fisher et al., 2017). The analytic results predict that the factor loadings coefficients will be identical under either specification. Table 1 shows that the coefficient estimates are the same.

Robustness of latent variable model to misspecified measurement model

Here, we illustrate the effect of two different measurement model misspecifications (e.g., M_1 or M_2) on the MIIV-2SLS latent variable parameter estimates in Figure 3. Again, we generated 1000 observations under the correct model and found the MIIV-2SLS estimates for both the correct and incorrect model specifications.

The estimated coefficients for the first example (see Figure 3a) are in Table 2. These results also confirm the previous analytic results, as the estimates of γ_{11} and β_{21} are

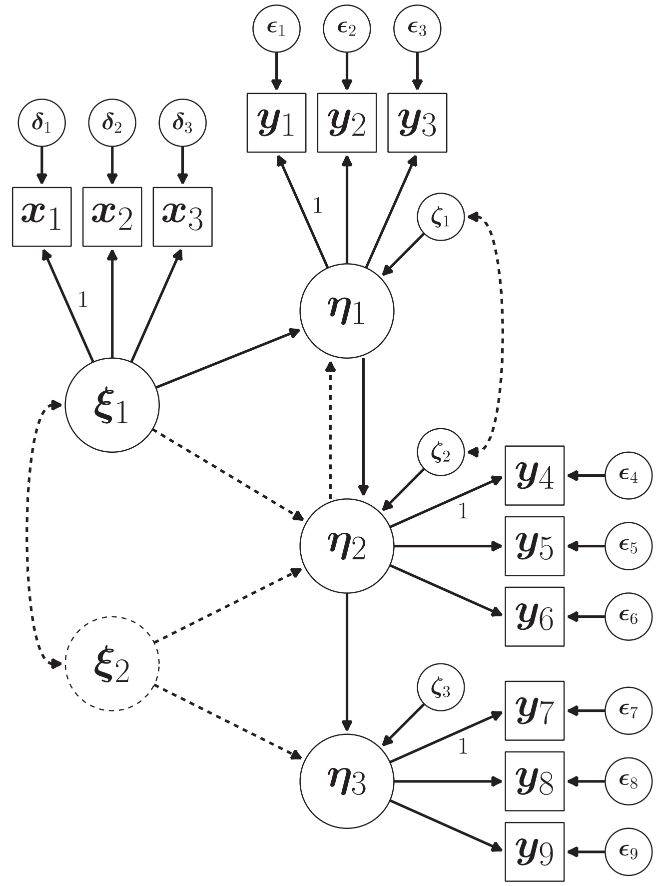

FIGURE 2 Misspecified latent variable model. Solid and dashed lines comprise the data-generating model; dashed lines represent those omitted in the misspecification.

robust to the omission of cross-loadings and covariances amongst the unique factors. It is worth mentioning even if we mistakenly omit correlations amongst the unique factors of the scaling indicators these estimates would still be robust.

In the second example (M_2), we again examine the impact of omitted correlations among the unique factors and omitted factor loadings (see Figure 3b). Now, y_3 loads on both η_1 and η_2. Unlike the previous example, this creates an indirect effect from ζ_2 to y_3, making y_3 an ineligible MIIV for the second structural equation under the true specification. Including y_3 as a MIIV due to model misspecification biases the parameter estimate for β_{21} and Table 3 shows that the estimate of β_{21} is 20% too large.[5]

[5] Fortunately when y_3 is incorrectly included as a MIIV in the η_1 equation, the Sargan's test (Kirby & Bollen, 2009) correctly rejects the null hypothesis($\chi^2 = 105$, $p<0.001$), detecting that at least one of the MIIVs correlates with the error.

TABLE 1
Measurement Model Estimates Under Correct and Incorrect Specification of Latent Variable Model (N = 1000)

		Latent Variable Model Specification	
Parameter	*Population*	*Correct*	*Incorrect*
λ_{y_2}	1	1.00	1.00
λ_{y_3}	1	1.00	1.00
λ_{y_5}	1	1.00	1.00
λ_{y_6}	1	0.99	0.99
λ_{y_8}	1	1.00	1.00
λ_{y_9}	1	1.02	1.02
λ_{x_2}	1	0.99	0.99
λ_{x_3}	1	1.00	1.00

TABLE 2
Measurement Model Estimates for Misspecification in Figure 3a

		Latent Variable Model Specification	
Parameter	*Population*	*Correct*	*Incorrect*
λ_{11}	1	1.05	1.05
β_{21}	1	1.03	1.03

TABLE 3
Latent Variable Coefficient Estimates Under Correct and Incorrect Measurement Model Specifications in Figure 3b

		Latent Variable Model Specification	
Parameter	*Population*	*Correct*	*Incorrect*
λ_{11}	1	1.02	1.02
β_{21}	1	0.97	1.20

EMPIRICAL EXAMPLE

To demonstrate these concepts, we present a series of models based on the industrialization and political democracy example initially presented in Bollen (1989). The data contain information on 75 developing countries from 1960 ("industrialization" and "political democracy" variables) as well as 1965 ("political democracy"). The model has three indicators of industrialization and the same four indicators measure political democracy in both years. With real empirical data, we do not know the true structure. But these data have the benefit of having been evaluated for a specific model in terms of its appropriateness, interpretability, and model fit. Specifically, we can examine misspecifications against the "original" baseline model wherein industrialization in 1960 (ξ_1) predicts political democracy in years 1960 and 1965, and political democracy in 1960 (η_1) predicts political democracy in 1965 (η_2; see Figure 4).

First, we investigate robustness to misspecification of the latent variable model. Here, one path is omitted (see Figure 4a) which induces a causal chain wherein ξ_1 predicts η_1 that in turn predicts η_2. Table 4 presents the estimates of the measurement model. As can be seen, the factor loadings are identical when estimated with the original or the causal chain model. Given

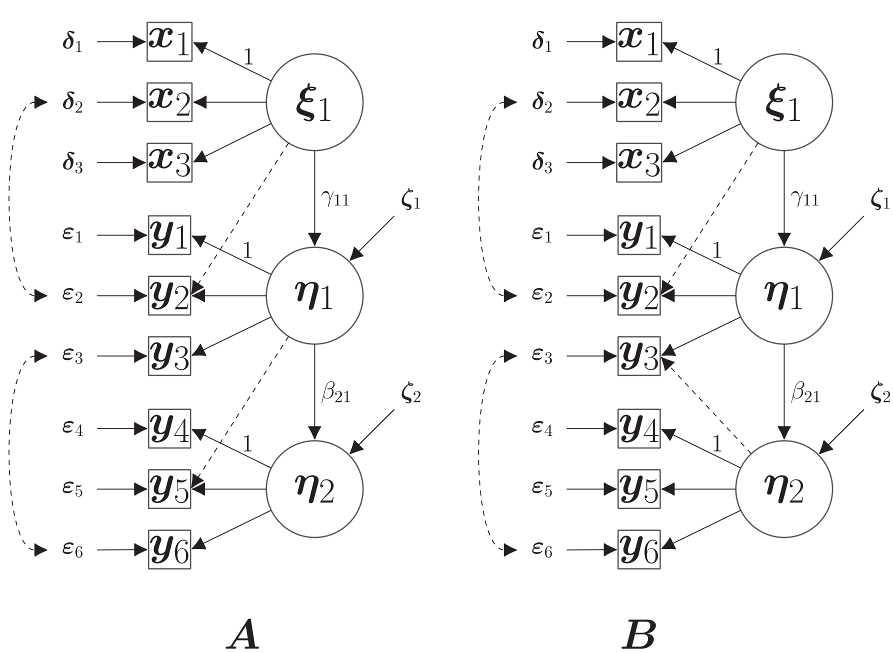

FIGURE 3 Misspecified measurement models. Solid and dashed lines comprise the data-generating model; dashed lines represent those omitted in the misspecification.

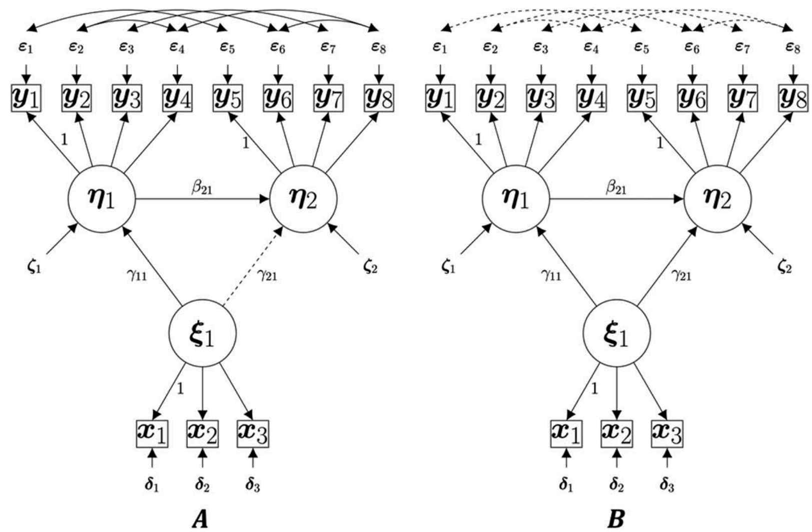

FIGURE 4 Political democracy and industrialization models. Solid and dashed lines comprise the data-generating model; dashed lines represent those omitted in the misspecification.

TABLE 4
Measurement Model Estimates for the Original and Misspecified (i.e., Causal Chain) Models

DV	EV	Original estimate	Original std. error	Causal chain estimate	Causal chain std. error
y_2	η_1	1.139	0.179	1.139	0.179
y_3	η_1	0.969	0.140	0.969	0.140
y_4	η_1	1.210	0.139	1.210	0.139
y_6	η_2	1.051	0.165	1.051	0.165
y_7	η_2	1.180	0.151	1.180	0.151
y_8	η_2	1.203	0.154	1.203	0.154
x_2	ξ_1	2.078	0.128	2.078	0.128
x_3	ξ_1	1.751	0.149	1.751	0.149

TABLE 5
Latent Variable Estimates for the Original and Misspecified Measurement Model Where Correlations Among Indicator Variables are Removed

DV	EV	Original estimate	Original std. error	Misspecified estimate	Misspecified std. error
η_1	ξ_1	1.261	0.426	1.261	0.426
η_2	ξ_1	1.123	0.312	1.123	0.312
η_2	η_1	0.724	0.101	0.724	0.101

Note. Estimates remain identical despite measurement model misspecification.

that the estimate for the relation between ξ_1 and η_2 is significant (see Table 5) and the Sargan test is marginally significant for equation η_2, including this path as is done in the original model appears to be warranted.

Next, we turn to measurement model misspecifications. As depicted in Figure 4b, the correlations among the unique factor of the indicator variables for political democracy are removed. One might expect these errors of the indicators to be correlated from 1960 to 1965; hence, the removal of them is referred to as the "misspecified model" herein. In following the analytic and simulation results, the latent variable model estimates were robust to this misspecification (see Table 5). However, in the measurement model, only the factor loading estimates for the industrialization latent variable are robust-unchanged for the misspecified model, though the change in the other loadings are small and the equation for two variables fail the Sargan test (Table 6).

TABLE 6
Measurement Model Estimates Change in Comparison to the Original Model (Table 4)

DV	EV	Estimate	Std. error	Sargan	df	P(Chi)
y_2	η_1	1.227	0.168	18.863	8	0.016
y_3	η_1	0.986	0.129	10.155	8	0.254
y_4	η_1	1.178	0.131	14.884	8	0.061
y_6	η_2	1.086	0.157	20.569	8	0.008
y_7	η_2	1.132	0.137	13.767	8	0.088
y_8	η_2	1.149	0.144	15.301	8	0.054
x_2	ξ_1	2.078	0.128	8.301	8	0.405
x_3	ξ_1	1.751	0.149	8.738	8	0.365

CONCLUSIONS

There is growing recognition of the importance of understanding structural robustness of estimators. This paper advances our knowledge of robustness of structural

TABLE 7
Summary of New Robustness Conditions for Structural Misspecifications Using the MIIV-2SLS Estimator

The MIIV-2SLS estimator of the coefficients of the measurement model is robust-unchanged to structural misspecifications in the latent variable model.
The MIIV-2SLS estimator of the coefficients of the latent variable model is robust-unchanged to having the incorrect covariance matrices among the unique factors (errors) of the measurement model for the nonscaling indicators.
The MIIV-2SLS estimator of the coefficients of the latent variable model is robust-unchanged to omitted cross-loadings as long as the omitted nonzero cross-loadings when present do not create indirect paths from the latent variable equation error to one or more of the MIIVs from the misspecified model.
The MIIV-2SLS estimator of the coefficients of the latent variable model is not robust-unchanged to omitted cross-loadings if the omitted cross-loadings when present create nonzero indirect paths from the latent variable equation error to one or more of the MIIVs from the misspecified model.
The MIIV-2SLS estimator of the coefficients of the latent variable model is robust-unchanged to the omission of correlated unique factors (errors) for the scaling indicators of latent variables that are included in the true latent variable equation.
The MIIV-2SLS estimator is robust-unchanged to the omission of any observed variable that is not part of the L2O transformed equation and is not among the MIIVs for that equation.

misspecifications of the MIIV-2SLS estimator by deriving a number of new robustness conditions. Table 7 provides a summary of these.

Among other things, we established that structural misspecification in the latent variable model does not affect the MIIV-2SLS estimates of the measurement model. We showed that omitted correlated errors of the nonscaling indicators in the measurement model do not affect the MIIV-2SLS coefficient estimates of the latent variable model. The story of omitted factor loadings is more complicated. As explained in the text and in Table 7, sometimes, the MIIV-2SLS coefficient estimates of the latent variable model are robust-unchanged, sometimes not.

Our new conditions are established for coefficients to be robust-unchanged so that we can accommodate the widely accepted view that models are approximations. We also mentioned the idea of robust-consistent for MIIV-2SLS estimates that remain consistent estimators of the parameters under two different models. Our conditions for robust-unchanged carry over to robust-consistent with the added proviso that the equation of interest is correctly specified and the MIIVs for that equation stay the same with or without the structural misspecifications that are elsewhere. Of course, with real data, the true equation is unknown. However, we have the Sargan (1958) test of whether the MIIVs in an overidentified equation are uncorrelated with the composite error. Failure of the test indicates failure of the model to lead to a consistent estimator of the equation's coefficients. Our robustness conditions suggest where to look for modifying the model.

One implication of our results ties to the one-step versus two-step estimation debate that periodically occurs in the SEM literature (e.g., Anderson & Gerbing, 1988; Hayduk & Glaser, 2000; Mulaik & Millsap, 2000). The debate concerns whether to estimate the measurement model prior to building the latent variable model versus simultaneously estimating both models together. Our results show that if a researcher uses the MIIV-2SLS estimator to tackle the measurement model first, then the results of the measurement model are not influenced by the latent variable model. Hence, a researcher who wants to concentrate on the measurement model first knows that structural misspecifications in the latent variable model do not alter the measurement model coefficient estimates in the MIIV-2SLS context.

A related implication is that some authors have advocated that the measurement model should be fit first and the latent variable second (McDonald, 2010; McDonald & Ho, 2002). They suggest this strategy because of their concern that the overall fit measures from the usual full information estimators like ML might mask the poor fit of the latent variable model by having a good fitting measurement model. Consistent with this claim, O'Boyle and Williams (2011) find that roughly half the studies in organizational research had a good fitting measurement model masking a poorly fitting latent variable model. If we use the MIIV-2SLS estimator, we can test the fit of individual equations from either the latent variable or measurement model and would be less likely to miss a poorly fitting latent variable model due to a good fitting measurement model because each overidentified equation has its own test statistic. As this suggests, it is possible to perform local tests of model fit rather than a single global fit measure. With the local tests of fit for MIIV-2SLS, a researcher can better determine those equations in the model that cause the most problems. And the robustness conditions we propose could prove helpful in locating the sources of bad fit.

Our results also are useful in interpreting the results of simulation studies that look at MIIV-2SLS. Our analytical conditions enable us to predict when estimates will be unchanged and when they will be altered by a structural misspecification. These analytical conditions should work with the polychoric instrumental variable estimator as well (e.g., Bollen & Maydeu-Oliveres, 2007; Jin et al., 2016; Nestler, 2013). This means that we do not need to use simulation studies to predict robustness but can use these analytical results.

Our robustness conditions could be extended in several directions. For instance, further examination of which estimates would change from which equations when dropping variables from an analysis could be done by checking whether the MIIVs for an equation change. It also should be possible to extend these results to the MIIV-GMM (Bollen et al., 2014) when multi-equations are estimated rather than a single equation. We hope that this paper provides a foundation upon which to further our understanding of if, when, and where misspecifications affect SEM estimates.

FUNDING

We gratefully acknowledge support from NIH National Institute of Biomedical Imaging and Bioengineering (Award Number: 1-R01-EB022904-01).

ORCID

Kenneth A. Bollen http://orcid.org/0000-0002-6710-3800

REFERENCES

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*, 411–423. doi:10.1037/0033-2909.103.3.411

Bauldry, S. (2014). miivfind: A command for identifying model-implied instrumental variables for structural equation models in Stata. *Stata Journal*, *14*, 60–75.

Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological Methodology, 17*, 37–69.

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

Bollen, K. A. (1996). An alternative 2SLS estimator for latent variable models. *Psychometrika*, *61*, 109–121. doi:10.1007/BF02296961

Bollen, K. A. (2001). Two-stage least squares and latent variable models: Simultaneous estimation and robustness to misspecifications. In R. Cudeck, S. D. Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and Future, A Festschrift in Honor of Karl Jöreskog* (pp. 119–138). Lincoln, IL: Scientific Software.

Bollen, K. A., & Bauer, D. J. (2004). Automating the selection of model-implied instrumental variables. *Sociological Methods & Research*, *32*(4), 425–452. doi:10.1177/0049124103260341

Bollen, K. A., Kirby, J. B., Curran, P. J., Paxton, P. M., & Chen, F. (2007). Latent variable models under misspecification: Two stage least squares (2SLS) and maximum likelihood (ML) estimators. *Sociological Methods and Research*, *36*, 46–86. doi:10.1177/0049124107301947

Bollen, K. A., Kolenikov, S., & Bauldry, S. (2014). Model-implied instrumental variable—generalized method of moments (MIIV-GMM) estimators for latent variable models. *Psychometrika*, *79*(1), 20–50. doi:10.1007/s11336-013-9335-3

Bollen, K. A., & Maydeu-Oliveres, A. (2007). A polychoric instrumental variable (PIV) estimator for structural equation models with categorical variables. *Psychometrika*, *72*, 309–326. doi:10.1007/s11336-007-9006-3

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*(1), 62–83.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Fisher, Z., Bollen, K. A., Gates, K., & Rönkkö, M. (2017). *MIIVsem: Model Implied Instrumental Variable (MIIV) estimation of structural equation models*. R package version 0.5.0.

Hayduk, L. A., & Glaser, D. N. (2000). Jiving the four-step, waltzing around factor analysis, and other serious fun. *Structural Equation Modeling*, *7*, 1–35. doi:10.1207/S15328007SEM0701_01

Jin, S., Luo, H., & Yang-Wallentin, F. (2016). A simulation study of polychoric instrumental variable estimation in structural equation models. *Structural Equation Modeling*, *23*(5), 680–694. doi:10.1080/10705511.2016.1189334

Jöreskog, K. G., & Sörbom, D. (2001). *LISREL 8: User's reference guide.* Uppsala, Sweden: Scientific Software International.

Kaplan, D. (1989). A study of the sampling variability and z-values of parameter estimates from misspecified structural equation models. *Multivariate Behavioral Research*, *24*, 41–57.

Kaplan, D., & Wenger, R. N. (1993). Asymptotic independence and separability in covariance structure models: Implications for specification error, power, and model modification. *Multivariate Behavioral Research*, *28*, 483–498.

Kirby, J., & Bollen, K. A. (2009). Using instrumental variable (IV) tests to evaluate model specification in latent variable structural equation models. *Sociological Methodology*, *39*, 327–355. doi:10.1111/j.1467-9531.2009.01217.x

Kolenikov, S., & Bollen, K. A. (2012). Testing negative error variances: Is a heywood case a symptom of misspecification? *Sociological Methods and Research*, *41*, 124–167. doi:10.1177/0049124112442138

MacCallum, R. C., Browne, M. W., & Cai, L. (2007). Factor analysis models as approximations. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

McDonald, R. P. (2010). Structural models and the art of approximation. *Perspectives on Psychological Science*, *5*, 675–686. doi:10.1177/1745691610388766

McDonald, R. P., & Ho, M. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*, 64–82. doi:10.1037/1082-989X.7.1.64

Mulaik, S. A. & Millsap, R. E. (2000). Doing the four-step right. *Structural Equation Modeling*, *7*(1), 36–73. doi:10.1207/S15328007SEM0701_02

Nestler, S. (2013). A Monte Carlo study comparing PIV, ULS and DWLS in the estimation of dichotomous confirmatory factor analysis. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 127–143. doi:10.1111/j.2044-8317.2012.02044.x

O'Boyle, E. H., Jr., & Williams, L. J. (2011). Decomposing model fit: measurement vs. Theory in organizational research using latent variables. *Journal of Applied Psychology*, *96*, 1–12. doi:10.1037/a0020539

Sargan, J. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, *26*, 393–415. doi:10.2307/1907619

Yuan, K.-H., Kouros, C. D., & Kelley, K. (2008). Diagnosis for covariance structure models by analyzing the path. *Structural Equation Modeling*, *15*, 564–602.

Yuan, K.-H., Marshall, L. L., & Bentler, P. M. (2003). Assessing the effect of model misspecifications on parameter estimates in structural equation models. *Sociological Methodology*, *33*, 241–265.