PSYCHOMETRIKA
https://doi.org/10.1007/s11336-021-09771-4

A UNIFIED MODEL-IMPLIED INSTRUMENTAL VARIABLE APPROACH FOR STRUCTURAL EQUATION MODELING WITH MIXED VARIABLES

SHAOBO JIN AND FAN YANG- WALLENTIN

UPPSALA UNIVERSITY

KENNETH A. BOLLEN

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

The model-implied instrumental variable (MIIV) estimator is an equation-by-equation estimator of structural equation models that is more robust to structural misspecifications than full information estimators. Previous studies have concentrated on endogenous variables that are all continuous (MIIV-2SLS) or all ordinal . We develop a unified MIIV approach that applies to a mixture of binary, ordinal, censored, or continuous endogenous observed variables. We include estimates of factor loadings, regression coefficients, variances, and covariances along with their asymptotic standard errors. In addition, we create new goodness of fit tests of the model and overidentification tests of single equations. Our simulation study shows that the proposed MIIV approach is more robust to structural misspecifications than diagonally weighted least squares (DWLS) and that both the goodness of fit model tests and the overidentification equations tests can detect structural misspecifications. We also find that the bias in asymptotic standard errors for the MIIV estimators of factor loadings and regression coefficients are often lower than the DWLS ones, though the differences are small in large samples. Our analysis shows that scaling indicators with low reliability can adversely affect the MIIV estimators. Also, using a small subset of MIIVs reduces small sample bias of coefficient estimates, but can lower the power of overidentification tests of equations.

Key words: MIIV, continuous variables, ordinal variables, goodness-of-fit test, overidentification test.

1. Introduction

Structural equation modeling (SEM) is widely used in social and behavior sciences. Endogenous observed variables (e.g., indicators) in a SEM model can be continuous, binary, ordinal, or censored. When all endogenous observed variables are continuous, systemwide maximum likelihood (ML) is the dominant estimator. When binary or ordinal observed endogenous variables are present, researchers frequently use unweighted least squares (ULS, Muthén, 1978) and diagonally weighted least squares (DWLS, Muthén et al., 1997). Typically, these estimators require the estimation of a polychoric correlation matrix as a first stage. In the second stage, they use the polychoric correlation matrix as input and apply systemwide estimators such as ML, ULS, and DWLS. They are systemwide in the sense that all parameter estimates for all equations in the whole structural equation model are estimated simultaneously. Despite their popularity, various studies (e.g., Bollen, 1996; Jin et al., 2016; Nestler, 2013; Yang-Wallentin et al., 2010) have shown that the systemwide estimators can spread the bias due to model misspecifications such as nonzero coefficients (“omitted variables") or covariances of errors mistakenly set to zero. Our focus is on these types of misspecifications.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11336-021-09771-4.

Correspondence should be made to Shaobo Jin, Department of statistics, Uppsala University, Uppsala, Sweden. Email: shaobo.jin@statistik.uu.se

© 2021 The Author(s)
Published online: 07 June 2021

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

To reduce the impacts of such model misspecifications, Bollen (1996) proposed a two-stage least squares (2SLS) approach using model-implied instrumental variables (MIIVs) when all endogenous observed variables are continuous. Bollen and Maydeu-Olivares (2007) extend this approach to models with observed ordinal endogenous variables and called it the polychoric instrumental variable (PIV) estimator. The MIIVs for the 2SLS and PIV estimators are easily found from the pool of variables in the SEM model by the algorithm described in Bollen (1996, page 114) and implemented in SAS by Bollen and Bauer (2004), in Stata by Bauldry (2014), and in R by Fisher et al. (2017). The instrumental variables from this approach are model-implied instrumental variables; in that according to the structure of the model, these MIIVs should be uncorrelated with the error of the equation to be estimated. 2SLS with MIIVs is denoted by MIIV-2SLS to distinguish it from the 2SLS with auxiliary, external instrumental variables that are not part of the original model. In practice, auxiliary instruments (Bollen, 2012) are more common than MIIVs so it is important to distinguish between them. Both MIIV-2SLS and PIV approaches estimate the parameters in an equation-by-equation manner, in which each indicator equation with its factor loadings and each latent variable equation with its regression coefficients is estimated separately from each other. Furthermore, the estimator is non-interative (Bollen, 1996), which eliminates the non-convergence issue that can occur with the systemwide estimators. The simulation study of Bollen et al. (2007) showed that the MIIV-2SLS estimator has accuracy comparable to the traditional ML estimator if the model is correctly specified and is often more robust if the model is misspecified. Similarly, Jin et al. (2016) and Nestler (2013) showed that the PIV estimator is as accurate as the ULS and DWLS estimators in the correctly specified model and is more accurate in the structurally misspecified model. The reader is directed to Bollen et al. (2018a) and Bollen (2020) for conditions under which the MIIV-2SLS and PIV approaches remain robust to structural misspecifications. Recently, the principle of MIIV-2SLS and PIV has been generalized by Bollen et al. (2014) to SEM with generalized method of moments estimators, by Fisher et al. (2019) to dynamic time-series models, by Nestler (2014) to handle equality constraints, by Nestler (2015a) to nonlinear SEM models, and by Nestler (2015b) to growth curve models, just to name a few.

Another advantage of the MIIV-2SLS and PIV method is the local, equation-specific test of model specification. The most common method to improve model fit and search for model misspecification with systemwide estimators is to sequentially apply the modification index (Sörbom, 1989). However, the modification index can lead to severe errors (MacCallum, 1986; MacCallum et al., 1992). Bollen (1996, pages 117-118) and Kirby and Bollen (2009) recommend the Sargan (Sargan, 1958) test for equation-by-equation overidentification tests of misspecification. Jin and Cao (2018) showed that these tests are not suitable for ordinal endogenous variables and proposed alternative overidentification tests for such variables.

The cited studies focused on situations where all endogenous variables were either continuous (MIIV-2SLS) or all were ordinal (PIV). To our knowledge, the only exception is a parallel study by Fisher and Bollen (2020), which also considered inference using MIIV with different types of observed variables. The main purpose of our study is to unify the MIIV-2SLS and PIV approaches and propose an approach for a SEM model with different variable types, such as continuous, ordinal, binary, or a mixture of different types. Hereafter, the unified approach is termed the MIIV approach. We will present the MIIV point estimators, which is in line with the estimator in Fisher and Bollen (2020). Two sets of standard errors are proposed. One set of standard errors matches those in Bollen (1996) for observed continuous variables in finite sample and is asymptotically equivalent to those in Bollen and Maydeu-Olivares (2007) for observed ordinal variables. The other is asymptotically equivalent to Bollen (1996) but equivalent to Bollen and Maydeu-Olivares (2007) in finite samples. Fisher and Bollen (2020) only focused on point estimation and standard error estimation. In contrast, we will also develop goodness-of-fit tests for the whole model and overidentification tests for individual equations. Similarly to the standard error estimators, one set

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

of overidentification tests for equations matches Kirby and Bollen (2009) for observed continuous variables and the other set matches Jin and Cao (2018) for observed ordinal variables.

The rest of the paper is organized as follows. We first briefly present the SEM model and the MIIV idea. Then, we develop the estimation of SEMs with the unified MIIV approach. Next, we provide model goodness-of-fit tests and overidentification tests of equations. We conduct a Monte Carlo simulation to investigate the small sample properties of the proposed approach. An empirical example illustrates our approach. A discussion and conclusion section ends the paper.

2. Structural Equation Modeling and Instrumental Variable

2.1. SEM Model

Consider the SEM model

$$\boldsymbol{y}^* = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{1}$$

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\zeta}, \tag{2}$$

where $\boldsymbol{y}^*$ is the vector of continuous variables, $\boldsymbol{\Lambda}$ is the matrix of factor loadings, $\boldsymbol{\eta}$ is the vector of latent variables, $\boldsymbol{B}$ is the matrix of latent regression coefficient, and $\boldsymbol{\varepsilon}$ and $\boldsymbol{\zeta}$ are disturbances. Without loss of generality, we assume that $\boldsymbol{y}^*$ has a zero mean. We can justify the zero mean assumption by forming mean deviation variables for the observed continuous variables and for observed binary and ordinal variables researchers commonly assume that their underlying variables have means of zero. We also assume that $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ are independent and that $\boldsymbol{\zeta}$ and $\boldsymbol{\varepsilon}$ are independent. The $\boldsymbol{\eta}$ vector consists of endogenous and exogenous latent variables with the exogenous variables independent of the errors in $\boldsymbol{\zeta}$ that correspond to the endogenous variables. Furthermore, the disturbances $\boldsymbol{\varepsilon}$ and $\boldsymbol{\zeta}$ are homoscedastic, where $\text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Theta}$ and $\text{var}(\boldsymbol{\zeta}) = \boldsymbol{\Psi}$.

The observed vector is denoted by $\boldsymbol{y}$, which is not necessarily the same as $\boldsymbol{y}^*$. If y_j, the jth entry in $\boldsymbol{y}$, is continuous, then $y_j = y_j^*$, where y_j^* is the jth entry in $\boldsymbol{y}^*$. If y_j is binary or ordinal, then it is obtained by categorizing the underlying continuous variable y_j^* according to the threshold values. If y_j is censored, then it is obtained by censoring y_j^* at a boundary. If $\boldsymbol{y}^*$ only contains observed continuous variables, we only need it to have finite moments up to the fourth order. If $\boldsymbol{y}^*$ contains any observed binary, ordinal, or censored variables, we assume $\boldsymbol{\eta}$, $\boldsymbol{\varepsilon}$, and $\boldsymbol{\zeta}$ follow a multivariate normal distribution.

Let $\boldsymbol{\Sigma}$ be the population covariance matrix of $\boldsymbol{y}^*$. If some entries in $\boldsymbol{y}^*$ are discrete, the corresponding variances are fixed to 1 for identification. Estimation of $\boldsymbol{S}$, the sample counterpart of $\boldsymbol{\Sigma}$, depends on the types of indicators. For example, using Olsson (1979), we can estimate the polychoric correlation for two ordinal variables, and using Olsson et al. (1982), we can estimate the polyserial correlation for one ordinal indicator and one continuous indicator. Throughout the paper, we assume that $\boldsymbol{S}$ is a consistent estimator of $\boldsymbol{\Sigma}$ and that

$$\sqrt{n}\left(\boldsymbol{s} - \boldsymbol{\sigma}\right) \xrightarrow{d} N\left(\boldsymbol{0}, \boldsymbol{\Upsilon}\right),$$

where n is the sample size, $\boldsymbol{s}$ is the vector of nonredundant free entries in $\boldsymbol{S}$, $\boldsymbol{\sigma}$ is the vector of corresponding free entries in $\boldsymbol{\Sigma}$, and $\boldsymbol{\Upsilon}$ is the asymptotic covariance matrix.

Our SEM model implies that $\boldsymbol{\Sigma}$ equals the implied covariance matrix of

$$\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right) = \boldsymbol{\Lambda}\left(\boldsymbol{I} - \boldsymbol{B}\right)^{-1}\boldsymbol{\Psi}\left(\boldsymbol{I} - \boldsymbol{B}\right)^{-\mathrm{T}}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Theta}, \tag{3}$$

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

where $\boldsymbol{\theta}$ is the vector of all model parameters (i.e., free parameters in $\boldsymbol{\Lambda}$, $\boldsymbol{B}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Theta}$) and $(\cdot)^{-\mathrm{T}}$ takes the transpose of the inverse of the enclosed matrix. The diagonal entries of $\boldsymbol{\Theta}$ that correspond to binary and ordinal indicators are not free parameters, but are restricted so that the corresponding entries in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ are 1 for identification. The traditional systemwide approaches estimate $\boldsymbol{\theta}$ by minimizing the fit function

$$T(\boldsymbol{\theta}) = (\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta}))^T \hat{\boldsymbol{W}} (\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})), \tag{4}$$

where $\boldsymbol{\sigma}(\boldsymbol{\theta})$ is the vector of free entries in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and the weight matrix $\hat{\boldsymbol{W}}$ is a consistent estimator of $\boldsymbol{W}$. Different weight matrices yield different estimators, e.g., $\boldsymbol{W} = \boldsymbol{I}$ in unweighted least squares (ULS; Muthén, 1978), $\boldsymbol{W}$ is the inverse of diagonal elements of $\boldsymbol{\Upsilon}$ in diagonally weighted least squares (DWLS; Muthén et al., 1997), and $\boldsymbol{W} = \boldsymbol{\Upsilon}^{-1}$ in weighted least squares (WLS; Browne, 1984).

2.2. *Latent to Observed Variable (L2O) Transformation*

Each latent variable must be given a scale or metric. To set the scale of $\boldsymbol{\eta}$, the most common approach is to choose one indicator per latent variable and to set its factor loading to one so that it becomes the scaling indicator. When possible, we should choose the scaling indicator to have factor complexity of one and a high R^2. Suppose that we partition $\boldsymbol{y}^*$ into $(\boldsymbol{y}_1^{*T}, \boldsymbol{y}_2^{*T})^T$ and $\boldsymbol{\varepsilon}$ into $(\boldsymbol{\varepsilon}_1^T, \boldsymbol{\varepsilon}_2^T)^T$ such that $\boldsymbol{y}_1^* = \boldsymbol{\eta} + \boldsymbol{\varepsilon}_1$ is for the scaling indicators and $\boldsymbol{y}_2^* = \boldsymbol{\Lambda}_2\boldsymbol{\eta} + \boldsymbol{\varepsilon}_2$ contains unknown factor loadings for the nonscaling indicators. Following Bollen and Maydeu-Olivares (2007), the SEM model is equivalent to

$$\begin{pmatrix} \boldsymbol{y}_2^* \\ \boldsymbol{y}_1^* \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Lambda}_2 \\ \boldsymbol{B} \end{pmatrix} \boldsymbol{y}_1^* + \begin{pmatrix} \boldsymbol{\varepsilon}_2 - \boldsymbol{\Lambda}_2\boldsymbol{\varepsilon}_1 \\ \boldsymbol{\zeta} + (\boldsymbol{I} - \boldsymbol{B})\boldsymbol{\varepsilon}_1 \end{pmatrix}, \tag{5}$$

which is simply a multivariate regression system with "observed" variables. This is referred to as the Latent-to-Observed (L2O) variable transformation (Bollen, 2019), because the original equation system with latent variables is transformed to one without latent variables. When we have noncontinuous endogenous observed variables, the L2O transformation results in underlying scaling variables rather than observed ones. Regardless, in most cases, this leads the regressors to be correlated with one or more parts of the composite error terms. We can use MIIVs to consistently estimate the regression coefficients in $\boldsymbol{B}$ and $\boldsymbol{\Lambda}_2$.

Equation (5) indicates that we can partition $\boldsymbol{\theta}$ into two vectors: $\boldsymbol{\theta}_1$ (free parameters in $\boldsymbol{\Lambda}_2$ and $\boldsymbol{B}$) and $\boldsymbol{\theta}_2$ (the nonredundant free parameters in $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$). To estimate $\boldsymbol{\theta}_1$, Bollen (1996) and Bollen and Maydeu-Olivares (2007) proposed to consider the equation system (5) in a row-by-row manner. Suppose that the jth row of the system (5) is

$$y_j^* = \boldsymbol{z}_j^{*T}\boldsymbol{\theta}_1^{(j)} + e_j, \tag{6}$$

where $\boldsymbol{z}_j^*$ is the right-hand side explanatory variable vector, and $\boldsymbol{\theta}_1^{(j)}$ is the vector of parameters. We assume that there exists a non-empty subset of $\boldsymbol{z}_j^*$ that is correlated with the error e_j. As the names implies, the MIIVs, denoted by $\boldsymbol{v}_j^*$, are selected from the variables $\boldsymbol{y}^*$, excluding y_j^* and the elements in $\boldsymbol{z}_j^*$ that have nonzero correlations with e_j. The valid MIIVs $\boldsymbol{v}_j^*$ must be correlated with the endogenous variable $\boldsymbol{z}_j^*$ as well as uncorrelated with the error term e_j, and the number of $\boldsymbol{v}_j^*$ must be no lower than the number of $\boldsymbol{z}_j^*$. We can find the pool of all valid MIIVs for a given equation by using the algorithms proposed by Bollen (1996). Due to space limitation, we direct

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

the readers to Bollen (2019) for examples of Eq. (6) and the algorithm of selecting valid IVs. It is worth mentioning that the choice of the scaling indicator will affect the choice of MIIVs and the estimates. Though the estimator remains consistent, its asymptotic variance could differ. As mentioned earlier, it is preferable to choose as scaling indicators those indicators that are thought to be most closely related to the latent variable that they measure.

Bollen (1996; 2001), Bollen et al. (2007), and Kirby and Bollen (2009) have investigated the case where $\boldsymbol{y}$ is continuous. In contrast, Bollen and Maydeu-Olivares (2007), Jin et al. (2016), Jin and Cao (2018), and Nestler (2013) have investigated the case where $\boldsymbol{y}$ is ordinal. Fisher and Bollen (2020) is the only study we know of that considered indicators of different types.

3. Estimation by the MIIV Approach

In this section, we propose a unified approach for estimating and testing models with continuous, binary, or ordinal endogenous observed variables. The approach applies as long as the sample moment matrix is a consistent estimator of the population moment matrix, with an estimate of the asymptotic covariance matrix of the elements of such matrices.

3.1. Point Estimator of θ_1

If Eq. (6) is correctly specified and the MIIVs are valid, then $\boldsymbol{\theta}_1^{(j)} = \boldsymbol{\gamma}_j(\boldsymbol{\sigma})$, where

$$\boldsymbol{\gamma}_j(\boldsymbol{\sigma}) = \left(\boldsymbol{\Sigma}_{vz,j}^T \boldsymbol{\Sigma}_{vv,j}^{-1} \boldsymbol{\Sigma}_{vz,j}\right)^{-1} \boldsymbol{\Sigma}_{vz,j}^T \boldsymbol{\Sigma}_{vv,j}^{-1} \boldsymbol{\Sigma}_{vy,j},$$

with $\boldsymbol{\Sigma}_{vz,j}$ being the covariance matrix between $\boldsymbol{v}_j^*$ and $\boldsymbol{z}_j^*$, and $\boldsymbol{\Sigma}_{vv,j}$ being the covariance matrix of $\boldsymbol{v}_j^*$, and $\boldsymbol{\Sigma}_{vy,j}$ being the covariance matrix between $\boldsymbol{v}_j^*$ and y_j^*. The MIIV estimator is

$$\hat{\boldsymbol{\theta}}_1^{(j)} = \boldsymbol{\gamma}_j(\boldsymbol{s}) = \left(\boldsymbol{S}_{vz,j}^T \boldsymbol{S}_{vv,j}^{-1} \boldsymbol{S}_{vz,j}\right)^{-1} \boldsymbol{S}_{vz,j}^T \boldsymbol{S}_{vv,j}^{-1} \boldsymbol{S}_{vy,j}, \tag{7}$$

where $\boldsymbol{S}_{\cdot,j}$ is the sample counterpart of $\boldsymbol{\Sigma}_{\cdot,j}$. The same estimator was proposed by Bollen and Maydeu-Olivares (2007) and Fisher and Bollen (2020). If all observed variables are continuous, the MIIV estimator (7) is equivalent to the estimator in Bollen (1996)

$$\hat{\boldsymbol{\theta}}_1^{(j)} = \left[\boldsymbol{Z}_j^T \boldsymbol{V}_j \left(\boldsymbol{V}_j^T \boldsymbol{V}_j\right)^{-1} \boldsymbol{V}_j^T \boldsymbol{Z}_j\right]^{-1} \boldsymbol{Z}_j^T \boldsymbol{V}_j \left(\boldsymbol{V}_j^T \boldsymbol{V}_j\right)^{-1} \boldsymbol{V}_j^T \boldsymbol{y}_j,$$

where $\boldsymbol{y}_j$, $\boldsymbol{Z}_j$, and $\boldsymbol{V}_j$ are the demeaned data matrices of y_j^*, $\boldsymbol{z}_j^*$, and $\boldsymbol{v}_j^*$, respectively. Similar to MIIV-2SLS and PIV, the MIIV estimator is a consistent estimator of $\boldsymbol{\theta}_1^{(j)}$, as long as Eq. (6) is correctly specified, the inverses in Eq. (7) exist, and the MIIVs are valid. This implies that we can maintain consistency of $\hat{\boldsymbol{\theta}}_1^{(j)}$ in a misspecified SEM model under certain conditions (e.g., Bollen, 2001; Bollen et al., 2018b; Bollen, 2020). In contrast, the systemwide estimators sometimes spread the bias from a poorly specified equation in one part of the model to a correctly specified equation in another part.

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

3.2. Standard Error of $\hat{\boldsymbol{\theta}}_1$

If we collect $\hat{\boldsymbol{\theta}}_1^{(j)}$ for all j, the MIIV estimator of $\boldsymbol{\theta}_1$ is $\hat{\boldsymbol{\theta}}_1 = \boldsymbol{\gamma}(\boldsymbol{s})$, where $\boldsymbol{\gamma}(\boldsymbol{\sigma})$ is the vector that stacks all $\boldsymbol{\gamma}_j(\boldsymbol{\sigma})$ on top of the other. The delta method indicates that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^*\right) = \boldsymbol{K}(\boldsymbol{\sigma})\sqrt{n}(\boldsymbol{s} - \boldsymbol{\sigma}) + o_{\mathrm{P}}(1), \tag{8}$$

where $\boldsymbol{\theta}_1^*$ is the asymptotic limit of $\hat{\boldsymbol{\theta}}_1$ and $\boldsymbol{K}(\boldsymbol{\sigma}) = \partial\boldsymbol{\gamma}(\boldsymbol{\sigma})/\partial\boldsymbol{\sigma}^T$. Then, we obtain

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_1^{(j)} - \boldsymbol{\theta}_1^{(j)*}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{K}_j(\boldsymbol{\sigma})\boldsymbol{\Upsilon}\boldsymbol{K}_j^T(\boldsymbol{\sigma})\right), \tag{9}$$

where $\boldsymbol{\theta}_1^{(j)*}$ is the asymptotic limit of $\hat{\boldsymbol{\theta}}_1^{(j)}$ and $\boldsymbol{K}_j(\boldsymbol{\sigma}) = \partial\boldsymbol{\gamma}_j(\boldsymbol{\sigma})/\partial\boldsymbol{\sigma}^T$. The asymptotic distribution (9) holds even when the MIIVs are invalid and $\boldsymbol{\theta}_1^{(j)*}$ is not the same as $\boldsymbol{\theta}_1^{(j)}$. To compute $\boldsymbol{K}_j(\boldsymbol{\sigma})$, the chain rule yields

$$\begin{aligned}\frac{\partial\boldsymbol{\gamma}_j(\boldsymbol{\sigma})}{\partial\sigma_i} &= \left(\boldsymbol{\Sigma}_{vz,j}^T\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{\Sigma}_{vz,j}\right)^{-1}\frac{\partial\boldsymbol{\Sigma}_{vz,j}^T\boldsymbol{\Sigma}_{vv,j}^{-1}}{\partial\sigma_i}\left[\boldsymbol{\Sigma}_{vy,j} - \boldsymbol{\Sigma}_{vz,j}\boldsymbol{\gamma}_j(\boldsymbol{\sigma})\right] \\ &\quad + \left(\boldsymbol{\Sigma}_{vz,j}^T\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{\Sigma}_{vz,j}\right)^{-1}\boldsymbol{\Sigma}_{vz,j}^T\boldsymbol{\Sigma}_{vv,j}^{-1}\left[\frac{\partial\boldsymbol{\Sigma}_{vy,j}}{\partial\sigma_i} - \frac{\partial\boldsymbol{\Sigma}_{vz,j}}{\partial\sigma_i}\boldsymbol{\gamma}_j(\boldsymbol{\sigma})\right],\end{aligned} \tag{10}$$

where σ_i is the ith entry in $\boldsymbol{\sigma}$.

To obtain the estimator of the standard error, $\boldsymbol{K}_j(\boldsymbol{\sigma})$ is estimated by $\boldsymbol{K}_j(\boldsymbol{s})$ and estimation of $\boldsymbol{\Upsilon}$ depends on the distributional assumption. In general, we can estimate $\boldsymbol{\Upsilon}$ using the approach in Muthén (1984). If all observed variables are ordinal, we can estimate $\boldsymbol{\Upsilon}$ from Jöreskog (1994). If all observed variables are continuous, we estimate $\boldsymbol{\Upsilon}$ using the fourth-order moments as in the asymptotic distribution-free approach (Browne, 1984). If all observed variables are normally distributed, we can estimate $\boldsymbol{\Upsilon}$ by $2\boldsymbol{D}^+(\boldsymbol{S}\otimes\boldsymbol{S})\boldsymbol{D}^{+T}$, where $\boldsymbol{D}$ is a duplication matrix and $\boldsymbol{D}^+ = \left(\boldsymbol{D}^T\boldsymbol{D}\right)^{-1}\boldsymbol{D}^T$ (Browne, 1987).

3.3. Equivalence of Standard Errors of $\hat{\boldsymbol{\theta}}_1$

The asymptotic distribution attained by Bollen and Maydeu-Olivares (2007) when all observed variables are ordinal is the same as the asymptotic distribution (9). The same distribution is also used by Fisher and Bollen (2020). As mentioned above, this asymptotic distribution is applicable regardless of the validity of the MIIVs. If Eq. (6) is correctly specified and all MIIVs are valid, $\boldsymbol{\Sigma}_{vy,j} = \boldsymbol{\Sigma}_{vz,j}\boldsymbol{\theta}_1^{(j)}$ and $\boldsymbol{\gamma}_j(\boldsymbol{\sigma}) = \boldsymbol{\theta}_1^{(j)}$, then the first term in equation (10) vanishes and $\boldsymbol{K}_j(\boldsymbol{\sigma})\boldsymbol{\Upsilon}\boldsymbol{K}_j^T(\boldsymbol{\sigma})$ reduces to

$$\left(\boldsymbol{\Sigma}_{vz,j}^T\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{\Sigma}_{vz,j}\right)^{-1}\boldsymbol{\Sigma}_{vz,j}^T\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{\Omega}_j\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{\Sigma}_{vz,j}\left(\boldsymbol{\Sigma}_{vz,j}^T\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{\Sigma}_{vz,j}\right)^{-1}, \tag{11}$$

where $\boldsymbol{g}_j(\boldsymbol{\sigma}) = \boldsymbol{\Sigma}_{vy,j} - \boldsymbol{\Sigma}_{vz,j}\boldsymbol{\theta}_1^{(j)}$ and

$$\boldsymbol{\Omega}_j = \left(\frac{\partial\boldsymbol{g}_j(\boldsymbol{\sigma})}{\partial\boldsymbol{\sigma}^T}\right)\boldsymbol{\Upsilon}\left(\frac{\partial\boldsymbol{g}_j(\boldsymbol{\sigma})}{\partial\boldsymbol{\sigma}^T}\right)^T.$$

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

The covariance matrix (11) can be estimated by

$$\left(\boldsymbol{S}_{vz,j}^{T}\boldsymbol{S}_{vv,j}^{-1}\boldsymbol{S}_{vz,j}\right)^{-1}\boldsymbol{S}_{vz,j}^{T}\boldsymbol{S}_{vv,j}^{-1}\hat{\boldsymbol{\Omega}}_{j}\boldsymbol{S}_{vv,j}^{-1}\boldsymbol{S}_{vz,j}\left(\boldsymbol{S}_{vz,j}^{T}\boldsymbol{S}_{vv,j}^{-1}\boldsymbol{S}_{vz,j}\right)^{-1}, \tag{12}$$

where $\hat{\boldsymbol{\Omega}}_j$ is the estimator of $\boldsymbol{\Omega}$ with $\boldsymbol{\Sigma}$ evaluated at $\boldsymbol{S}$ and $\boldsymbol{\Upsilon}$ evaluated at some estimator $\hat{\boldsymbol{\Upsilon}}$.

If all observed variables are continuous and all MIIVs are valid, Bollen (1996) showed that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{1}^{(j)}-\boldsymbol{\theta}_{1}^{(j)}\right)\xrightarrow{d}N\left(\mathbf{0},\varphi_{j}^{2}\boldsymbol{\Sigma}_{\hat{z}\hat{z},j}^{-1}\right), \tag{13}$$

where $\boldsymbol{\Sigma}_{\hat{z}\hat{z},j}=\boldsymbol{\Sigma}_{vz,j}^{T}\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{\Sigma}_{vz,j}$ is the asymptotic covariance matrix of the predicted right-hand side endogenous variable of Eq. (6) and $\varphi_j^2=\boldsymbol{\Sigma}_{yy,j}-2\boldsymbol{\theta}^{(j)T}\boldsymbol{\Sigma}_{zy,j}+\boldsymbol{\theta}^{(j)T}\boldsymbol{\Sigma}_{zz,j}\boldsymbol{\theta}^{(j)}$ is the variance of the error term e_j. We expect the asymptotic covariance matrix $\varphi_j^2\boldsymbol{\Sigma}_{\hat{z}\hat{z},j}^{-1}$ to be equivalent to (11) when all observed variables are continuous and all MIIVs are valid, since both are asymptotic covariance matrices of $\hat{\boldsymbol{\theta}}_1^{(j)}$. It is worth emphasizing that we need all MIIVs to be valid in order for (11) and (13) to be valid, but neither the observed continuous variables nor the disturbances need to be normally distributed (Greene, 2002, page 77). We estimate the asymptotic covariance matrix $\varphi_j^2\boldsymbol{\Sigma}_{\hat{z}\hat{z},j}^{-1}$ by $\hat{\varphi}_j^2\left(\boldsymbol{S}_{vz,j}^{T}\boldsymbol{S}_{vv,j}^{-1}\boldsymbol{S}_{vz,j}\right)^{-1}$, where $\hat{\varphi}_j^2=\boldsymbol{S}_{yy,j}-2\hat{\boldsymbol{\theta}}_1^{(j)T}\boldsymbol{S}_{zy,j}+\hat{\boldsymbol{\theta}}_1^{(j)T}\boldsymbol{S}_{zz,j}\hat{\boldsymbol{\theta}}_1^{(j)}$.

Despite the equivalence of $\varphi_j^2\boldsymbol{\Sigma}_{\hat{z}\hat{z},j}^{-1}$ and equation (11), their finite sample estimators may still differ, as revealed in the following proposition. For ease of presentation, all proofs are placed in the appendix.

Proposition 1. *Suppose that all observed variables are continuous and* $\boldsymbol{\Upsilon}$ *is estimated by* $2\boldsymbol{D}^{+}\left(\boldsymbol{S}\otimes\boldsymbol{S}\right)\boldsymbol{D}^{+T}$. *Then,*

$$\hat{\boldsymbol{\Omega}}_{j}=\hat{\varphi}_{j}^{2}\boldsymbol{S}_{vv,j}+\left(\boldsymbol{S}_{vy,j}-\boldsymbol{S}_{vz,j}\hat{\boldsymbol{\theta}}_{1}^{(j)}\right)\left(\boldsymbol{S}_{vy,j}-\boldsymbol{S}_{vz,j}\hat{\boldsymbol{\theta}}_{1}^{(j)}\right)^{T},$$

and the estimator (12) is the same as $\hat{\varphi}_j^2\left(\boldsymbol{S}_{vz,j}^{T}\boldsymbol{S}_{vv,j}^{-1}\boldsymbol{S}_{vz,j}\right)^{-1}$.

Proposition 1 shows that the estimator $2\boldsymbol{D}^{+}\left(\boldsymbol{S}\otimes\boldsymbol{S}\right)\boldsymbol{D}^{+T}$ equates the estimator (12) and the standard error in Bollen (1996) in finite samples. Such an estimator is a consistent estimator of $\boldsymbol{\Upsilon}$ under the normality assumption. In contrast, if the normality assumption is relaxed and $\boldsymbol{\Upsilon}$ is estimated from the sample fourth-order moments, the estimator (12) is not necessarily equal to $\hat{\varphi}_j^2\left(\boldsymbol{S}_{vz,j}^{T}\boldsymbol{S}_{vv,j}^{-1}\boldsymbol{S}_{vz,j}\right)^{-1}$. For Proposition 1 to hold, we do not even need the MIIVs to be valid. Regardless of the validity of MIIVs, we always have $\boldsymbol{S}_{vz,j}^{T}\boldsymbol{S}_{vv,j}^{-1}\left(\boldsymbol{S}_{vy,j}-\boldsymbol{S}_{vz,j}\hat{\boldsymbol{\theta}}_{1}^{(j)}\right)=\mathbf{0}$, which makes the estimator (12) the same as $\hat{\varphi}_j^2\left(\boldsymbol{S}_{vz,j}^{T}\boldsymbol{S}_{vv,j}^{-1}\boldsymbol{S}_{vz,j}\right)^{-1}$. As revealed in the following proposition, if we relax the assumption that $\boldsymbol{\Upsilon}$ is estimated by $2\boldsymbol{D}^{+}\left(\boldsymbol{S}\otimes\boldsymbol{S}\right)\boldsymbol{D}^{+T}$ but assume that all MIIVs are valid, the finite sample equivalence still holds.

Proposition 2. *Suppose that all observed variables are continuous with finite fourth-order moments, Eq. (6) is correctly specified, and all MIIVs are valid. Then,* $\boldsymbol{\Omega}_j=\varphi_j^2\boldsymbol{\Sigma}_{vv,j}$.

Proposition 2 implies that we can estimate $\boldsymbol{\Omega}_j$ by $\hat{\varphi}_j^2\boldsymbol{S}_{vv,j}$ to attain the finite sample equivalence of (12) and $\hat{\varphi}_j^2\left(\boldsymbol{S}_{vz,j}^{T}\boldsymbol{S}_{vv,j}^{-1}\boldsymbol{S}_{vz,j}\right)^{-1}$. It is also worth mentioning that Proposition 2 does not require the observed continuous variables to be normally distributed.

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

3.4. Point Estimator of θ_2

To estimate $\boldsymbol{\theta}_2$ given $\hat{\boldsymbol{\theta}}_1$, we minimize the least-squares fit function $T\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)$, where $T\left(\boldsymbol{\theta}\right)$ is given by (4). The weight matrices commonly used for systemwide estimators can naturally be used here. Bollen and Maydeu-Olivares (2007) and Fisher and Bollen (2020) used ULS. It is worth mentioning that the estimator of $\boldsymbol{\theta}_2$ is not robust against model misspecification, since we use the systemwide estimation here [see Bollen and Maydeu-Olivares (2007) page 321 and Nestler (2015b)]. It is also worth mentioning that non-convergence is not an issue when estimating $\boldsymbol{\theta}_1$, since it is based on the closed-form estimator (7). However, when $T\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)$ is minimized to estimate $\boldsymbol{\theta}_2$, non-convergence becomes a potential issue for the MIIV approach as well. However, we can always estimate $\hat{\boldsymbol{\theta}}_1$ even if non-convergence prevents us from estimating $\boldsymbol{\theta}_2$

3.5. Standard Error of $\hat{\theta}_2$

Bollen and Maydeu-Olivares (2007) derived the asymptotic distribution of the PIV estimator $\hat{\boldsymbol{\theta}}_2$ under the ULS fit function. Their equation (31) was taken by Fisher and Bollen (2020) in equation (33). Under certain conditions, it is shown in the appendix that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2\right) = \boldsymbol{C}\left(\boldsymbol{\theta}, \boldsymbol{\sigma}\right)\sqrt{n}\left(\boldsymbol{s} - \boldsymbol{\sigma}\right) + o_{\mathrm{P}}\left(1\right), \tag{14}$$

where $\boldsymbol{C}\left(\boldsymbol{\theta}, \boldsymbol{\sigma}\right) = \boldsymbol{H}\left(\boldsymbol{\theta}\right)\left[\boldsymbol{I} - \boldsymbol{J}_1\left(\boldsymbol{\theta}\right)\boldsymbol{K}\left(\boldsymbol{\sigma}\right)\right]$ with $\boldsymbol{H}\left(\boldsymbol{\theta}\right) = \left[\boldsymbol{J}_2^T\left(\boldsymbol{\theta}\right)\boldsymbol{W}\boldsymbol{J}_2\left(\boldsymbol{\theta}\right)\right]^{-1}\boldsymbol{J}_2^T\left(\boldsymbol{\theta}\right)\boldsymbol{W}$, $\boldsymbol{J}_1\left(\boldsymbol{\theta}\right) = \partial\boldsymbol{\sigma}\left(\boldsymbol{\theta}\right)/\partial\boldsymbol{\theta}_1^T$ and $\boldsymbol{J}_2\left(\boldsymbol{\theta}\right) = \partial\boldsymbol{\sigma}\left(\boldsymbol{\theta}\right)/\partial\boldsymbol{\theta}_2^T$. Expansion of (14) implies

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2\right) \xrightarrow{d} N\left(\boldsymbol{0},\ \boldsymbol{C}\left(\boldsymbol{\theta}, \boldsymbol{\sigma}\right)\boldsymbol{\Upsilon}\boldsymbol{C}^T\left(\boldsymbol{\theta}, \boldsymbol{\sigma}\right)\right). \tag{15}$$

We estimate the standard errors from distribution (15) by evaluating $\boldsymbol{\theta}$, $\boldsymbol{\sigma}$, and $\boldsymbol{\Upsilon}$ at $\hat{\boldsymbol{\theta}}$, $\boldsymbol{s}$, and $\hat{\boldsymbol{\Upsilon}}$, respectively. Our $\boldsymbol{H}$ matrix is more general than the $\boldsymbol{H}$ matrix in equation (33) of Fisher and Bollen (2020) because we allow the use of a more general weight matrix $\boldsymbol{W}$ to develop the standard errors whereas they assume $\boldsymbol{W} = \boldsymbol{I}$ as is true for ULS. [1]

4. MIIV Tests

4.1. Goodness-of-fit Tests for Model

Bollen and Maydeu-Olivares (2007) propose goodness-of-fit tests for the model as a whole for PIV. We develop similar model test statistics for the MIIV approach here. We show in the appendix that, if the model is correctly specified, $nT\left(\boldsymbol{\theta}\right)$ evaluated at the MIIV estimates converges in distribution to a weighted sum of independent Chi-square random variables with 1 degree of freedom, where the weights are the eigenvalues of

$$\boldsymbol{M} = \boldsymbol{\Upsilon}^{1/2}\left[\boldsymbol{I} - \boldsymbol{J}\left(\boldsymbol{\theta}\right)\begin{pmatrix}\boldsymbol{K}\left(\boldsymbol{\sigma}\right)\\ \boldsymbol{C}\left(\boldsymbol{\theta}, \boldsymbol{\sigma}\right)\end{pmatrix}\right]^T\boldsymbol{W}\left[\boldsymbol{I} - \boldsymbol{J}\left(\boldsymbol{\theta}\right)\begin{pmatrix}\boldsymbol{K}\left(\boldsymbol{\sigma}\right)\\ \boldsymbol{C}\left(\boldsymbol{\theta}, \boldsymbol{\sigma}\right)\end{pmatrix}\right]\boldsymbol{\Upsilon}^{1/2}, \tag{16}$$

[1] Below their eq. (31), Fisher and Bollen (2020) list the weight matrix as equal to the inverse of the asymptotic covariance matrix of the polychoric covariance matrix. This is a typographical error and the weight matrix should be $\boldsymbol{W} = \boldsymbol{I}$, making it the ULS estimator (private communication with authors).

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

with $\boldsymbol{J}(\boldsymbol{\theta}) = \partial\boldsymbol{\sigma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^T$. Hence, we can apply the Satorra and Bentler (1994) adjustments to assess the overall goodness of fit of the model. In particular, the mean-scaled statistic and the mean-variance adjusted statistic are

$$T_{\mathrm{m}} = \frac{\mathrm{nr}}{\mathrm{tr}\left\{\hat{\boldsymbol{M}}\right\}} T\left(\hat{\boldsymbol{\theta}}\right) \quad \text{and} \quad T_{\mathrm{mv}} = \frac{n\mathrm{tr}\left\{\hat{\boldsymbol{M}}\right\}}{\mathrm{tr}\left\{\hat{\boldsymbol{M}}^2\right\}} T\left(\hat{\boldsymbol{\theta}}\right),$$

respectively, where r is the difference between the number of nonredundant free entries in $\boldsymbol{\Sigma}$ and the number of free parameters. The mean-scaled statistic is approximated by a Chi-square distribution with r degrees of freedom and the mean-variance adjusted statistic is approximated by a Chi-square distribution with $\left[\mathrm{tr}\left\{\hat{\boldsymbol{M}}\right\}\right]^2/\mathrm{tr}\left\{\hat{\boldsymbol{M}}^2\right\}$ degrees of freedom.

4.2. *Overidentification Test for Equations*

We mentioned earlier that an advantage of the MIIV approach is that we have an overidentification tests for every overidentified equation in the system (5). This provides a test of the validity of the MIIVs and the validity of the model specification. The reader is directed to Bollen (2019) for the implications of such a test. In this section, we present the asymptotic Chi-square distributed overidentification test that is applicable to continuous, ordinal, and binary endogenous observed variables. The derivation by and large follows the derivation in Jin and Cao (2018).

If equation (6) is correctly specified and the MIIVs are valid, $\boldsymbol{g}_j(\boldsymbol{\sigma}) = \boldsymbol{0}$. Then the delta method indicates that

$$\sqrt{n}\left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j}\boldsymbol{\theta}_1^{(j)}\right) = \sqrt{n}\left(\boldsymbol{g}_j(\boldsymbol{s}) - \boldsymbol{g}_j(\boldsymbol{\sigma})\right) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Omega}). \tag{17}$$

It is easy to show that

$$\boldsymbol{Q}_j = \boldsymbol{I} - \boldsymbol{\Omega}_j^{-1/2}\boldsymbol{\Sigma}_{vz,j}\left(\boldsymbol{\Sigma}_{vz,j}^T\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{\Sigma}_{vz,j}\right)^{-1}\boldsymbol{\Sigma}_{vz,j}^T\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{\Omega}_j^{1/2}.$$

is an idempotent matrix, but not necessarily symmetric. Let $\hat{\boldsymbol{Q}}_j$ be an estimator of $\boldsymbol{Q}_j$ by replacing $\boldsymbol{\Sigma}$ with $\boldsymbol{S}$ and $\boldsymbol{\theta}_1^{(j)}$ with $\hat{\boldsymbol{\theta}}_1^{(j)}$. Together with Eq. (7), it can be shown that

$$\sqrt{n}\hat{\boldsymbol{\Omega}}_j^{-1/2}\left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j}\hat{\boldsymbol{\theta}}_1^{(j)}\right) = \sqrt{n}\,\hat{\boldsymbol{Q}}_j\hat{\boldsymbol{\Omega}}_j^{-1/2}\boldsymbol{S}_{vy,j} = \sqrt{n}\,\hat{\boldsymbol{Q}}_j\hat{\boldsymbol{\Omega}}_j^{-1/2}\left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j}\boldsymbol{\theta}_1^{(j)}\right), \tag{18}$$

where the second equality holds since $\sqrt{n}\,\hat{\boldsymbol{Q}}_j\hat{\boldsymbol{\Omega}}_j^{-1/2}\boldsymbol{S}_{vz} = \boldsymbol{0}$. However, the asymptotic variance of (18) is $\boldsymbol{Q}_j\boldsymbol{Q}_j^T$, which is not of full rank. Let $\boldsymbol{G}_j$ and $\hat{\boldsymbol{G}}_j$ be the Moore–Penrose inverses of $\boldsymbol{Q}_j\boldsymbol{Q}_j^T$ and $\hat{\boldsymbol{Q}}_j\hat{\boldsymbol{Q}}_j^T$, respectively. The following theorem shows that

$$F = n\left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j}\hat{\boldsymbol{\theta}}_1^{(j)}\right)^T\hat{\boldsymbol{\Omega}}_j^{-1/2}\hat{\boldsymbol{G}}_j\hat{\boldsymbol{\Omega}}_j^{-1/2}\left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j}\hat{\boldsymbol{\theta}}_1^{(j)}\right) \tag{19}$$

is asymptotically Chi-square distributed.

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

Theorem 1. *As $n \to \infty$, F converges in distribution to a Chi-square distribution with $L_j - K_j$ degrees of freedom, provided that Eq. (6) is correctly specified and all MIIVs are valid, where L_j is the number of MIIVs, K_j is the number of explanatory variables in $\tilde{\boldsymbol{z}}_j$, and $L_j - K_j > 0$.*

The expression of F is the same as the equation (9) in Jin and Cao (2018) for observed ordinal variables. We have shown in Theorem 1 that it is also valid for a mixture of different types of variables. We can interpret the test statistic F as the generalized Wald statistic of Andrews (1987).

4.3. Connection with Sargan's Chi-Square

The test statistic (19) is asymptotically Chi-square in general, including the case where all observed variables are continuous. The Sargan's Chi-square statistic

$$F_{\text{Sargan}} = \frac{\boldsymbol{e}_j^T \boldsymbol{V}_j \left(\boldsymbol{V}_j^T \boldsymbol{V}_j\right)^{-1} \boldsymbol{V}_j^T \boldsymbol{e}_j}{\boldsymbol{e}_j^T \boldsymbol{e}_j / n},$$

applies when all observed variables are continuous, where $\boldsymbol{e}_j = \boldsymbol{y}_j - \boldsymbol{Z}_j \hat{\boldsymbol{\theta}}_1^{(j)}$ is the residual vector from Eq. (6). Equivalently,

$$F_{\text{Sargan}} = n \left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j} \hat{\boldsymbol{\theta}}_1^{(j)}\right)^T \frac{\boldsymbol{S}_{vv,j}^{-1}}{\hat{\varphi}_j^2} \left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j} \hat{\boldsymbol{\theta}}_1^{(j)}\right). \tag{20}$$

The asymptotic distribution of F_{Sargan} is derived by Sargan (1958). It is not surprising that F and F_{Sargan} are asymptotically equivalent, if all observed variables are continuous. The asymptotic distributions of F and F_{Sargan} essentially depend on the distribution of $\sqrt{n}\left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j} \boldsymbol{\theta}_1^{(j)}\right)$. If all observed variables are continuous and all MIIVs are valid,

$$\sqrt{n}\left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j} \boldsymbol{\theta}_1^{(j)}\right) = \sqrt{n}\left(\frac{1}{n} \boldsymbol{V}_j^T \boldsymbol{y}_j - \frac{1}{n} \boldsymbol{V}_j^T \boldsymbol{Z}_j \boldsymbol{\theta}_1^{(j)}\right) \xrightarrow{d} N\left(\boldsymbol{0}, \varphi_j^2 \boldsymbol{\Sigma}_{vv,j}\right).$$

Otherwise, the distribution (17) applies.

Despite of the asymptotic equivalence, F and F_{Sargan} are not necessarily the same due to the finite sample estimator of $\boldsymbol{\Omega}_j$. A properly chosen estimator of $\boldsymbol{\Omega}_j^{-1/2} \boldsymbol{G}_j \boldsymbol{\Omega}_j^{-1/2}$ is needed to yield the established test statistics in the literature. By Proposition 2, if $\boldsymbol{\Omega}_j$ is estimated by $\hat{\varphi}_j^2 \boldsymbol{S}_{vv,j}$, then $F = F_{\text{Sargan}}$ at any sample size. Alternatively, $\boldsymbol{\Omega}_j$ can be estimated by

$$\tilde{\boldsymbol{\Omega}}_j = \hat{\boldsymbol{\Omega}}_j - \left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j} \hat{\boldsymbol{\theta}}_1^{(j)}\right)\left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j} \hat{\boldsymbol{\theta}}_1^{(j)}\right)^T.$$

Consequently,

$$\tilde{F} = n \left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j} \hat{\boldsymbol{\theta}}_1^{(j)}\right)^T \tilde{\boldsymbol{\Omega}}_j^{-1/2} \tilde{\boldsymbol{G}}_j \tilde{\boldsymbol{\Omega}}_j^{-1/2} \left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j} \hat{\boldsymbol{\theta}}_1^{(j)}\right) \tag{21}$$

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

is asymptotically the same as F, where $\tilde{\boldsymbol{G}}_j$ is the Moore–Penrose inverse of $\tilde{\boldsymbol{Q}}_j \tilde{\boldsymbol{Q}}_j^T$ with

$$\tilde{\boldsymbol{Q}}_j = \boldsymbol{I} - \tilde{\boldsymbol{\Omega}}_j^{-1/2} \boldsymbol{S}_{vz,j} \left(\boldsymbol{S}_{vz,j}^T \boldsymbol{S}_{vv,j}^{-1} \boldsymbol{S}_{vz,j} \right)^{-1} \boldsymbol{S}_{vz,j}^T \boldsymbol{S}_{vv,j}^{-1} \tilde{\boldsymbol{\Omega}}_j^{1/2}. \tag{22}$$

If all variables are normal and $\hat{\boldsymbol{\Upsilon}} = 2\boldsymbol{D}^+ (\boldsymbol{S} \otimes \boldsymbol{S}) \boldsymbol{D}^{+T}$, then $\tilde{\boldsymbol{\Omega}}_j$ reduces to $\hat{\varphi}_j^2 \boldsymbol{S}_{vv,j}$ (Proposition 1) and $\tilde{F}$ is exactly the same as F_{Sargan} at any sample size.

4.4. Satorra-Bentler Adjusted Overidentification Tests of Equations

Besides the Chi-square distributed test statistic, Jin and Cao (2018) proposed alternative test statistics in the spirit of Satorra and Bentler (1994) and showed that the Satorra–Bentler-based statistics tend to have better small sample properties than the Chi-square distributed test statistic. In this section, the test statistic in Jin and Cao (2018) is extended to the mixture of different types of variables as well.

We can express the Sargan's Chi-square statistic F_{Sargan} as

$$F_{\text{Sargan}} = \frac{n \left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j} \hat{\boldsymbol{\theta}}_1^{(j)} \right)^T \boldsymbol{S}_{vv,j}^{-1} \left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j} \hat{\boldsymbol{\theta}}_1^{(j)} \right)}{\boldsymbol{S}_{yy,j} - \hat{\boldsymbol{\theta}}_1^{(j)T} \boldsymbol{S}_{zy,j} - \boldsymbol{S}_{zy,j}^T \hat{\boldsymbol{\theta}}_1^{(j)} + \hat{\boldsymbol{\theta}}_1^{(j)T} \boldsymbol{S}_{zz,j} \hat{\boldsymbol{\theta}}_1^{(j)}}. \tag{23}$$

Jin and Cao (2018) termed (23) as "Naive Sargan's Chi-square statistic" and showed that its asymptotic distribution is a weighted sum of independent Chi-square random variables with 1 degrees of freedom, if all observed variables are ordinal. The next theorem generalizes their results.

Theorem 2. *The asymptotic distribution of* F_{Sargan} *is a weighted sum of independent Chi-square random variables with 1 degrees of freedom, if all MIIVs are valid. The weights are the eigenvalues of* $\boldsymbol{\Pi}$*, where*

$$\boldsymbol{\Pi}_j = \varphi_j^{-2} \boldsymbol{\Omega}_j^{1/2} \left[\boldsymbol{\Sigma}_{vv,j}^{-1} - \boldsymbol{\Sigma}_{vv,j}^{-1} \boldsymbol{\Sigma}_{vz,j} \left(\boldsymbol{\Sigma}_{vz,j}^T \boldsymbol{\Sigma}_{vv,j}^{-1} \boldsymbol{\Sigma}_{vz,j} \right)^{-1} \boldsymbol{\Sigma}_{vz,j}^T \boldsymbol{\Sigma}_{vv,j}^{-1} \right] \boldsymbol{\Omega}_j^{1/2}.$$

Theorem 2 implies that we can apply the Satorra–Bentler-type adjustments to F_{Sargan}. In particular, the mean-scaled statistic and and the mean-variance adjusted statistic are

$$F_{\text{m}} = \frac{L_j - K_j}{\text{tr}\left\{ \hat{\boldsymbol{\Pi}}_j \right\}} F_{\text{Sargan}} \quad \text{and} \quad F_{mv} = \frac{\text{tr}\left\{ \hat{\boldsymbol{\Pi}}_j \right\}}{\text{tr}\left\{ \hat{\boldsymbol{\Pi}}_j^2 \right\}} F_{\text{Sargan}},$$

respectively, where $\hat{\boldsymbol{\Pi}}_j$ is the estimator of $\boldsymbol{\Pi}_j$, replacing $\boldsymbol{\Sigma}$ by $\boldsymbol{S}$ and $\boldsymbol{\theta}_1^{(j)}$ by $\hat{\boldsymbol{\theta}}_1^{(j)}$. We can approximate the mean-scaled statistic by a Chi-square distribution with $L_j - K_j$ degrees of freedom and approximate the mean-variance adjusted statistic by a Chi-square distribution with $\left[\text{tr}\left\{ \hat{\boldsymbol{\Pi}}_j \right\} \right]^2 / \text{tr}\left\{ \hat{\boldsymbol{\Pi}}_j^2 \right\}$ degrees of freedom.

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

If the observed variables are ordinal, the mean-scaled and mean-variance adjusted statistics derived by Jin and Cao (2018) are

$$\frac{L_j - K_j}{\text{tr}\left(\hat{\boldsymbol{\Delta}}_j\right)} F_{\text{Sargan}} \quad \text{and} \quad \frac{\text{tr}\left\{\hat{\boldsymbol{\Delta}}_j\right\}}{\text{tr}\left\{\hat{\boldsymbol{\Delta}}_j^2\right\}} F_{\text{Sargan}}, \tag{24}$$

respectively, where $\hat{\boldsymbol{\Delta}}_j$ is an estimator of

$$\boldsymbol{\Delta}_j = \frac{1}{1 - \boldsymbol{\Sigma}_{yz,j}\boldsymbol{\theta}_1^{(j)} - \boldsymbol{\theta}_1^{(j)T}\boldsymbol{\Sigma}_{yz,j}^T + \boldsymbol{\theta}_1^{(j)T}\boldsymbol{\Sigma}_{zz,j}\boldsymbol{\theta}_1^{(j)}} \boldsymbol{\Sigma}_{vv,j}^{-1/2} \left(\frac{\partial \boldsymbol{h}_j(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}^T}\right) \boldsymbol{\Upsilon} \left(\frac{\partial \boldsymbol{h}_j(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}^T}\right)^T \boldsymbol{\Sigma}_{vv,j}^{-1/2}$$

with $\boldsymbol{h}_j(\boldsymbol{\sigma}) = \boldsymbol{\Sigma}_{vy,j} - \boldsymbol{\Sigma}_{vz,j}\boldsymbol{\gamma}_j(\boldsymbol{\sigma})$. Despite that $\boldsymbol{\Pi}_j$ is not necessarily the same as $\boldsymbol{\Delta}_j$, the following corollary shows their connections.

Corollary 1. *$tr\left\{\boldsymbol{\Delta}_j\right\} = tr\left\{\boldsymbol{\Pi}_j\right\}$ and $tr\left\{\boldsymbol{\Delta}_j^2\right\} = tr\left\{\boldsymbol{\Pi}_j^2\right\}$ provided that observed variables are ordinal and all MIIVs are valid. However, their finite sample estimators differ by a function of $\boldsymbol{h}_j(\boldsymbol{s})$.*

Corollary 1 indicates that F_{m} and F_{mv} are asymptotically the same as the statistics developed by Jin and Cao (2018), if only ordinal variables are observed and all MIIVs are valid. If all observed variables are continuous and all MIIVs are valid, $\boldsymbol{\Pi}_j = \boldsymbol{Q}_j$, and F_{m}, F_{mv}, and F_{Sargan} are asymptotically the same. The implication is that F_{m} and F_{mv} can always be computed as test statistics and be approximated by their corresponding Chi-square distributions. Similarly to the Chi-square distributed test statistic, the consistent estimator $\tilde{\boldsymbol{\Omega}}_j$ needs to be used to ensure small sample equivalence of F_{m}, F_{mv}, and F_{Sargan} under the assumptions in Proposition 1, whereas $\hat{\boldsymbol{\Omega}}_j$ only ensures asymptotic equivalence. Hereafter, the mean-scaled statistic and mean-variance adjusted statistic using $\tilde{\boldsymbol{\Omega}}_j$ are denoted by $\tilde{F}_{\text{m}}$ and $\tilde{F}_{\text{mv}}$, respectively.

5. Monte Carlo Simulation

In this section a simulation study is conducted to investigate the finite sample properties of the proposed MIIV approach. The simulation is performed in R (R Core Team, 2020) and is built on the packages lavaan (Rosseel, 2012) and MIIVsem (Fisher et al., 2017). [2] Due to space limitation, we present only limited results in the subsequent section.

5.1. Simulation Design

To explore some of our proposed statistics, we use the SEM model in Li (2016) with two exogenous and three endogenous latent variables. The latent regression of the true data generating process with standardized coefficients is shown in Fig. 1. Every latent variable is measured by three indicators. The indicators of η_1, η_3, and η_5 are continuous, whereas the indicators of η_2 and η_4 are ordinal with five categories. The response probabilities are 0.04, 0.05, 0.21, 0.46, and 0.24, which is the slightly asymmetric condition in Li (2016). Each ordinal variable has an underlying continuous variable, and we set the error variances of the latter so that their total variances equal

[2] The R package can be found in the online supporting documents. It will also be integrated to the MIIVsem package in the future.

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

one. For simplicity, we also let the variances of continuous indicators equal one, but the error variances of continuous indicators are freely estimated. The factor loadings are set to 0.8, 0.65, and 0.5 with corresponding R^2 of indicators being 0.64, 0.42, and 0.25.

Given the scaling indicator, MIIVs are automatically generated from the hypothesized model using the package MIIVsem. Previous studies (e.g., Bollen et al., 2007; Jin and Cao, 2018) suggested that when the sample size is small, using fewer IVs results in less bias than using all IVs. Hence, we consider three estimators in the simulation study, i.e., MIIV with one more IV than the number of endogenous variables (denoted by MIIV1), MIIV with all possible IVs (denoted by MIIVall), and DWLS. We use the same set of MIIVs for estimation and in the model goodness-of-fit tests and the equation overidentification tests.

Since the MIIV approach uses the scaling indicators to set the scale of $\boldsymbol{\eta}$, we explore two choices of the scaling indicator, i.e., the indicators with standardized factor loadings 0.8 and the indicators with standardized factor loadings 0.5. The former sets the scale of $\boldsymbol{\eta}$ using the indicator with the highest R^2, whereas the latter uses the indicator with the lowest R^2. We also use the same MIIV scaling indicators for DWLS, so that we can compare the effects of scaling on different estimators. Fisher and Bollen (2020) have shown that the scaling indicator with a higher Shea (1997)'s R^2 contributes to lower biases in the MIIV approach. The R^2 of the scaling indicator measures the reliability of that indicator, whereas the Shea (1997)'s R^2 is a diagnostic for "weak" instrumental variables. Choosing the scaling indicator as the one with the lowest reliability or lowest Shea (1997)'s R^2 weakens the instrument compared to choosing the scaling indicator with the higher values. MIIVall uses all MIIVs in estimation. MIIV1 uses one more than the minimum number of MIIVs and these are chosen to maximize the Shea (1997)'s R^2. Figure 2 illustrates the values of Shea (1997)'s R^2 for the data generating process. There are ten equations for factor loadings and three equations for latent regression coefficients in our model. Figure 2 shows that the maximum Shea (1997)'s R^2 never exceeds 0.5 and the highest values occur when the scaling indicator has the highest R^2 value. The Shea (1997)'s R^2 is considerably lower for the scaling indicators with the lowest R^2s. Indeed for two of these equations, the Shea (1997)'s R^2 value is lower than 0.1. Although the high R^2 scaling indicator condition generally exceeds the low R^2 scaling indicator condition, the last two equations for the latent variable model have values of 0.1 or less. Hence, the simulation allows us to examine how very weak instrumental variables affect the estimates.

We fit both correctly specified and misspecified models. The hypothesized model is correctly specified, if it includes all paths in Figure 1. The misspecified model omits the dashed path b_{43}. For each model, six sample sizes are considered, namely, $n = 200$, 400, 800, 1200, 2000, and 3200. The number of replication is 10, 000, for each combination of the hypothesized model, sample size and choice of scaling indicator.

5.2. *Outcome Measures*

If only the coefficients and factor loadings of $\boldsymbol{\theta}_1$ are of interest, MIIV always converges. Non-convergence is only a potential problem for MIIV when $\boldsymbol{\theta}_2$ is estimated, but it is always a potential problem for DWLS. In the current study, $\boldsymbol{\theta}_2$ is estimated using the DWLS fit function, given the MIIV estimator of $\boldsymbol{\theta}_1$. Even for the converged estimates, the covariance matrices of error terms and factors are possibly non-positive definite. Both non-convergence and non-positive definite solutions are regarded as improper and are discarded in the current study.

To investigate the accuracy of the MIIV estimator, we compute the relative bias $100 \times$ median of $\left\{\theta^{-1}\left(\hat{\theta}_r - \theta\right), r = 1, ..., R\right\}$, where $\hat{\theta}_r$ is the estimate of θ in the rth replication and R is the number of proper replications. We use the median instead of the mean to prevent outliers from affecting the central tendency estimate in the smaller sample sizes. We focus on relative bias greater than 5 percent bias. The relative bias is also used to investigate the accuracy of the

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

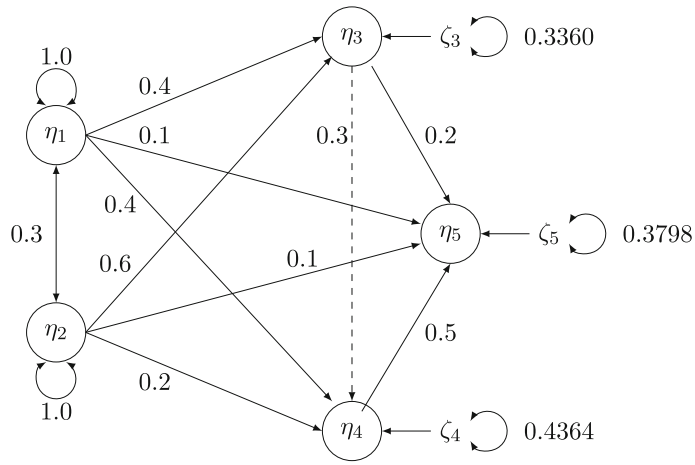


FIGURE 1.
Path diagram of the latent regression part of the SEM model in the simulation study. The dashed line is present in the true model but is omitted in the misspecified model. The population values are the standardized coefficients.

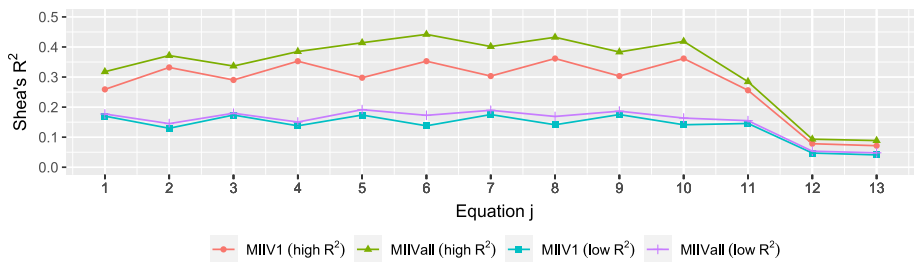


FIGURE 2.
Shea (1997)'s R^2 for each equation in the population model.

standard error estimates. However, we do not know the true value of the standard error. Hence, we use the sample standard deviation of the estimates as the pseudo-true value. Due to the presence of outliers, we use the interquartile range to trim the estimates, where the interquartile range is calculated by the third empirical quartile minus the first empirical quartile. Any estimate of a parameter is trimmed if the estimate is less than the first empirical quartile minus 3 times the interquartile range or larger than the third empirical quartile plus the 3 times the interquartile range. Since the converged replications with negative definite covariance matrices also reflect sampling fluctuation, the standard deviation of the converged estimates after trimming is treated as robust estimate of the standard error.

Concerning the goodness-of-fit test for the model as a whole, both the mean scaled test T_{m} and the mean-variance adjusted test T_{mv} are investigated. For DWLS, only the mean-scaled test is used. The goodness-of-fit tests of models are only computed for proper solutions. Regarding the overidentification tests for equations, the Chi-square tests ((19) and (21)), the mean scaled tests (F_{m} and $\tilde{F}_{\mathrm{m}}$), and the mean-variance adjusted tests (F_{mv} and $\tilde{F}_{\mathrm{mv}}$) are investigated. We will only focus on the overidentification equation tests for the misspecified model, since it contains both correctly and incorrectly specified equations when estimating $\boldsymbol{\theta}_1$. Because the overidentification equation tests does not require estimation of $\boldsymbol{\theta}_2$, we compute them for all converged solutions, including the solutions with negative covariance matrices. We use the empirical percentage of test statistics that exceed the χ^2 critical value at the significance level 0.05 for all test statistics.

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

TABLE 1.
Percentage of converged solutions with positive definite covariance matrices.

Model	R^2	Method	Sample size				
			200	400	800	1200	2000
Correctly specified	High	DWLS	99.15	100.00	100.00	100.00	100.00
		MIIV1	96.49	99.84	100.00	100.00	100.00
		MIIVall	98.56	99.97	100.00	100.00	100.00
	Low	DWLS	99.20	99.99	100.00	100.00	100.00
		MIIV1	87.70	98.22	99.96	99.97	100.00
		MIIVall	89.45	98.29	99.96	100.00	100.00
Misspecified	High	DWLS	98.46	99.97	100.00	100.00	100.00
		MIIV1	96.23	99.84	100.00	100.00	100.00
		MIIVall	98.74	99.99	100.00	100.00	100.00
	Low	DWLS	98.44	99.99	100.00	100.00	100.00
		MIIV1	87.55	97.77	99.86	100.00	100.00
		MIIVall	91.00	98.47	99.94	100.00	100.00

5.3. *Percentage of Proper Solutions*

Table 1 shows that MIIV and DWLS generally have similar percentages of proper solutions when the R^2 of the scaling indicator is high. If the R^2 is low and n is low, MIIV can yield more improper solutions than DWLS. Recall that MIIV always converges if only $\boldsymbol{\theta}_1$ is estimated. Hence, the improper solutions for MIIV originates from estimating $\boldsymbol{\theta}_2$.

5.4. *Correctly Specified Model*

When the model is correctly specified, all estimators are consistent estimators. We report the average of the absolute value of the relative bias for each of the four parameter sets ($\boldsymbol{\Lambda}$, $\boldsymbol{B}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Theta}$). Figure 3 shows that the biases of all estimators converges toward zero as sample size grows. Under the high R^2 condition, the average absolute value of the relative bias of the parameter estimates is low for DWLS, MIIVall and MIIV1. In contrast, when both the R^2 and sample size are low, DWLS maintains low bias, while MIIVall and MIIV1 exhibit more bias that diminishes as the sample size increases. It is also clear that MIIV1 tends to be less biased than MIIVall, especially when the scaling indicator has a low R^2. When the R^2 is low, MIIVall has the highest average absolute value of the relative bias for the latent variable equation error variances ($\boldsymbol{\Psi}$).

Figure 4 reveals the averaged absolute value of relative bias of the standard error estimates for each parameter set. We found some extreme outliers at the smaller sample sizes, so we trimmed values that were greater than three times the interquartile range above the third quartile or below the first quartile. This resulted in no more than 1.3% of cases trimmed which occurred when $n = 200$. The percentage trimmed rapidly decreased toward zero as n increased. For MIIV, both standard errors from (9) and (12) are expected to be accurate in the correctly specified model. If the R^2 of the scaling indicator is high, all standard errors are accurate. The only exception is that the relative bias of the standard errors is larger for the latent variable regression coefficients ($\boldsymbol{B}$) for DWLS at the smaller sample size. If the R^2 is low, the pattern is more complex. Under these conditions, the relative bias of the standard errors of the factor loadings and latent variable regression coefficients are lowest for MIIV1 (Eq. 9), MIIV1 (Eq. 12), and MIIVall (Eq. 12). The bias is greatest for MIIVall (equation 9) and DWLS, though these biases diminish as sample size increases. Continuing with the low R^2 results, the bias of the standard errors of the variance

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

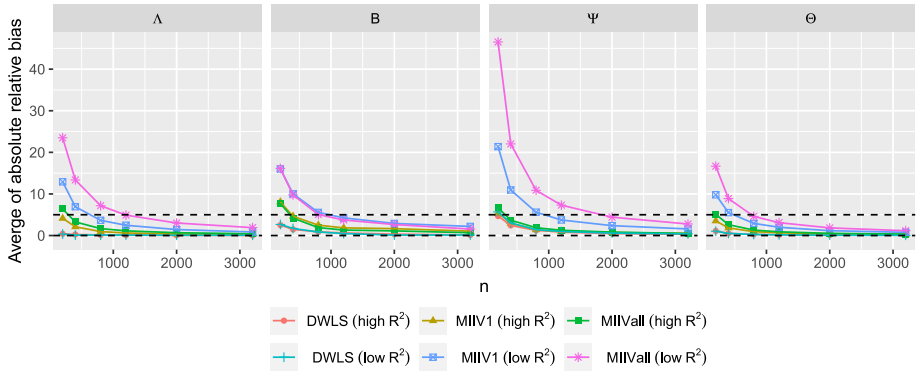


FIGURE 3.
Averaged absolute value of the relative bias of the parameter estimators when the model is correctly specified. Dashed lines at 0 and 5 percent relative bias.

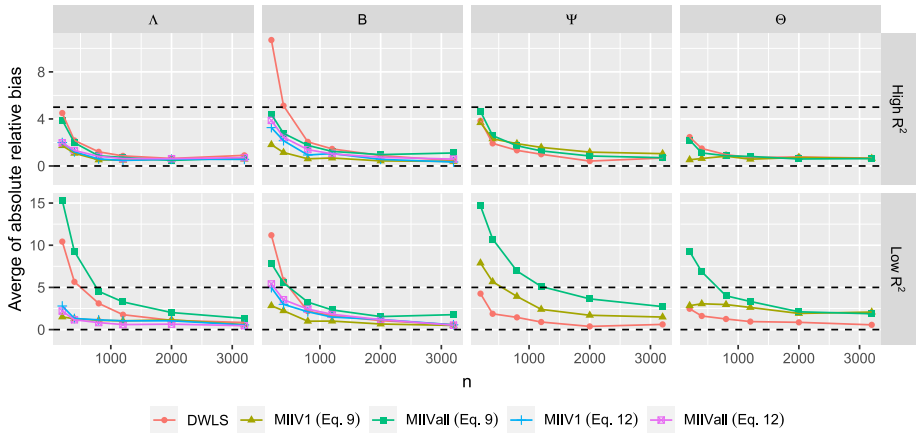


FIGURE 4.
Averaged absolute value of the relative bias of the standard error estimators when the model is correctly specified. Note: The MIIV standard errors of $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{B}}$ are computed from Eq. 9 or 12. The MIIV standard errors of $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\Theta}}$ are computed from Eq. 15, hence MIIV Eq. 9 is the same as MIIV Eq. 12.

parameter estimates are highest for MIIVall followed by MIIV1 and DWLS. These too diminish as the sample size increases. When MIIV1 is used, the standard error estimator (12) tends to be similar to (9). However, the standard error estimator (9) tends to be more biased than (12) for MIIVall.

Given that the model is correctly specified, we can explore the empirical size of the goodness-of-fit tests for models by computing the percentage of test statistics that exceed the critical value of a Chi-square with significance level 0.05. Figure 5 shows that using different number of IVs does not have strong effects on the empirical size. The mean-variance adjusted statistics for MIIV are the closest to the nominal level of all of the model test statistic (Jin et al., 2016). The size of the mean-scaled statistic for DWLS tends to be too high at about the same magnitude for the high and low R^2 scaling indicators. The mean-scaled statistics for MIIV are the least accurate, tending to be somewhat higher under the high R^2 condition and much higher under the low R^2 condition.

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

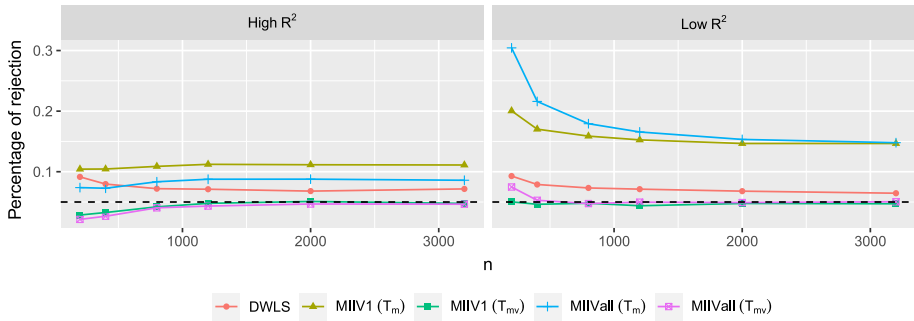


FIGURE 5.
Percentages of rejection of the goodness-of-fit tests when the model is correctly specified. The significance level is 0.05 (dashed line).

5.5. *Misspecified Model*

In the misspecified model, b_{43} is mistakenly fixed to 0, indicating that the latent regression with η_4 as the dependent variable ($j = 12$) is misspecified. Bollen (2001) and Bollen et al. (2018a) give the conditions under which the MIIV-2SLS and PIV estimator of coefficients will be robust under structural misspecifications such as in this model. Using their results, we know that MIIV factor loading ($\mathbf{\Lambda}$) estimates are identical under the correct and incorrect specifications and hence will have identical properties. Similarly, the MIIV estimates of the coefficients of b_{51}, b_{52}, b_{53} and b_{54} from the η_5 equation are robust to the structural misspecification. In contrast, the MIIV estimates of b_{41} and b_{42} from the latent regression for η_4 are not robust due to the omission of η_3 from the η_4 equation. In addition, the MIIV estimates of b_{31} and b_{32} from the latent regression for η_3 (with $j = 11$) are not robust. The reason is that in the misspecified model, it mistakenly appears that y_{10}, y_{11}, and y_{12}, the indicators of η_4, are suitable MIIVs because the path from ζ_3 to these indicators is cutoff by setting b_{43} to zero. Hence, we expect MIIVall to be consistent for $\mathbf{\Lambda}$, b_{51}, b_{52}, b_{43} and b_{54}, but inconsistent for b_{31}, b_{32}, b_{41}, and b_{42}. In contrast, MIIV1 can be consistent for b_{31} and b_{32} as long as y_{10}, y_{11}, and y_{12} are excluded. These robustness conditions describe the status of the factor loadings and latent variable coefficients, but there are no analogous analytic conditions for the variance parameters.

MIIV1 and MIIVall versions of MIIV are robust in the estimation of the factor loadings ($\mathbf{\Lambda}$). Therefore, the results of estimating the factor loadings for MIIV1 and MIIVall should be the same in the misspecified as they were in the correctly specified model, since the misspecification in the latent regression model does not affect the MIIVs for the measurement model. Figure 6 confirms this. Both MIIV1 and MIIVall yield consistent estimators of $\mathbf{\Lambda}$ regardless of the R^2 of the scaling indicator. DWLS does not seem to spread the bias to the measurement model and in the smaller samples the bias is even lower than that of MIIV1 and MIIVall, though the differences are slight in the bigger samples. In contrast, DWLS spreads the bias over all regression coefficients from the entire latent regression part. The DWLS average of absolute bias ranges from a low of 6 to 8 percentage bias for b_{31} and b_{32} to 50 percent bias for b_{51}, b_{52}, b_{53} and b_{54} and 70 percent bias for b_{41} and b_{42}. The MIIV1 and MIIVall estimators of b_{51}, b_{52}, b_{53} and b_{54} have large small sample biases, but decrease toward zero as n increases. The MIIV1 estimates of b_{31} and b_{32} have low biases, whereas the MIIVall estimator has a large bias if the scaling indicator has a large R^2. It is however interesting to see that the small sample bias of MIIVall is low if the scaling indicator has a low R^2. Hence, the MIIVall estimator of b_{31} and b_{32} can be inconsistent, but the asymptotic bias can still be low. When it comes to the effect of the scaling indicator, using a strong

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

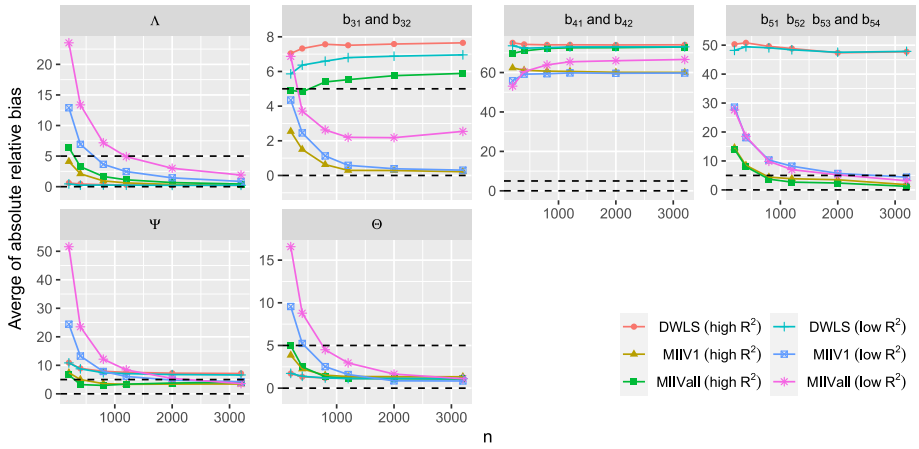


FIGURE 6.
Averaged absolute value of the relative bias of the parameter estimators when the model is misspecified.

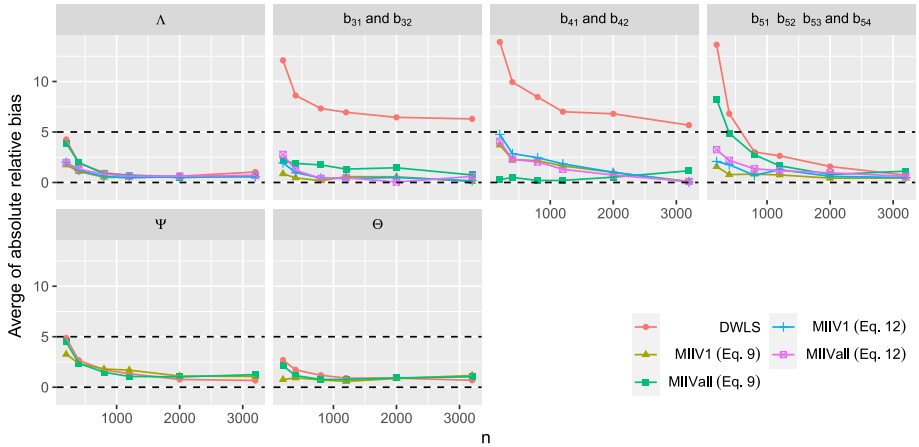
scaling indicator generally yields a lower bias than using a weak one for all parameters that can be consistently estimated.

If the model is misspecified, we still expect the standard errors from Eq. (9) to correctly quantify the variation of $\hat{\boldsymbol{\theta}}_1$, even though $\hat{b}_{31}$, $\hat{b}_{32}$, $\hat{b}_{41}$, and $\hat{b}_{42}$ are not consistent estimators of the true population value. However, the standard errors from (12) can be biased for these same coefficients, since it needs the MIIVs to be valid. From Eq. 10, if $\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{\text{vz},j}\boldsymbol{\gamma}_j$ $(\boldsymbol{s})$ is far away from $\mathbf{0}$, the bias is large. Otherwise the bias is small. It is seen from Figure 7 that all MIIV standard errors have low biases as n increases, regardless of the R^2. In contrast, the DWLS standard errors tend to be biased for the estimates of B. It is interesting to see that the standard error (12) also yields a low bias when the MIIVs are invalid. Nevertheless, we still prefer standard errors from Eq. 9, as it is theoretically valid. The bias of estimated standard errors is often higher if the R^2 of the scaling indicator is low. To produce Fig. 7, the percentage of trimmed replications is generally small (e.g., no more than 1.7% are trimmed when $n = 200$ and no more than 0.3% are trimmed when $n = 400$).

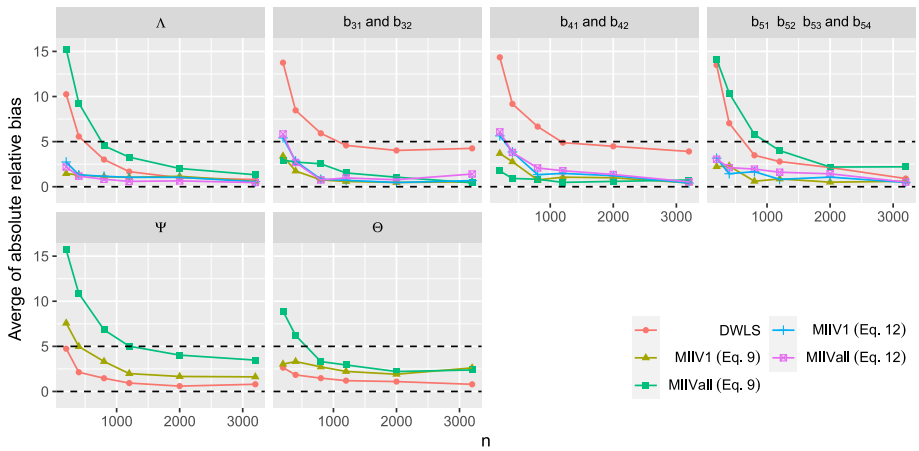
The empirical powers of the goodness-of-fit tests for the full model are in Fig. 8. Recall that the mean and variance adjusted test statistics (T_{mv}) for MIIV1 and MIIVall had the most accurate Type I error probabilities in the correctly specified model under both the high and low R^2 conditions. The T_{m} test statistic from MIIV1 and MIIVall and the DWLS test statistic rejected too frequently with the correct model. For the misspecified model, using a strong scaling indicator tends to yield higher power than using a weak scaling indicator for all MIIV and DWLS test statistics. Nevertheless as n increases, the power of all tests tends to increase. Under the high R^2 condition, MIIV1 (T_{m}) has the highest power with the test statistics MIIV1 (T_{mv}), DWLS, and MIIVall (T_{m}) the next highest. MIIVall (T_{mv}) has the lowest power. Under the low R^2 condition, MIIV1 (T_{m}), MIIVall (T_{m}), and DWLS have the highest power followed by MIIV1 (T_{mv}) and MIIVall (T_{mv}). If we want the best combination of accurate Type I error and high statistical power under the high R^2 condition, then MIIV1 (T_{mv}) is the best choice. The situation is more ambiguous for the low R^2 condition.

The MIIV estimators also provide overidentification tests for equations, one for each overidentified equation. Due to space limitaion, we only present the overidentification tests for the η_3 $(j = 11)$, η_4 $(j = 12)$, and η_5 $(j = 13)$ equations in the latent variable model. As discussed previously, we expect the empirical percentage rejection to approach 1 as n grows for equations η_3 $(j = 11)$ and η_4 $(j = 12)$ for MIIVall and to approach the nominal level 0.05 for η_5 $(j = 13)$

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press



(a) High R^2



(b) Low R^2

FIGURE 7.
Averaged absolute value of the relative bias of the standard error estimators when the model is misspecified and the R^2 of the scaling indicator is high. Note: The MIIV standard errors of $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{B}}$ are computed from equation (9) or (12). The MIIV standard errors of $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\Theta}}$ are computed from Eq. 15, hence MIIVEq. 9 is the same as MIIV(Eq. 12).

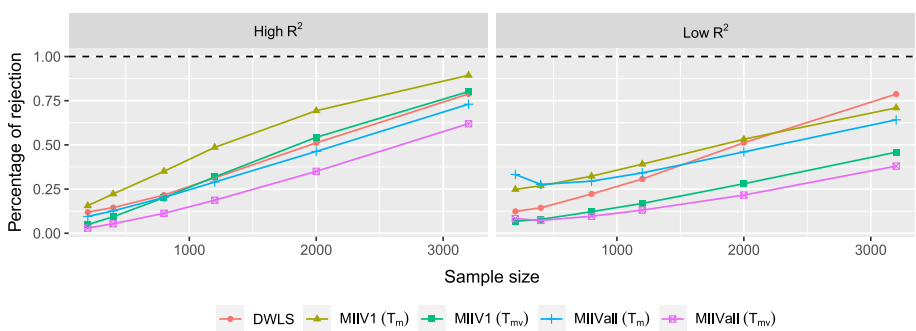
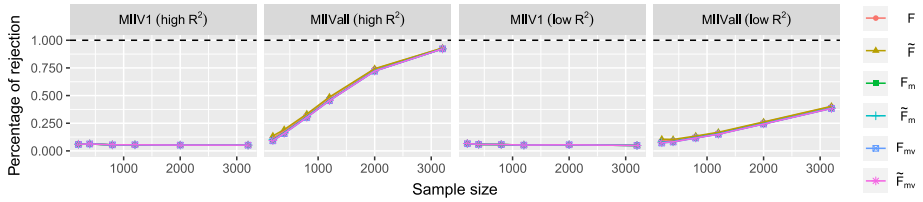
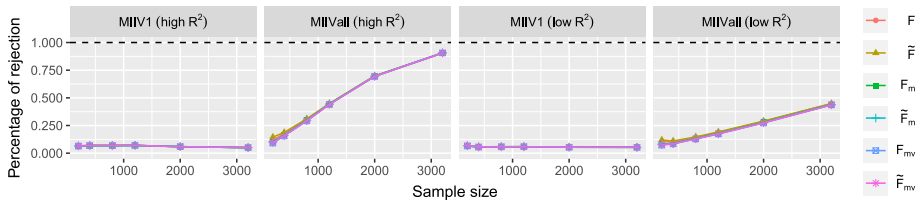


FIGURE 8.
Percentages of rejection of the goodness-of-fit tests when the model is misspecified. The significance level is 0.05.

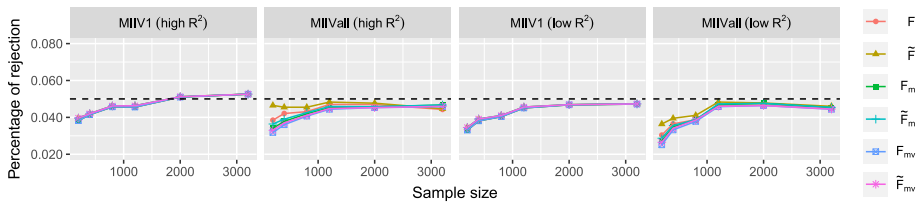
https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press



(a) Equation $j = 11$ for the latent regression with η_3 as the response variable



(b) Equation $j = 12$ for the latent regression with η_4 as the response variable



(c) Equation $j = 13$ for the latent regression with η_5 as the response variable

FIGURE 9.
Overidentification test for the η_3 ($j = 11$), η_4 ($j = 12$), and η_5 ($j = 13$) equations in the latent variable model at the significance level 0.05.

. Figure 9 illustrates that the sizes of all proposed overidentification tests become closer to the significance level 0.05 for the correctly specified η_5 equation with the rate of convergence more rapid if the R^2 of the scaling indicator is high. The power of the MIIVall tests for the η_3 ($j = 11$) and η_4 ($j = 12$) equations increase as n increases. However, the power of the MIIV1 tests is still far from 1, which is in line with the results in Jin and Cao (2018). It is also seen that the MIIV1 tests of F, F_{m}, and F_{mv} perform almost the same and that the MIIV1 tests of $\tilde{F}$, $\tilde{F}_{\text{m}}$, and $\tilde{F}_{\text{mv}}$ perform almost the same. Hence, they are almost visually indistinguishable in the figure. For MIIVall, using a strong scaling indicator yields a higher power than using a weak scaling indicator.

6. Empirical Example

To demonstrate the use of our MIIV estimator, we consider the Reisenzein (1986) dataset with 138 observations that is available in the MIIVsem package. It is hypothesized that the effect of Controllability (perceived controllability over a situation) on Help (decision to help another person) is mediated by Sympathy and Anger, shown in Fig. 10. Each factor is measured by three indicators. Every observed variable of Controllability and Anger has nine categories and are treated as continuous. The observed variables of Sympathy and Help are measured by either the five-point or nine-point Likert scales. They are converted to five-point Likert scales and are treated as categorical.

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

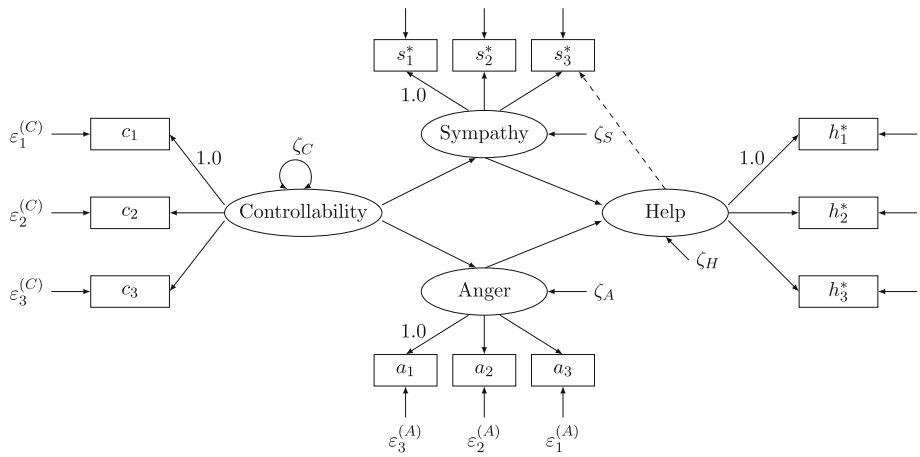


FIGURE 10.
Path diagonal of the the Reisenzein (1986) dataset.

TABLE 2.
Goodness-of-fit tests of the Reisenzein (1986) data set.

Test statistic	MIIV1			MIIVall			DWLS		
	Value	df	p value	Value	df	p value	Value	df	p value
Without the cross loading (s_3^* on *Help*)									
Mean	146.72	50.00	<0.0001	140.53	50.00	<0.0001	164.26	50.00	<0.0001
Mean-Var	39.45	13.44	<0.0001	42.78	15.22	<0.0001	51.09	15.55	<0.0001
With the cross loading (s_3^* on *Help*)									
Mean	36.54	49.00	0.91	48.48	49.00	0.49	44.09	49.00	0.67
Mean-Var	10.39	13.93	0.73	14.00	14.15	0.46	16.17	17.97	0.58

Mean Mean-scaled test, *Mean-Var* Mean-variance adjusted test.

We first use MIIV1, MIIVall and DWLS to estimate the model in Fig. 10 without the cross-loading s_3^* on Help. For both MIIV1 and MIIVall, there exists eight indicator equations with factor loadings (with left-hand side variables $c_2, c_3, s_2^*, s_3^*, a_2, a_3^*, h_2^*, h_3^*$) and three latent variable equations with regression coefficients (with left-hand side variables Sympathy, Anger, and Help). It is seen from the top panel of Table 2 that all goodness-of-fit tests reject the null hypothesis that the model fits the data well. However, the robust CFI, TLI, and RMSEA with DWLS are 0.996, 0.995, and 0.053, indicating a reasonable fit. In contrast, Table 3 shows that the MIIVall specification tests are highly significant when used to test the equations for s_3^* and Help. Here, the Bonferroni correction is used for simplicity to adjust the effect of multiple testing. Hence, the equations can be misspecified leading to invalid MIIVs. If we add the cross-loading s_3^* on Help, all goodness-of-fit tests become insignificant (bottom panel of Table 2). It is seen from Table 4 that MIIV and DWLS yield similar point estimates and standard errors when the revised model is fitted.

7. Discussion and Conclusion

We proposed a unified MIIV approach that handles a mixture of continuous, ordinal, or binary observed endogenous variables in SEM. Our method only requires a consistent estimator

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

TABLE 3.
Overidentification tests of the Reisenzein (1986) data set without the cross-loading s_3^* on Help. The significant tests after the Bonferroni correction is boldfaced.

Test	Content	Left-hand side variable in each equation										
		c_2	c_3	s_2^*	s_3^*	a_2	a_3	h_2^*	h_3^*	*Sympathy*	*Anger*	*Help*
F	Value	17.48	6.83	12.56	55.68	9.79	15.03	8.33	15.27	10.34	14.43	**19.91**
	df	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	4.00	4.00	**5.00**
	p_Value	0.04	0.65	0.18	0.00	0.37	0.09	0.50	0.08	0.04	0.01	**0.00**
$\tilde{F}$	Value	20.02	7.19	13.82	93.34	10.54	16.87	8.86	17.18	11.18	**16.12**	**23.26**
	df	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	4.00	**4.00**	**5.00**
	p_Value	0.02	0.62	0.13	0.00	0.31	0.05	0.45	0.05	0.02	**0.00**	**0.00**
F_m	Value	12.51	11.32	15.49	29.25	3.43	15.50	10.74	9.78	9.64	11.37	**18.24**
	df	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	4.00	4.00	**5.00**
	p_Value	0.19	0.25	0.08	0.00	0.94	0.08	0.29	0.37	0.05	0.02	**0.00**
$\tilde{F}_m$	Value	12.64	11.42	15.69	29.96	3.44	15.70	10.84	9.86	9.81	11.61	**18.73**
	df	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	4.00	4.00	**5.00**
	p_Value	0.18	0.25	0.07	0.00	0.94	0.07	0.29	0.36	0.04	0.02	**0.00**
F_{mv}	Value	9.33	8.00	12.86	20.63	2.19	9.74	8.22	8.06	8.76	10.24	**16.50**
	df	6.71	6.36	7.47	6.35	5.75	5.66	6.89	7.42	3.63	3.60	**4.53**
	p_Value	0.21	0.27	0.09	0.00	0.88	0.12	0.30	0.37	0.05	0.03	**0.00**
$\tilde{F}_{mv}$	Value	9.36	8.13	13.08	20.55	2.19	9.82	8.31	8.06	8.87	10.37	**16.91**
	df	6.67	6.41	7.50	6.17	5.73	5.63	6.90	7.36	3.62	3.57	**4.51**
	p_Value	0.20	0.26	0.09	0.00	0.88	0.11	0.30	0.36	0.05	0.03	**0.00**

of the covariance matrix and an asymptotic covariance matrix of its elements. We provide point estimators of all parameters and their asymptotic standard errors. Furthermore, we provide model goodness of fit test statistics and local tests of overidentified equations. The latter of which correspond to the classic Sargan's Chi-square test and the tests in Jin and Cao (2018). In addition, we give Satorra–Bentler-type modifications to these test statistics. The simulation study shows that the small sample properties of the proposed MIIV approach is generally in line with the theoretical results. That is, the MIIV estimators are more robust to structural misspecifications than are the systemwide estimators; the overidentification equation tests provide useful local tests of fit, and the model goodness of fit tests provide useful diagnostics on global fit. The performance of these MIIV tests is best with strong MIIVs and deteriorates if the MIIVs are weak.

The MIIV procedure is applicable to a large class of latent variable models, as long as they can be expressed as (1) and (2). Among these are the confirmatory factor analysis model (Jin et al., 2016; Jin and Cao, 2018; Nestler, 2013), the latent growth model (Nestler, 2015b), and specification tests for nonlinear terms in the latent variable model (Nestler, 2015a). The MIIV estimation essentially depends on $\boldsymbol{S}$, which can be interpreted as a method of moment. Various latent variable models such as the item response theory models are expressed from the likelihood perspective. When using the probit link, some connections between the MIIV estimates and item response theory parameters are possible (Takane and de Leeuw, 1987). But for item response-type models in general the connections to a MIIV approach are less clear.

The finite sample equivalence among various test statistics depend on the estimator of $\boldsymbol{\Omega}$. The estimators $\hat{\boldsymbol{\Omega}}$ and $\tilde{\boldsymbol{\Omega}}$ are asymptotically the same but differ by a consistent estimator of $\mathbf{0}$ if all MIIVs are valid. If some MIIVs are not valid, $\hat{\boldsymbol{\Omega}}$ and $\tilde{\boldsymbol{\Omega}}$ are not always asymptotically the same. In other words, the established equivalence is only for the null distribution. A rigorous power analysis and extensive simulation studies to investigate the distributions of the test statistics under

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

TABLE 4.
Point estimates and the standard errors of the Reisenzein (1986) dataset with the cross-loading s_3^* on Help. For MIIV, the standard error of $\hat{\boldsymbol{\theta}}_1$ are computed from (9).

Parameter	DWLS		MIIV1		MIIVall	
	Est.	SE	Est.	SE	Est.	SE
Parameters in $\boldsymbol{\theta}_1$						
Controllability→ c_2	1.089	0.193	1.088	0.154	1.042	0.153
Controllability→ c_3	1.275	0.262	1.169	0.188	1.149	0.188
Sympathy→ s_2^*	0.761	0.060	0.756	0.054	0.734	0.056
Sympathy→ s_3^*	0.603	0.058	0.603	0.077	0.555	0.076
Anger→ a_2	0.933	0.154	0.909	0.128	0.906	0.134
Anger→ a_3	0.885	0.130	0.881	0.101	0.882	0.116
Help→ h_2^*	0.951	0.031	0.953	0.032	0.955	0.033
Help→ h_3^*	0.806	0.040	0.819	0.040	0.829	0.038
Help→ s_3^*	0.339	0.054	0.338	0.064	0.361	0.065
Controllability→Sympathy	−0.278	0.053	−0.267	0.044	−0.279	0.045
Controllability→Anger	0.651	0.149	0.652	0.133	0.636	0.132
Sympathy→Help	0.265	0.081	0.340	0.096	0.360	0.104
Anger→Help	−0.243	0.050	−0.212	0.048	−0.199	0.048
Parameters in $\boldsymbol{\theta}_2$						
var (ζ_C)	4.965	1.699	5.132	1.534	5.206	1.599
var (ζ_S)	0.610	0.102	0.601	0.083	0.590	0.086
var (ζ_A)	1.981	0.596	2.144	0.621	2.293	0.650
var (ζ_H)	0.524	0.071	0.507	0.075	0.500	0.075
var $\left(\varepsilon_1^{(C)}\right)$	2.103	0.579	1.935	0.350	1.861	0.315
var $\left(\varepsilon_2^{(C)}\right)$	2.104	0.734	1.918	0.402	2.333	0.424
var $\left(\varepsilon_3^{(C)}\right)$	0.824	0.858	1.881	0.346	2.021	0.484
var $\left(\varepsilon_1^{(A)}\right)$	1.442	0.402	1.201	0.326	1.129	0.328
var $\left(\varepsilon_2^{(A)}\right)$	1.730	0.389	1.717	0.276	1.678	0.269
var $\left(\varepsilon_3^{(A)}\right)$	1.739	0.385	1.582	0.309	1.520	0.330

Est. Estimate, *SE* Standard error.

the alternative hypothesis and the effects of using different $\boldsymbol{\Omega}$ estimators in order to make further recommendations.

We also explored weak instruments in our simulation. An IV is commonly viewed as weak if its correlation with the endogenous variable is small. With the presence of weak IVs, it is known in the econometrics literature that the 2SLS estimator is biased, and the overidentification test has the wrong size when even a small correlation between instruments and equation errors exist (e.g., Bound et al., 1995; Hahn and Hausman, 2003; 2005). In our simulation, we considered both MIIV1 and MIIVall. The weak IVs are not excluded from MIIVall, but MIIV1 only incorporates the strongest IVs for the given data set. We found that scaling indicators with low R^2s were associated with weak instruments. A number of authors have proposed diagnostics for weak instruments including Hahn and Hausman (2002), Kleibergen and Paap (2006), Olea and Pflueger (2013), Shea (1997), and Stock and Yogo (2005). Fisher and Bollen (2020) showed that the accuracy of estimation is negatively related with the Shea's partial R^2. Further studies are needed to provide guidelines on what to do when all IVs are weak or when some IVs are strong but others are weak. Finally, we recognize the limits of any simulation study and encourage other simulation

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

studies that explore a wider variety of conditions to determine the degree to which our findings generalize.

Appendix: Mathematical Proof

Proof of Proposition 1. Let σ_{vz} be an element in $\boldsymbol{\Sigma}_{\text{vz},j}$, σ_{vv} be an element in $\boldsymbol{\Sigma}_{\text{vv},j}$, and σ_{vy} be an element in $\boldsymbol{\Sigma}_{\text{vy},j}$. Then

$$\frac{\partial \boldsymbol{g}_j(\boldsymbol{\sigma})}{\partial \sigma_{\text{vz}}} = -\frac{\partial \boldsymbol{\Sigma}_{\text{vz},j}}{\partial \sigma_{\text{vz}}}\boldsymbol{\theta}_1^{(j)}, \quad \frac{\partial \boldsymbol{g}_j(\boldsymbol{\sigma})}{\partial \sigma_{vv}} = \mathbf{0}, \quad \frac{\partial \boldsymbol{g}_j(\boldsymbol{\sigma})}{\partial \sigma_{vy}} = \frac{\partial \boldsymbol{\Sigma}_{vy,j}}{\partial \sigma_{vy}}.$$

For any other σ that does not belong to $\boldsymbol{\Sigma}_{\text{vz},j}$, $\boldsymbol{\Sigma}_{\text{vv},j}$, nor $\boldsymbol{\Sigma}_{\text{vy},j}$, $\partial \boldsymbol{g}_j(\boldsymbol{\sigma})/\partial\sigma = 0$. In matrix form, the nonzero part in $\partial \boldsymbol{g}_j(\boldsymbol{\sigma})/\partial \boldsymbol{\sigma}^T$ is essentially

$$\left(\boldsymbol{I} \quad -\boldsymbol{\theta}_1^{(j)T} \otimes \boldsymbol{I}\right),$$

subject to a permutation of columns, where $\otimes$ is the Kronecker product. If $\hat{\boldsymbol{\Upsilon}} = 2\boldsymbol{D}^{+}(\boldsymbol{S} \otimes \boldsymbol{S})\boldsymbol{D}^{+T}$, the estimator of $\boldsymbol{\Upsilon}$ that corresponds to the nonzero part in $\partial \boldsymbol{g}_j(\boldsymbol{s})/\partial \boldsymbol{\sigma}^T$ is

$$\begin{pmatrix} \boldsymbol{S}_{\text{yy},j}\boldsymbol{S}_{\text{vv},j} & \boldsymbol{S}_{\text{zy},j}^T \otimes \boldsymbol{S}_{\text{vv},j} \\ \boldsymbol{S}_{\text{zy},j} \otimes \boldsymbol{S}_{\text{vv},j} & \boldsymbol{S}_{\text{zz},j} \otimes \boldsymbol{S}_{\text{vv},j} \end{pmatrix} + \begin{pmatrix} \boldsymbol{S}_{\text{vy},j}\boldsymbol{S}_{\text{vy},j}^T & \boldsymbol{S}_{\text{vz},j} \otimes \boldsymbol{S}_{\text{vy},j}^T \\ \boldsymbol{S}_{\text{vz},j}^T \otimes \boldsymbol{S}_{\text{vu},j} & \hat{\boldsymbol{A}}_j \end{pmatrix},$$

where $\hat{\boldsymbol{A}}_j$ is a block matrix such that the (i,k)th block is $\boldsymbol{S}_{\text{vz}_k,j}\boldsymbol{S}_{\text{vz}_i,j}^T$. Some algebra shows that

$$\begin{aligned} \frac{\partial \boldsymbol{g}_j(\boldsymbol{s})}{\partial \boldsymbol{\sigma}^T}\hat{\boldsymbol{\Upsilon}}\left(\frac{\partial \boldsymbol{g}_j(\boldsymbol{s})}{\partial \boldsymbol{\sigma}^T}\right)^T =& \left(\boldsymbol{S}_{\text{yy},j} - 2\hat{\boldsymbol{\theta}}_1^{(j)T}\boldsymbol{S}_{\text{zy},j} + \hat{\boldsymbol{\theta}}_1^{(j)T}\boldsymbol{S}_{\text{zz},j}\hat{\boldsymbol{\theta}}_1^{(j)}\right)\boldsymbol{S}_{\text{vv},j} + \boldsymbol{S}_{\text{vy},j}\boldsymbol{S}_{\text{vy},j}^T \\ & - \left(\boldsymbol{S}_{\text{vz},j}\hat{\boldsymbol{\theta}}_1^{(j)}\right) \otimes \boldsymbol{S}_{\text{vy},j}^T - \left(\hat{\boldsymbol{\theta}}_1^{(j)T}\boldsymbol{S}_{\text{vz},j}^T\right) \otimes \boldsymbol{S}_{\text{vy},j} \\ & + \left(\hat{\boldsymbol{\theta}}_1^{(j)T} \otimes \boldsymbol{I}\right)\hat{\boldsymbol{A}}_j\left(\hat{\boldsymbol{\theta}}_1^{(j)} \otimes \boldsymbol{I}\right) \\ =& \hat{\varphi}_j^2\boldsymbol{S}_{\text{vv},j} + \left(\boldsymbol{S}_{\text{vy},j} - \boldsymbol{S}_{\text{vz},j}\hat{\boldsymbol{\theta}}_1^{(j)}\right)\left(\boldsymbol{S}_{\text{vy},j} - \boldsymbol{S}_{\text{vz},j}\hat{\boldsymbol{\theta}}_1^{(j)}\right)^T, \end{aligned}$$

where the last equality holds because of

$$\left(\hat{\boldsymbol{\theta}}_1^{(j)T} \otimes \boldsymbol{I}\right)\hat{\boldsymbol{A}}_j\left(\hat{\boldsymbol{\theta}}_1^{(j)} \otimes \boldsymbol{I}\right) = \boldsymbol{S}_{vz,j}\hat{\boldsymbol{\theta}}_1^{(j)}\hat{\boldsymbol{\theta}}_1^{(j)T}\boldsymbol{S}_{vz,j}^T$$

by expanding the elements in the Kronecker product. Then, the proposition holds because

$$\boldsymbol{S}_{vz,j}^T\boldsymbol{S}_{vv,j}^{-1}\left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j}\hat{\boldsymbol{\theta}}_1^{(j)}\right) = \mathbf{0}.$$

□

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

Proof of Proposition 2. From the proof of Proposition 1, the nonzero part in $\partial \boldsymbol{g}_j(\boldsymbol{\sigma})/\partial \boldsymbol{\sigma}^T$ is essentially

$$\left(\boldsymbol{I} \quad -\boldsymbol{\theta}_1^{(j)T} \otimes \boldsymbol{I}\right),$$

subject to a permutation of columns. If all observed variables are continuous, the submatrix in $\boldsymbol{\Upsilon}$ corresponding to $\boldsymbol{\Sigma}_{vy,j}$ and $\boldsymbol{\Sigma}_{vz,j}$ is of the form

$$\begin{pmatrix} E\left(y_j^{*2}\boldsymbol{v}_j^*\boldsymbol{v}_j^{*T}\right) - E\left(y_j^*\boldsymbol{v}_j^*\right)E\left(y_j^*\boldsymbol{v}_j^{*T}\right) & E\left[\left(y_j^*\boldsymbol{z}_j^{*T}\right)\otimes\left(\boldsymbol{v}_j^*\boldsymbol{v}_j^{*T}\right)\right] - E\left(y_j^*\otimes\boldsymbol{v}_j^*\right)E\left(\boldsymbol{z}_j^{*T}\otimes\boldsymbol{v}_j^{*T}\right) \\ E\left[\left(y_j^*\boldsymbol{z}_j^*\right)\otimes\left(\boldsymbol{v}_j^*\boldsymbol{v}_j^{*T}\right)\right] - E\left(\boldsymbol{z}_j^*\otimes\boldsymbol{v}_j^*\right)E\left(y_j^*\otimes\boldsymbol{v}_j^{*T}\right) & E\left[\left(\boldsymbol{z}_j^*\boldsymbol{z}_j^{*T}\right)\otimes\left(\boldsymbol{v}_j^*\boldsymbol{v}_j^{*T}\right)\right] - E\left(\boldsymbol{z}_j^*\otimes\boldsymbol{v}_j^*\right)E\left(\boldsymbol{z}_j^{*T}\otimes\boldsymbol{v}_j^{*T}\right) \end{pmatrix}.$$

Some algebra shows that

$$\begin{aligned} \frac{\partial \boldsymbol{g}_j(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}^T}\boldsymbol{\Upsilon}\left(\frac{\partial \boldsymbol{g}_j(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}^T}\right)^T &= E\left(y_j^{*2}\boldsymbol{v}_j^*\boldsymbol{v}_j^{*T}\right) - 2E\left[y_j^*\left(\boldsymbol{z}_j^{*T}\boldsymbol{\theta}_1^{(j)}\right)\left(\boldsymbol{v}_j^*\boldsymbol{v}_j^{*T}\right)\right] \\ &\quad + E\left[\left(\boldsymbol{z}_j^{*T}\boldsymbol{\theta}_1^{(j)}\right)^2\left(\boldsymbol{v}_j^*\boldsymbol{v}_j^{*T}\right)\right] \\ &\quad - \left(\boldsymbol{\Sigma}_{vy,j} - \boldsymbol{\Sigma}_{vz,j}\boldsymbol{\theta}_1^{(j)}\right)\left(\boldsymbol{\Sigma}_{vy,j} - \boldsymbol{\Sigma}_{vz,j}\boldsymbol{\theta}_1^{(j)}\right)^T \\ &= E\left(\boldsymbol{v}_j^*\boldsymbol{v}_j^{*T}e_j^2\right) - \left(\boldsymbol{\Sigma}_{vy,j} - \boldsymbol{\Sigma}_{vz,j}\boldsymbol{\theta}_1^{(j)}\right)\left(\boldsymbol{\Sigma}_{vy,j} - \boldsymbol{\Sigma}_{vz,j}\boldsymbol{\theta}_1^{(j)}\right)^T. \end{aligned}$$

If the model is correctly specified and all MIIVs are valid, then $E\left(e_j^2\boldsymbol{v}_j^*\boldsymbol{v}_j^{*T}\right) = \varphi_j^2\boldsymbol{\Sigma}_{vv,j}$ and $\boldsymbol{\Sigma}_{vy,j} - \boldsymbol{\Sigma}_{vz,j}\boldsymbol{\theta}_1^{(j)} = \boldsymbol{0}$, which completes the proof. □

Proof of Equation (14). The assumptions that we need include (1) $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$, (2) $\boldsymbol{s} - \boldsymbol{\sigma}\left(\hat{\boldsymbol{\theta}}\right) \xrightarrow{p} \boldsymbol{0}$, (3) $\partial^3 T(\boldsymbol{\theta})/\partial\theta_{2,j}\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^T$ is bounded in probability in an open set $\mathbb{O}$ that contains $\boldsymbol{\theta}_2$ for any $\theta_{2,j}$, the jth entry in $\boldsymbol{\theta}_2$, and (4) $\partial^2\boldsymbol{\sigma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}^T$ is also bounded in $\mathbb{O}$. Assumptions (1) and (2) hold if the model is correctly specified and all MIIVs are valid. Assumptions (3) and (4) hold since $T(\boldsymbol{\theta})$ and $\boldsymbol{\sigma}(\boldsymbol{\theta})$ have continuous higher-order partial derivatives given by our SEM model.

Since $\boldsymbol{\sigma}(\boldsymbol{\theta})$ has continuous third-order partial derivatives with respect to $\boldsymbol{\theta}$, the standard Taylor expansion yields

$$\begin{aligned} 0 &= \frac{\partial T\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2\right)}{\partial\theta_{2,j}} \\ &= \frac{\partial T\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)}{\partial\theta_{2,j}} + \frac{\partial^2 T\left(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_2^T}\left(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2\right) \\ &\quad + \frac{1}{2}\left(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2\right)^T\frac{\partial^3 T\left(\hat{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^T}\left(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2\right), \end{aligned} \tag{25}$$

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

where $\theta_{2,j}$ is the jth entry of $\boldsymbol{\theta}_2$ and $\tilde{\boldsymbol{\theta}}_2$ lies between $\hat{\boldsymbol{\theta}}_2$ and $\boldsymbol{\theta}_2$. Similarly, the standard Taylor expansion yields

$$\frac{\partial T\left(\hat{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2\right)}{\partial\theta_{2,j}}=\frac{\partial T\left(\boldsymbol{\theta}\right)}{\partial\theta_{2,j}}+\frac{\partial^2 T\left(\boldsymbol{\theta}\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_1^T}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1\right)+\frac{1}{2}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1\right)^T\frac{\partial^3 T\left(\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1^T}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1\right),\tag{26}$$

and

$$\frac{\partial^2 T\left(\hat{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_2}=\frac{\partial^2 T\left(\boldsymbol{\theta}\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_2}+\frac{\partial^3 T\left(\breve{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_2^T}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1\right),\tag{27}$$

where both $\tilde{\boldsymbol{\theta}}_1$ and $\breve{\boldsymbol{\theta}}_1$ lie between $\hat{\boldsymbol{\theta}}_1$ and $\boldsymbol{\theta}_1$. Then, after plugging (26) and (27) into (25), we obtain

$$\begin{aligned}&\sqrt{n}\left[\frac{\partial^2 T\left(\boldsymbol{\theta}\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_2^T}+\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1\right)^T\frac{\partial^3 T\left(\breve{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_2^T}+\frac{1}{2}\left(\hat{\boldsymbol{\theta}}_2-\boldsymbol{\theta}_2\right)^T\frac{\partial^3 T\left(\hat{\boldsymbol{\theta}}_1,\tilde{\boldsymbol{\theta}}_2\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^T}\right]\left(\hat{\boldsymbol{\theta}}_2-\boldsymbol{\theta}_2\right)\\&=-\sqrt{n}\frac{\partial T\left(\boldsymbol{\theta}\right)}{\partial\theta_{2,j}}-\frac{\partial^2 T\left(\boldsymbol{\theta}\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_1^T}\sqrt{n}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1\right)-\frac{1}{2}\sqrt{n}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1\right)^T\frac{\partial^3 T\left(\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1^T}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1\right).\end{aligned}$$

By the assumptions mentioned in the beginning of the proof and $\hat{\boldsymbol{W}}\xrightarrow{p}\boldsymbol{W}$, we obtain

$$\begin{aligned}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1\right)^T\frac{\partial^3 T\left(\breve{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_2^T}&\xrightarrow{p}0,\\\left(\hat{\boldsymbol{\theta}}_2-\boldsymbol{\theta}_2\right)^T\frac{\partial^3 T\left(\hat{\boldsymbol{\theta}}_1,\tilde{\boldsymbol{\theta}}_2\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^T}&\xrightarrow{p}0,\\\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1\right)^T\frac{\partial^3 T\left(\tilde{\boldsymbol{\theta}}_1,\boldsymbol{\theta}_2\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1^T}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1\right)&\xrightarrow{p}0.\end{aligned}$$

Hence,

$$\frac{\partial^2 T\left(\boldsymbol{\theta}\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_2^T}\sqrt{n}\left(\hat{\boldsymbol{\theta}}_2-\boldsymbol{\theta}_2\right)=-\sqrt{n}\frac{\partial T\left(\boldsymbol{\theta}\right)}{\partial\theta_{2,j}}-\frac{\partial^2 T\left(\boldsymbol{\theta}\right)}{\partial\theta_{2,j}\partial\boldsymbol{\theta}_1^T}\sqrt{n}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1\right)+o_{\mathrm{P}}\left(1\right).$$

By collecting all entries in $\boldsymbol{\theta}_2$, we obtain

$$\frac{\partial^2 T\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^T}\sqrt{n}\left(\hat{\boldsymbol{\theta}}_2-\boldsymbol{\theta}_2\right)=-\sqrt{n}\frac{\partial T\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}_2}-\frac{\partial^2 T\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_1^T}\sqrt{n}\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1\right)+o_{\mathrm{P}}\left(1\right).$$

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

By equation (8), we further obtain

$$\frac{\partial^2 T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^T} \sqrt{n}\left(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2\right) = -\sqrt{n}\frac{\partial T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} - \frac{\partial^2 T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_1^T} \boldsymbol{K}(\boldsymbol{\sigma}) \sqrt{n}(\boldsymbol{s} - \boldsymbol{\sigma}) + o_{\mathrm{P}}(1). \tag{28}$$

Note that the partial derivatives in (28) are

$$\begin{aligned}
\frac{\partial T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} &= -2\boldsymbol{J}_2^T(\boldsymbol{\theta}) \hat{\boldsymbol{W}}(\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})), \\
\frac{\partial^2 T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_1^T} &= 2\boldsymbol{J}_2^T(\boldsymbol{\theta}) \hat{\boldsymbol{W}} \boldsymbol{J}_1(\boldsymbol{\theta}) - 2\frac{\partial^2 \boldsymbol{\sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_1^T} \left\{\boldsymbol{I} \otimes \left[\hat{\boldsymbol{W}}(\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta}))\right]\right\}, \\
\frac{\partial^2 T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^T} &= 2\boldsymbol{J}_2^T(\boldsymbol{\theta}) \hat{\boldsymbol{W}} \boldsymbol{J}_2(\boldsymbol{\theta}) - 2\frac{\partial^2 \boldsymbol{\sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^T} \left\{\boldsymbol{I} \otimes \left[\hat{\boldsymbol{W}}(\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta}))\right]\right\},
\end{aligned}$$

where $\boldsymbol{s} - \boldsymbol{\sigma}\left(\hat{\boldsymbol{\theta}}\right) \xrightarrow{p} \mathbf{0}$ under our assumption. Then we get the expansion (14). □

Proof of the distribution of $nT\left(\hat{\boldsymbol{\theta}}\right)$. If the model is correctly specified, expansions (8) and (14) imply

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) = \begin{pmatrix} \boldsymbol{K}(\boldsymbol{\sigma}) \\ \boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{\sigma}) \end{pmatrix} \sqrt{n}(\boldsymbol{s} - \boldsymbol{\sigma}) + o_{\mathrm{P}}(1).$$

The delta method implies

$$\sqrt{n}\left(\boldsymbol{\sigma}\left(\hat{\boldsymbol{\theta}}\right) - \boldsymbol{\sigma}\right) = \boldsymbol{J}(\boldsymbol{\theta}) \begin{pmatrix} \boldsymbol{K}(\boldsymbol{\sigma}) \\ \boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{\sigma}) \end{pmatrix} \sqrt{n}(\boldsymbol{s} - \boldsymbol{\sigma}) + o_{\mathrm{P}}(1),$$

where $\boldsymbol{J}(\boldsymbol{\theta}) = \partial \boldsymbol{\sigma}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}^T$. Hence,

$$\begin{aligned}
\sqrt{n}\left(\boldsymbol{s} - \boldsymbol{\sigma}\left(\hat{\boldsymbol{\theta}}\right)\right) &= \sqrt{n}(\boldsymbol{s} - \boldsymbol{\sigma}) - \sqrt{n}\left(\boldsymbol{\sigma}\left(\hat{\boldsymbol{\theta}}\right) - \boldsymbol{\sigma}\right) \\
&= \left[\boldsymbol{I} - \boldsymbol{J}(\boldsymbol{\theta}) \begin{pmatrix} \boldsymbol{K}(\boldsymbol{\sigma}) \\ \boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{\sigma}) \end{pmatrix}\right] \sqrt{n}(\boldsymbol{s} - \boldsymbol{\sigma}) + o_{\mathrm{P}}(1).
\end{aligned}$$

Then, $nT\left(\hat{\boldsymbol{\theta}}\right)$ converges in distribution to a weighted sum of independent Chi-square random variables with 1 degree of freedom, where the weights are the eigenvalues of (16). □

Proof of Theorem 1. Theorem 2 in Styan (1970) indicates that the quadratic form is Chi-square distributed if and only if

$$\left(\boldsymbol{Q}\boldsymbol{Q}^T\right) \boldsymbol{G}\left(\boldsymbol{Q}\boldsymbol{Q}^T\right) \boldsymbol{G}\left(\boldsymbol{Q}\boldsymbol{Q}^T\right) = \left(\boldsymbol{Q}\boldsymbol{Q}^T\right) \boldsymbol{G}\left(\boldsymbol{Q}\boldsymbol{Q}^T\right), \tag{29}$$

$$\text{and} \quad \operatorname{rank}\left\{\left(\boldsymbol{Q}\boldsymbol{Q}^T\right) \boldsymbol{G}\left(\boldsymbol{Q}\boldsymbol{Q}^T\right)\right\} = \operatorname{tr}\left\{\boldsymbol{G}\left(\boldsymbol{Q}\boldsymbol{Q}^T\right)\right\}, \tag{30}$$

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

where rank {} and tr {} denote the rank and the trace of the enclosed matrix, respectively. Since $\boldsymbol{G}$ is the Moore–Penrose inverse of $\boldsymbol{Q}\boldsymbol{Q}^T$, condition (29) is naturally satisfied. Hence, it suffices to show condition (30). It can be easily shown that $\boldsymbol{G}\left(\boldsymbol{Q}\boldsymbol{Q}^T\right)$ is an idempotent matrix when $\boldsymbol{G}$ is a Moore–Penrose inverse. Hence,

$$\text{tr}\left\{\boldsymbol{G}\left(\boldsymbol{Q}\boldsymbol{Q}^T\right)\right\} = \text{rank}\left\{\boldsymbol{G}\left(\boldsymbol{Q}\boldsymbol{Q}^T\right)\right\} = \text{rank}\left\{\boldsymbol{Q}\boldsymbol{Q}^T\right\},$$

where the second equality holds because of Corollary 9.3.8 in Harville (1997). Note that, as the Moore–Penrose inverse,

$$\text{rank}\left\{\left(\boldsymbol{Q}\boldsymbol{Q}^T\right)\boldsymbol{G}\left(\boldsymbol{Q}\boldsymbol{Q}^T\right)\right\} = \text{rank}\left\{\boldsymbol{Q}\boldsymbol{Q}^T\right\}.$$

Hence, condition (30) is also satisfied. Further, as a gram matrix,

$$\text{rank}\left\{\boldsymbol{Q}\boldsymbol{Q}^T\right\} = \text{rank}\left\{\boldsymbol{Q}\right\}.$$

Recall that $\boldsymbol{Q}$ is idempotent, then the degrees of freedom can be simplified to

$$\text{tr}\left\{\boldsymbol{G}\left(\boldsymbol{Q}\boldsymbol{Q}^T\right)\right\} = \text{tr}\left\{\boldsymbol{Q}\right\} = L_j - K_j.$$

That is, F is asymptotically Chi-square distributed with $L_j - K_j$ degrees of freedom. □

Proof of Theorem 2. Equation (18) indicates that

$$\sqrt{n}\hat{\varphi}_j^{-1}\boldsymbol{S}_{vv,j}^{-1/2}\left(\boldsymbol{S}_{vu,j} - \boldsymbol{S}_{vz,j}\hat{\boldsymbol{\theta}}^{(j)}\right) = \sqrt{n}\hat{\varphi}_j^{-1}\boldsymbol{S}_{vv,j}^{-1/2}\hat{\boldsymbol{\Omega}}^{1/2}\hat{\boldsymbol{Q}}\hat{\boldsymbol{\Omega}}^{-1/2}\left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j}\boldsymbol{\theta}^{(j)}\right).$$

Hence, the quadratic form becomes

$$\begin{aligned} &n\left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j}\hat{\boldsymbol{\theta}}^{(j)}\right)^T\left(\hat{\varphi}_j^{-2}\boldsymbol{S}_{vv,j}^{-1}\right)\left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j}\hat{\boldsymbol{\theta}}^{(j)}\right) \\ &\quad = n\left[\hat{\boldsymbol{\Omega}}^{-1/2}\left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j}\boldsymbol{\theta}^{(j)}\right)\right]^T\left(\hat{\varphi}_j^{-2}\hat{\boldsymbol{Q}}^T\hat{\boldsymbol{\Omega}}^{1/2}\boldsymbol{S}_{vv,j}^{-1}\hat{\boldsymbol{\Omega}}^{1/2}\hat{\boldsymbol{Q}}\right)\left[\hat{\boldsymbol{\Omega}}^{-1/2}\left(\boldsymbol{S}_{vy,j} - \boldsymbol{S}_{vz,j}\boldsymbol{\theta}^{(j)}\right)\right], \end{aligned}$$

and its weight matrix is symmetric. Hence, the distribution (17) indicates that it converges in distribution to a weighted sum of independent Chi-square random variables with 1 degrees of freedom each and the weights are the eigenvalues of $\varphi_j^{-2}\boldsymbol{Q}^T\boldsymbol{\Omega}^{1/2}\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{\Omega}^{1/2}\boldsymbol{Q}$. Finally, plugging in the expression of $\boldsymbol{Q}$ into the weights yields the expression of $\boldsymbol{\Pi}$ in the theorem. □

Proof of Corollary 1. For any $\boldsymbol{\sigma}$ values, the expression of $\boldsymbol{\Pi}_j$ directly indicates

$$\varphi_j^2\text{tr}\left\{\boldsymbol{\Pi}_j\right\} = \text{tr}\left\{\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{P}_j\left(\boldsymbol{\sigma}\right)\boldsymbol{\Omega}_j\right\} \text{ and } \varphi_j^4\text{tr}\left\{\boldsymbol{\Pi}_j^2\right\} = \text{tr}\left\{\left(\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{P}_j\left(\boldsymbol{\sigma}\right)\boldsymbol{\Omega}_j\right)^2\right\},$$

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

where $\boldsymbol{P}_j(\boldsymbol{\sigma}) = \boldsymbol{I} - \boldsymbol{\Sigma}_{vz,j}\left(\boldsymbol{\Sigma}_{vz,j}^T\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{\Sigma}_{vz,j}\right)^{-1}\boldsymbol{\Sigma}_{vz,j}^T\boldsymbol{\Sigma}_{vv,j}^{-1}$. The chain rule yields

$$\begin{aligned}\frac{\partial \boldsymbol{h}_j(\boldsymbol{\sigma})}{\partial \sigma_i} = \boldsymbol{P}_j(\boldsymbol{\sigma})\left[\frac{\partial \boldsymbol{\Sigma}_{vy,j}}{\partial \sigma_i} - \frac{\partial \boldsymbol{\Sigma}_{vz,j}}{\partial \sigma_i}\boldsymbol{\gamma}_j(\boldsymbol{\sigma})\right] \\ -\boldsymbol{\Sigma}_{vz,j}\left(\boldsymbol{\Sigma}_{vz,j}^T\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{\Sigma}_{vz,j}\right)^{-1}\frac{\partial \boldsymbol{\Sigma}_{vz,j}^T\boldsymbol{\Sigma}_{vv,j}^{-1}}{\partial \sigma_i}\boldsymbol{h}_j(\boldsymbol{\sigma}),\end{aligned} \tag{31}$$

where σ_i is the ith entry in $\boldsymbol{\sigma}$. If the model is correctly specified and all MIIVs are valid, then $\boldsymbol{\gamma}_j(\boldsymbol{\sigma}) = \boldsymbol{\theta}_1^{(j)}$ and $\boldsymbol{h}_j(\boldsymbol{\sigma}) = \boldsymbol{0}$. Hence, we obtain

$$\frac{\partial \boldsymbol{h}_j(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}^T} = \boldsymbol{P}_j(\boldsymbol{\sigma})\frac{\partial \boldsymbol{g}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}^T}.$$

Consequently,

$$\begin{aligned}\varphi_j^2 \mathrm{tr}\left\{\boldsymbol{\Delta}_j\right\} &= \mathrm{tr}\left\{\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{P}_j(\boldsymbol{\sigma})\boldsymbol{\Omega}_j\boldsymbol{P}_j^T(\boldsymbol{\sigma})\right\} \\ &= \mathrm{tr}\left\{\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{P}_j(\boldsymbol{\sigma})\boldsymbol{\Omega}_j\right\}, \\ \varphi_j^4 \mathrm{tr}\left\{\boldsymbol{\Delta}_j^2\right\} &= \mathrm{tr}\left\{\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{P}_j(\boldsymbol{\sigma})\boldsymbol{\Omega}_j\boldsymbol{P}_j^T(\boldsymbol{\sigma})\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{P}_j(\boldsymbol{\sigma})\boldsymbol{\Omega}_j\boldsymbol{P}_j^T(\boldsymbol{\sigma})\right\} \\ &= \mathrm{tr}\left\{\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{P}_j(\boldsymbol{\sigma})\boldsymbol{\Omega}_j\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{P}_j(\boldsymbol{\sigma})\boldsymbol{\Omega}_j\right\},\end{aligned}$$

where the last equality holds since $\boldsymbol{P}_j^T(\boldsymbol{\sigma})\boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{P}_j(\boldsymbol{\sigma}) = \boldsymbol{\Sigma}_{vv,j}^{-1}\boldsymbol{P}_j(\boldsymbol{\sigma})$. Therefore, $\mathrm{tr}\left\{\boldsymbol{\Pi}_j\right\} = \mathrm{tr}\left\{\boldsymbol{\Delta}_j\right\}$ and $\mathrm{tr}\left\{\boldsymbol{\Pi}_j^2\right\} = \mathrm{tr}\left\{\boldsymbol{\Delta}_j^2\right\}$.

On the other hand, if the population $\boldsymbol{\Sigma}$ is estimated by $\boldsymbol{S}$ and $\boldsymbol{\theta}_1^{(j)}$ by $\hat{\boldsymbol{\theta}}_1^{(j)}$, then $\boldsymbol{h}_j(\boldsymbol{s})$ is not necessarily $\boldsymbol{0}$. From equation 31, the difference between $\mathrm{tr}\left\{\hat{\boldsymbol{\Delta}}\right\}$ and $\mathrm{tr}\left\{\hat{\boldsymbol{\Pi}}\right\}$ and the difference between $\mathrm{tr}\left\{\hat{\boldsymbol{\Delta}}^2\right\}$ and $\mathrm{tr}\left\{\hat{\boldsymbol{\Pi}}^2\right\}$ depends on $\boldsymbol{h}_j(\boldsymbol{s})$. □

Acknowledgments

We would like to thank the editor and the anonymous reviewers for their constructive suggestions to improve the article.

Funding Open access funding provided by Uppsala University. Jin and Yang-Wallentin gratefully acknowledge partial support of this research from the Swedish Research Council award 2017-01175. Bollen gratefully acknowledges partial support of this research from the National Institutes of Health award 1R21MH119572-01, P2CHD050924, and P30 AG066615.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Andrews, D. W. K. (1987). Asymptotic results for generalized wald tests. *Econometric Theory*, *3*, 348–358.

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

Bollen, K. A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, *61*, 109–121.

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* (pp. 119–138). Lincolnwood, IL: Scientific Software.

Bollen, K. A. (2012). Instrumental variables in sociology and the social sciences. *Annual Review of Sociology*, *38*(1), 37–72.

Bollen, K. A. (2019). Model implied instrumental variables (MIIVs): An alternative orientation to structural equation modeling. *Multivariate Behavioral Research*, *54*, 31–46.

Bollen, K. A. (2020). When good loadings go bad: Robustness in factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*, 515–524.

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

Bollen, K. A., Gates, K. M., & Fisher, Z. (2018a). Robustness conditions for MIIV-2SLS when the latent variable or measurement model is structurally misspecified. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 848–859.

Bollen, K. A., Gates, K. M., & Fisher, Z. (2018b). Robustness conditions for MIIV-2SLS when the latent variable or measurement model is structurally misspecified. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 848–859.

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 & Research*, *48*, 48–86.

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*, 20–50.

Bollen, K. A., & Maydeu-Olivares, A. (2007). A polychoric instrumental variable (PIV) estimator for structural equation models with categorical variables. *Psychometrika*, *72*, 309–326.

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *Journal of the American Statistical Association*, *90*, 443–450.

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

Browne, M. W. (1987). Robustness of statistical inference in factor analysis and related models. *Biometrika*, *74*, 375–384.

Fisher, Z., Bollen, K., Gates, K., & Rönkkö, M. (2017). *Model Implied Instrumental Variable (MIIV) Estimation of Structural Equation Models*. R package version 0.5.3.

Fisher, Z. F., & Bollen, K. A. (2020). An instrumental variable estimator for mixed indicators: Analytic derivatives and alternative parametrizations. *Psychometrika, 85*(3), 660–683. https://doi.org/10.1007/s11336-020-09721-6.

Fisher, Z. F., Bollen, K. A., & Gates, K. M. (2019). A limited information estimator for dynamic factor models. *Multivariate Behavioral Research, 54*(2), 246–263. https://doi.org/10.1080/00273171.2018.1519406.

Greene, W. H. (2002). *Econometric Analysis* (5th ed.). Upper Saddle River, New Jersey: Prentice Hall.

Hahn, J., & Hausman, J. (2002). A new specification test for the validity of instrumental variables. *Econometrica*, *70*, 163–189.

Hahn, J., & Hausman, J. (2003). Weak instruments: Diagnosis and cures in empirical econometrics. *The American Economic Review*, *93*, 118–125.

Hahn, J., & Hausman, J. (2005). Estimation with valid and invalid instruments. *Annales d'Economie et de Statistique*, *79*, 25–57.

Harville, D. A. (1997). *Matrix Algebra from a Stastatistic's Perspective*. New York: Springer.

Jin, S., & Cao, C. (2018). Selecting polychoric instrumental variables in confirmatory factor analysis: An alternative specification test and effects of instrumental variables. *British Journal of Mathematical and Statistical Psychology*, *71*, 387–413.

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

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, *59*, 381–389.

Kirby, J. B., & Bollen, K. A. (2009). Using instrumental variable (IV) tests to evaluate model specification in latent variable structural equation models. *Sociological Methodology*, *39*, 327–355.

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*, 97–126.
Li, C.-H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, *21*, 369–387.
MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, *100*, 107–120.
MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490–504.
Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, *43*, 551–560.
Muthén, B. (1984). A general structural equation model with dichotomous ordered categorical and continuous latent indicators. *Psychometrika*, *49*, 115–132.
Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Retrieved from https://www.statmodel.com/download/Article_075.pdf. Accessed by 2013-09-12.
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*, 127–143.
Nestler, S. (2014). How the 2SLS/IV estimator can handle equality constraints in structural equation models: A system-of-equations approach. *British Journal of Mathematical and Statistical Psychology*, *67*, 353–369.
Nestler, S. (2015a). A specification error test that uses instrumental variables to detect latent quadratic and latent interaction effects. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*, 542–551.
Nestler, S. (2015b). Using instrumental variables to estimate the parameters in unconditional and conditional second-order latent growth models. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*, 461–473.
Olea, J. L. M., & Pflueger, C. E. (2013). A robust test for weak instruments. *Journal of Business and Economic Statistics*, *31*, 358–369.
Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*, 443–460.
Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. *Psychometrika*, *47*, 337–347.
R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
Reisenzein, R. (1986). A structural equation analysis of Weiner's attribution-affect model of helping behavior. *Journal of Personality and Social Psychology*, *50*, 1123–1133.
Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.
Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, *26*, 393–415.
Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. V. Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks: SAGE Publications.
Shea, J. (1997). Instrument relevance in multivariate linear models: A simple measure. *The Review of Economics and Statistics*, *79*, 348–352.
Sörbom, D. (1989). Model modification. *Psychometrika*, *54*, 371–384.
Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and inference for econometric models: essays in honor of Thomas J. Rothenberg* (pp. 80–108). New York: Cambridge University Press.
Styan, G. P. H. (1970). Notes on the distribution of quadratic forms in singular normal variables. *Biometrika*, *57*, 567–572.
Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*, 393–408.
Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*, 392–423.

Manuscript Received: 17 JAN 2020
Final Version Received: 23 MAR 2021
Accepted: 18 MAY 2021

https://doi.org/10.1007/s11336-021-09771-4 Published online by Cambridge University Press