🔓 OPEN ACCESS

On Inconsistency of the Overidentification Test for the Model-implied Instrumental Variable Approach

Shaobo Jin

Uppsala University

ABSTRACT

In the context of structural equation modeling, the model-implied instrumental variable (MIIV) approach has been shown to be more robust against model misspecification than the systemwide approaches (e.g., maximum likelihood and least squares). Besides the goodness-of-fit tests that test the fit of the entire hypothesized covariance structure, the overidentification tests for MIIV can be used to test model specification on an equation-by-equation basis. However, it is known in the econometrics literature that the overidentification tests are inconsistent against general misspecification, if it is used to test a zero correlation between the instrumental variables and the error terms. In this paper, we show that such inconsistency can also occur for the MIIV approach. Numerical examples where the powers of the tests converge to the size are presented. Theoretical results are proved to support the numerical findings. Implications on when the overidentification tests are consistent are also presented.

1. Introduction

Normal-theory maximum likelihood and least squares are commonly used to fit a structural equation modeling (SEM) model. Both approaches are systemwide approaches in the sense that all parameters are estimated simultaneously, provided that the sample covariance matrix is estimated in an earlier stage. However, various studies (e.g., Bollen, 1996; Jin et al., 2016; Nestler, 2013; Yang-Wallentin et al., 2010) have shown that the systemwide estimators can be largely biased in a misspecified model. Throughout the paper, we refer to model misspecification as nonzero parameters are mistakenly fixed to zero, such as omitted factor loading, latent regression path or error covariances.

An alternative to the systemwide approaches is the model-implied instrumental variable (MIIV) approach. Some examples of the MIIV approach are the two-stage least squares (2SLS, Bollen, 1996) for continuous indicators and the polychoric instrumental variable (PIV, Bollen & Maydeu-Olivares, 2007) approach for ordinal indicators. They are implemented in the R package MIIVsem (Fisher et al., 2017). The idea of the MIIV approach has been applied 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. Recently, Fisher and Bollen (2020) and Jin et al. (2021) extended the MIIV approach to the mixture of continuous and categorical indicators.

An advantage of the MIIV approach is its robustness against certain model misspecification. From the simulation perspective, various studies (e.g., Bollen et al., 2007; Jin et al., 2016; Nestler, 2013) showed that the MIIV estimator is as accurate as the systemwide estimators in the correctly specified models and is more accurate in the misspecified models with omitted parameters. From the theoretical perspective, Bollen et al. (2018) and Bollen (2020) presented the conditions under which the MIIV estimator remains asymptotically unbiased in the misspecified models.

Another advantage of the MIIV approach is that the overidentification tests can be applied to investigate model specification in an equation-by-equation manner, without the need of fitting the whole model. Kirby and Bollen (2009) investigated the Sargan (1958) test when all indicators are continuous. Jin and Cao (2018) proposed alternative overidentification tests for ordinal variables. Jin et al. (2021) further generalized the tests to a mixture of different types of indicators. All these studies suggested to use as many MIIVs as possible when performing the overidentification tests. In particular, the simulation study in Jin and Cao (2018) showed that the overidentification test can have a very low power when the degrees of freedom of the overidentification test is one.

The current study is concerned with a pitfall of the overidentification tests. In many econometric textbooks, the null hypothesis of the overidentification tests is often formulated as a nonzero correlation between the instrumental variables (IVs) and the error term (e.g., Cameron & Trivedi, 2005; Wooldridge, 2002, 2013). Even Sargan (1958, p. 404) formulated the null hypothesis as "there is a relationship between the suggested variables with a residual independent of all the instrumental variables." However, Newey (1985) showed

that all chi-square overidentification tests are inconsistent against general misspecification. In other words, there exist examples where the IVs are correlated with the error term but the power of the chi-square overidentification test converges to the size of the test, instead of one. Stock (2015) stated that the maintained hypothesis behind the overidentification tests is that there exist enough exogenous IVs. When the test statistic is invariant to which IVs to be included in the maintained hypothesis, the test is often used as an omnibus test for failure in exogeneity. Parente and Santos Silva (2012) further suggested that it is more appropriate to interpret the overidentification tests as tests for whether the IVs can identify the same parameter.

The main contribution of our study is to investigate implications of inconsistent overidentification tests in the context of SEM. The result that the overidentification tests are inconsistent tests against endogeneity of IVs is not entirely new in the econometrics literature. Various studies (e.g., De Blander, 2008; Guggenberger, 2012; Parente & Santos Silva, 2012) have investigated the hypotheses under which the overidentification tests remain consistent. However, the IVs in econometric studies are mostly external, whereas the IVs in the MIIV approach are internal since they are searched from the indicators of the hypothesized SEM model. By utilizing the model structure, we can draw more insightful conclusions. To be more specific, the main purpose is two-fold. First, our study serves as a cautionary note that the overidentification tests developed for the MIIV approach can also be inconsistent against endogeneity of IVs. Second, we will derive conditions under which the overidentification tests remain consistent against endogeneity. These conditions will help researchers to understand the risks of using overidentification tests to test the validity of IVs when the MIIV approach is used. In the context of confirmatory factor analysis (CFA), Table 4 in Jin and Cao (2018) has shown that the power can converge to the size. However, they did not investigate the implications behind.

The rest of the paper is organized as follows. We first briefly present the SEM model and the MIIV idea. Second, we conduct a Monte Carlo simulation to illustrate inconsistency of the overidentification tests. Then we provide theoretical results supporting our findings in the simulation study and present the implications. A discussion and conclusion section ends the paper.

2. The MIIV Approach

Consider the SEM model

$$y^* = \Lambda \eta + \varepsilon, \tag{1}$$
$$\eta = B\eta + \zeta, \tag{2}$$

where y^* is the vector of continuous variables, Λ is the matrix of factor loadings, η is the vector of latent variables, B is the matrix of latent regression coefficients, ε is the disturbance of the measurement model with $\text{var}(\varepsilon) = \Theta$, and ζ is disturbance of the latent regression model with $\text{var}(\zeta) = \Psi$. Both ε and ζ are homoscedastic in the model. We follow

Jin et al. (2021) and assume that $E(y^*) = 0$. Instead of observing y^*, the observed vector is denoted by y. If the jth element of y, denoted by y_j, is continuous, then $y_j = y_j^*$, where y_j^* is the jth element in y^*. If y_j is ordinal, then y_j^* is the underlying continuous variable and y_j is obtained by discretizing y_j^* based on some threshold values.

Throughout the paper, we always assume that η is independent of ε and that ζ is independent of ε. Depending on y, different distributional assumptions are needed. If y only consists of observed continuous variables, we assume that y^* has finite moments up to the fourth order. If some elements of y are ordinal, we assume that the corresponding underlying continuous variables are normal. These assumptions are similar to those in Jin et al. (2021).

In the MIIV approach, the scale of η is set by the scaling indicators, namely, choosing one indicator per latent variable and setting the factor loading to one. Suppose that y^* is partitioned into y_1^* and y_2^* such that $y_1^* = \eta + \varepsilon_1$ is used for identification and $y_2^* = \Lambda_2 \eta + \varepsilon_2$ contains unknown factor loadings. If we substitute $y_1^* - \varepsilon_1$ for η in the SEM model, we obtain a regression system with observed variables

$$\begin{pmatrix} y_2^* \\ y_1^* \end{pmatrix} = \begin{pmatrix} \Lambda_2 \\ B \end{pmatrix} y_1^* + \begin{pmatrix} \varepsilon_2 - \Lambda_2 \varepsilon_1 \\ \zeta + (I - B)\varepsilon_1 \end{pmatrix}, \tag{3}$$

which is referred to as the Latent-to-Observed (L2O) variable transformation by Bollen (2019). Equation (3) indicates that we can partition the parameter vector θ into two vectors: regression coefficient θ_1 (free parameters in Λ_2 and B) and dispersion parameter θ_2 (free parameters in Ψ and Θ). That is, $\theta^T = (\theta_1^T, \theta_2^T)$.

In most cases, y_1^* is correlated with the composite error term. Hence, IV regression is used to estimate the parameters in θ_1 in a row-by-row manner (Bollen, 1996; Bollen & Maydeu-Olivares, 2007). Suppose that the jth row of the system (Equation (3)) is

$$y_j^* = z_j^{*T} \theta_1^{(j)} + e_j, \tag{4}$$

where $\theta_1^{(j)}$ is the vector of unknown parameters and z_j^* is subset of y_1^* that is associated with $\theta_1^{(j)}$. For notational simplicity, we will suppress j when no confusion should arise.

The following assumptions are imposed in order to use IV regression. First, a non-empty subset of z^* is correlated with e. Second, there exists a set of IVs, denoted by v^*, which is correlated with z^* and uncorrelated with e. Third, the number of elements in v^*, denoted by L, is no lower than the number of elements in z^*, denoted by K. In the MIIV approach, the IVs are selected from y^*. The reader is directed to Bollen (2019) for a detailed introduction of the L2O transformation and the algorithm of searching MIIVs.

Let Σ be the population covariance matrix of y^* and S be its sample counterpart. We always assume that s is a consistent estimator of σ such that

$$\sqrt{n}(s - \sigma) \xrightarrow{d} N(0, \Upsilon), \tag{5}$$

where n is the sample size, s is the vector of free elements in S, σ is the vector of free elements in Σ, and Υ is the asymptotic covariance matrix. Let

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

where $\boldsymbol{\Sigma}_{vz} = \text{cov}(\boldsymbol{v}^{*}, \boldsymbol{z}^{*})$, $\boldsymbol{\Sigma}_{vv} = \text{var}(\boldsymbol{v}^{*})$, and $\boldsymbol{\Sigma}_{vy} = \text{cov}(\boldsymbol{v}^{*}, \boldsymbol{y}^{*})$. Then, the MIIV estimator of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = \gamma(\boldsymbol{s})$. Because of the consistency of $\boldsymbol{s}$, $\gamma(\boldsymbol{s})$ is a consistent estimator of $\gamma(\boldsymbol{\sigma})$. It is also a consistent estimator of the true $\boldsymbol{\theta}$, if Equation (4) is correctly specified, the MIIVs are valid, and both $\boldsymbol{\Sigma}_{vz}^{T}\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{vz}$ and $\boldsymbol{S}_{vz}^{T}\boldsymbol{S}_{vv}^{-1}\boldsymbol{S}_{vz}$ are invertible. However, if some MIIVs are invalid, then $\gamma(\boldsymbol{\sigma})$ is not necessarily equal to $\boldsymbol{\theta}$.

When all observed variables are continuous, Kirby and Bollen (2009) proposed to use the Sargan (1958) chi-square statistic

$$F_{Sargan} = \frac{n}{\hat{\varphi}^{2}}(\boldsymbol{S}_{vy} - \boldsymbol{S}_{vz}\hat{\boldsymbol{\theta}})^{T}\boldsymbol{S}_{vv}^{-1}(\boldsymbol{S}_{vy} - \boldsymbol{S}_{vz}\hat{\boldsymbol{\theta}}),$$

to test the specification of the jth equation, where $\hat{\varphi}^{2} = S_{yy} - 2\hat{\boldsymbol{\theta}}^{T}\boldsymbol{S}_{zy} + \hat{\boldsymbol{\theta}}^{T}\boldsymbol{S}_{zz}\hat{\boldsymbol{\theta}}$. When all observed variables are ordinal, Jin and Cao (2018) showed that the asymptotic distribution of F_{Sargan} is a weighted sum of chi-square distributions with one degree of freedom and proposed to apply the Satorra and Bentler (1994) adjustments to F_{Sargan}. They also proposed an asymptotic chi-square distributed statistic. Recently, Jin et al. (2021) proposed the overidentification tests where the observed variables can be a mixture of different types. When Equation (4) is correctly specified and all MIIVs are valid, they showed that

$$F = n(\boldsymbol{S}_{vy} - \boldsymbol{S}_{vz}\hat{\boldsymbol{\theta}})^{T}\hat{\boldsymbol{\Omega}}^{-1/2}(\boldsymbol{QQ}^{T})^{-}\hat{\boldsymbol{\Omega}}^{-1/2}(\boldsymbol{S}_{vy} - \boldsymbol{S}_{vz}\hat{\boldsymbol{\theta}})$$

is asymptotically chi-square distributed with $L-K>0$ degrees of freedom, where

$$\boldsymbol{\Omega} = \boldsymbol{C}(\boldsymbol{\sigma})\boldsymbol{\Upsilon}\boldsymbol{C}^{T}(\boldsymbol{\sigma}),$$
$$\boldsymbol{Q} = \boldsymbol{I} - \boldsymbol{\Omega}^{-1/2}\boldsymbol{\Sigma}_{vz}(\boldsymbol{\Sigma}_{vz}^{T}\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{vz})^{-1}\boldsymbol{\Sigma}_{vz}^{T}\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Omega}^{1/2},$$

and $(\boldsymbol{QQ}^{T})^{-}$ is the Moore–Penrose inverses of $\boldsymbol{QQ}^{T}$, with the ith column in $\boldsymbol{C}(\boldsymbol{\sigma})$ being

$$\frac{\partial \boldsymbol{\Sigma}_{vy}}{\partial \sigma_{i}} - \frac{\partial \boldsymbol{\Sigma}_{vz}}{\partial \sigma_{i}}\gamma(\boldsymbol{\sigma})$$

and σ_{i} being the ith element in $\boldsymbol{\sigma}$. Under the same conditions, Jin et al. (2021) showed that the asymptotic distribution of F_{Sargan} is a weighted sum of chi-square distributions with one degree of freedom. The weights are the eigenvalues of

$$\boldsymbol{\Pi} = \varphi^{-2}\boldsymbol{\Omega}^{1/2}[\boldsymbol{\Sigma}_{vv}^{-1} - \boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{vz}(\boldsymbol{\Sigma}_{vz}^{T}\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{vz})^{-1}\boldsymbol{\Sigma}_{vz}^{T}\boldsymbol{\Sigma}_{vv}^{-1}]\boldsymbol{\Omega}^{1/2}.$$

Hence, they also proposed the mean-scaled statistic and the mean-variance adjusted statistic as

$$F_{m} = \frac{L-K}{\text{tr}\{\hat{\boldsymbol{\Pi}}\}}F_{Sargan} \text{ and } F_{mv} = \frac{\text{tr}\{\hat{\boldsymbol{\Pi}}\}}{\text{tr}\{\hat{\boldsymbol{\Pi}}^{2}\}}F_{Sargan},$$

respectively. As proved by Jin et al. (2021), if all indicators are ordinal, these tests are asymptotically equivalent to the tests in Jin and Cao (2018).

3. Numerical Examples

Before we present the theoretical results, we will present two examples where different misspecifications remain undetected even if the MIIVs are correlated with the composite error term. The MIIVs are always searched from the hypothesized model, instead of the true data generating process (DGP). The first example considers a four-factor SEM model (see Figure 1), and the second example considers a three-factor CFA model (see Figure 4). The population values are regarded as realistic values by Li (2016). In both examples, we consider a mixture of continuous indicators and five-category Likert-scale indicators. The probabilities of belonging to each category are 0.04, 0.05, 0.21, 0.46, and 0.24, which is the slightly asymmetric setting in Li (2016). The variance of the continuous indicators and the underlying continuous variables are set to 1. Five sample sizes are considered, i.e., $n = 300$, 600, 1,500, 2,500, and 3,600. The number of replications is 10, 000 for each sample size and model misspecification. The significance level is set to 0.05. If the overidentification tests are consistent against the
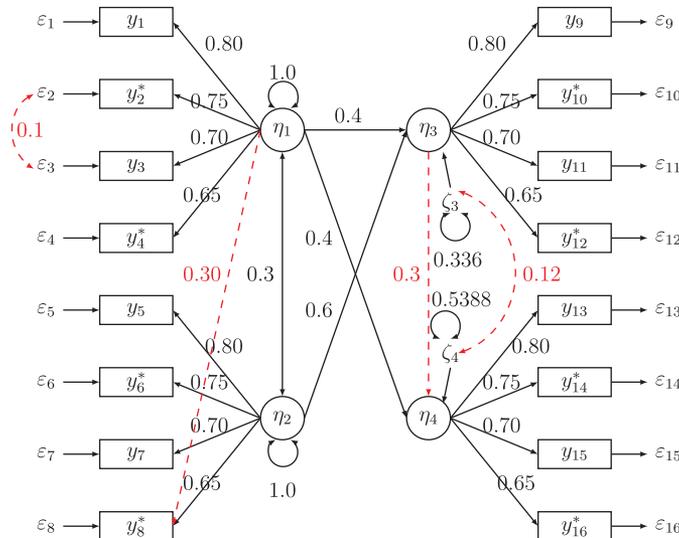

Figure 1. Path diagram of a structural equation model with standardized parameter values. One dashed line is omitted in each misspecified model, yielding four misspecified models. The starred indicators are ordinal, whereas the others are continuous. The scaling indicators are always y_1, y_5, y_9, and y_{13}.

misspecified models, we expect the empirical proportion of rejection to converge toward 1.

Example 1: SEM

Data are generated from Figure 1 with all paths. Four misspecified models are considered: each model drops one dashed line in Figure 1. Consequently, we can investigate the models with omitted cross loading, latent regression path, measurement model error correlation, or latent regression error correlation. As such, this example aims to show that inconsistency of tests can occur for any misspecification. The simulation results in Jin and Cao (2018) showed that the overidentification test with one degree of freedom is inconsistent against omitted factor loadings. Our example here also aims to show that such inconsistency can also occur for other degrees of freedom.

If λ_{81} is omitted when estimating λ_{82} (Figure 2a), the equation to be estimated can be expressed as

$$y_8^* = \lambda_{82} y_5^* + \varepsilon_8 - \lambda_{82}\varepsilon_5 + \lambda_{81}\eta_1.$$

In the hypothesized model where $\lambda_{81} = 0$, all indicators expect y_5^* and y_8^* are not correlated with the hypothesized composite error $\varepsilon_8 - \lambda_{82}\varepsilon_5$. Hence, they will be selected as MIIVs, when we search MIIVs from the hypothesized model. However, due to the correlations among η in the true model, all indicators will be correlated with the actual composite error $\varepsilon_8 - \lambda_{82}\varepsilon_5 + \lambda_{81}\eta_1$. Hence, all indicators are
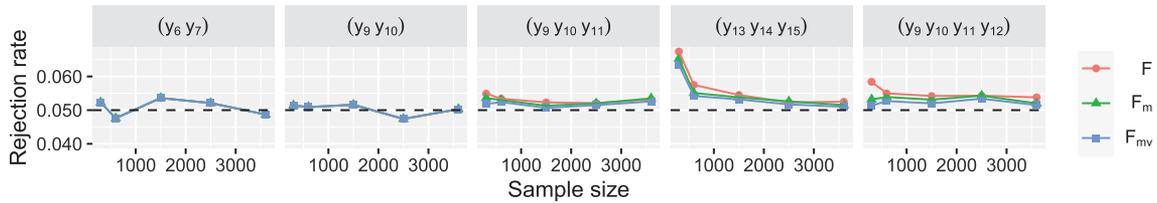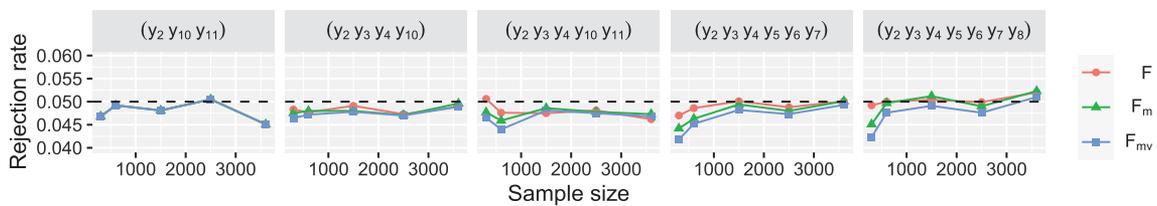

(a) $y_8^* = \lambda_{81}\eta_1 + \lambda_{82}\eta_2 + \varepsilon_8$ in the true model but $\lambda_{81} = 0$ in the hypothesized model
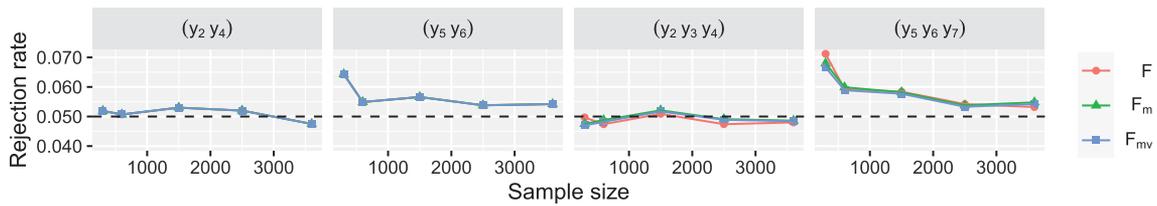
(b) $\eta_4 = b_1\eta_1 + b_2\eta_3 + \zeta_4$ in the true model but $b_2 = 0$ in the hypothesized model
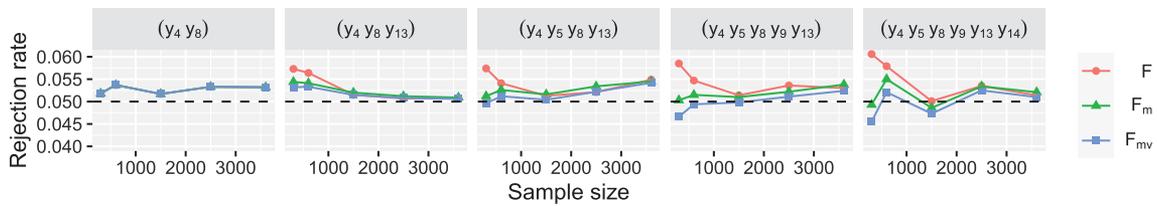
(c) $y_2^* = \lambda_{21}\eta_1 + \varepsilon_2$ and $\mathrm{cor}(\varepsilon_2, \varepsilon_3) \neq 0$ in the true model but $\mathrm{cor}(\varepsilon_2, \varepsilon_3) = 0$ in the hypothesized model

(d) $\eta_4 = b_1\eta_1 + b_2\eta_3 + \zeta_4$ and $\mathrm{cor}(\zeta_3, \zeta_4) \neq 0$ in the true model but $\mathrm{cor}(\zeta_3, \zeta_4) = 0$ in the hypothesized model

Figure 2. Empirical rejection rate of the inconsistent overidentification tests. The MIIVs used for the overidentification tests are stated in the figure title.

invalid MIIVs. Likewise, if b_2 is omitted when estimating b_1 (Figure 2b), η_3 is included in the composite error, hence all MIIVs obtained from the hypothesized model with $b_2 = 0$ are invalid MIIVs. If $\text{cor}(\varepsilon_2, \varepsilon_3)$ is omitted when estimating λ_{21} (Figure 2c), y_3^* to y_{16}^* are selected as MIIVs. However, y_3^* is an invalid MIIV, due to a nonzero $\text{cor}(\varepsilon_2, \varepsilon_3)$. If $\text{cor}(\zeta_3, \zeta_4)$ is omitted when estimating b_1 and b_2 (Figure 2d), y_2^* to y_8^* and y_{10}^* to y_{12}^* are selected as MIIVs. However, y_{10}^* to y_{12}^* are invalid, since they rely on η_4 that is correlated with ζ_4 through ζ_3.

It is seen from Figure 2 that the rejection rates approach 0.05 for various combinations of MIIVs in the misspecified model. Such a low power can be caused by a zero correlation between the MIIVs and the composite error (Figure 2c and the last two plots in Figure 2d). In such a case, some MIIVs remain valid, even though the SEM model is misspecified. As expected, using only valid MIIVs will not detect misspecification. However, including invalid MIIVs does not guaranteed that the omitted parameters can be detected (Figure 2a,b and the first three plots in Figure 2d). It is seen from Figure 3 that the overidentification tests can be consistent if other sets of MIIVs are used. In particular, if all MIIVs are used, then the tests are still consistent.
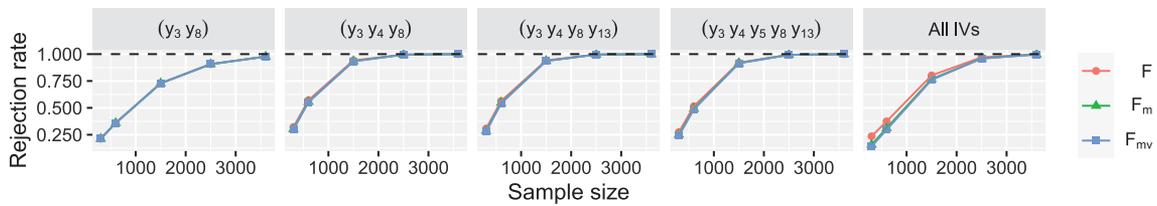
Example 2: CFA

Our second example aims to illustrate a case where using all MIIVs still yields an inconsistent test. Data are generated
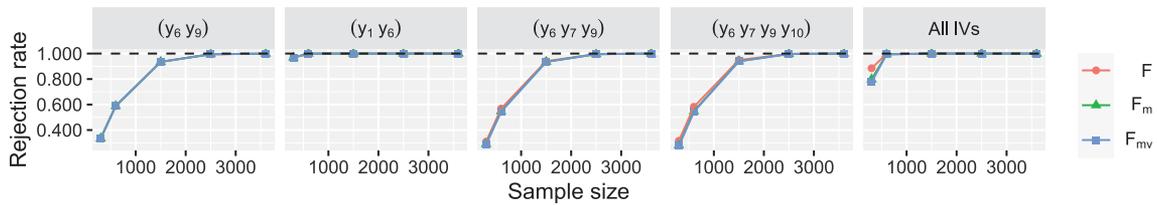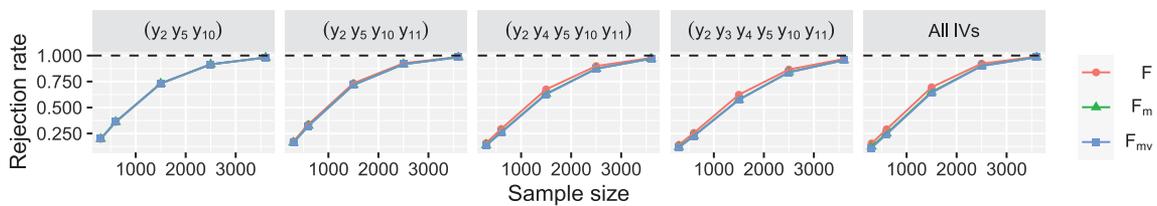

(a) $y_8^* = \lambda_{81}\eta_1 + \lambda_{82}\eta_2 + \varepsilon_8$ in the true model but $\lambda_{81} = 0$ in the hypothesized model
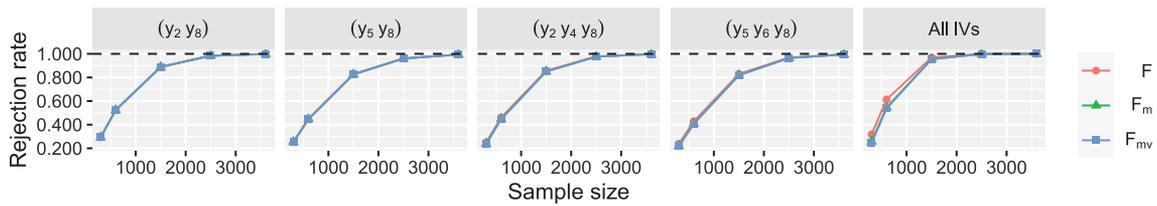
(b) $\eta_4 = b_1\eta_1 + b_2\eta_3 + \zeta_4$ in the true model but $b_2 = 0$ in the hypothesized model

(c) $y_2^* = \lambda_{21}\eta_1 + \varepsilon_2$ and $\text{cor}(\varepsilon_2, \varepsilon_3) \neq 0$ in the true model but $\text{cor}(\varepsilon_2, \varepsilon_3) = 0$ in the hypothesized model

(d) $\eta_4 = b_1\eta_1 + b_2\eta_3 + \zeta_4$ and $\text{cor}(\zeta_3, \zeta_4) \neq 0$ in the true model but $\text{cor}(\zeta_3, \zeta_4) = 0$ in the hypothesized model

Figure 3. Empirical rejection rate of the consistent overidentification tests. The MIIVs used for the overidentification tests are stated in the figure title.
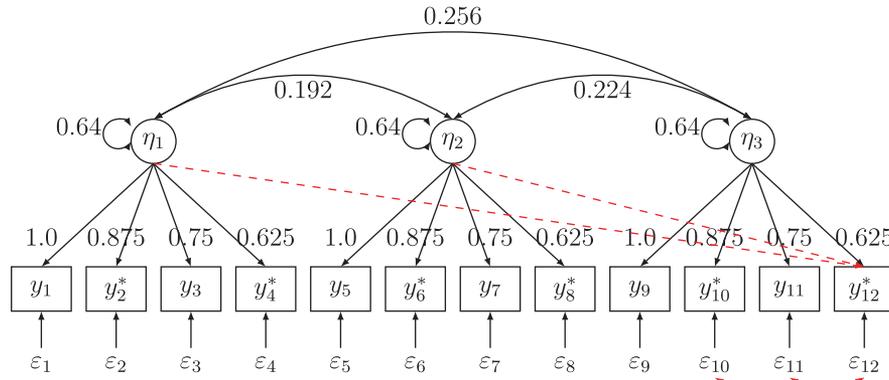
4. Inconsistency of Overidentification Tests

The starting point of IV regression to estimate $\boldsymbol{\theta}_1$ in Equation (4) is the moment condition

$$E[\boldsymbol{v}^*(y^* - \boldsymbol{z}^{*T}\boldsymbol{\theta}_1)] = \boldsymbol{0}. \qquad (8)$$

This condition holds if $\boldsymbol{v}^*$ is not correlated with the composite error term. The overidentification tests are often described as testing the validity of Equation (8). Since the IVs in the MIIV approach are searched from the indicators in the hypothesized model, violating Equation (8) means that the model is misspecified. Our numerical examples above show that misspecification can remain undetected even when the MIIVs are invalid. In fact, the result that the overidentification tests are not consistent against misspecification is not entirely new, especially in the econometrics literature. Various studies such as De Blander (2008), Guggenberger (2012), Newey (1985), and Parente and Santos Silva (2012) are devoted to inconsistency of overidentification tests.

4.1. Newey's Result

Under the assumption that Equation (8) is locally misspecified, Newey (1985) derived the asymptotic distribution of the overidentification test. For a better discussion of the inconsistency, we apply the arguments in Newey (1985) to the MIIV approach, but without the local drift in the DGP. Our notations here by and large follow Jin et al. (2021).

Let $\boldsymbol{h}(\boldsymbol{\sigma}) = \boldsymbol{\Sigma}_{vy} - \boldsymbol{\Sigma}_{vz}\boldsymbol{\gamma}(\boldsymbol{\sigma})$. The delta method to Equation (5) indicates that

$$\sqrt{n}\boldsymbol{h}(\boldsymbol{s}) = \sqrt{n}\boldsymbol{h}(\boldsymbol{\sigma}) + \frac{\partial \boldsymbol{h}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}^T}\sqrt{n}(\boldsymbol{s} - \boldsymbol{\sigma}) + o_P(1), \qquad (9)$$

where $\boldsymbol{h}(\boldsymbol{s}) = \boldsymbol{S}_{vy} - \boldsymbol{S}_{vz}\boldsymbol{\gamma}(\boldsymbol{s})$. By the chain rule, the ith column in the partial derivative $\partial \boldsymbol{h}(\boldsymbol{\sigma})/\partial \boldsymbol{\sigma}^T$ can be computed by

$$\frac{\partial \boldsymbol{h}(\boldsymbol{\sigma})}{\partial \sigma_i} = \boldsymbol{P}(\boldsymbol{\sigma})\boldsymbol{c}_i(\boldsymbol{\sigma}) - \boldsymbol{\Sigma}_{vz}(\boldsymbol{\Sigma}_{vz}^T\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{vz})^{-1}\frac{\partial \boldsymbol{\Sigma}_{vz}^T\boldsymbol{\Sigma}_{vv}^{-1}}{\partial \sigma_i}\boldsymbol{h}(\boldsymbol{\sigma}), \quad (10)$$

where $\boldsymbol{P}(\boldsymbol{\sigma}) = \boldsymbol{I} - \boldsymbol{\Sigma}_{vz}(\boldsymbol{\Sigma}_{vz}^T\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{vz})^{-1}\boldsymbol{\Sigma}_{vz}^T\boldsymbol{\Sigma}_{vv}^{-1}$ and $\boldsymbol{c}_i(\boldsymbol{\sigma})$ is the ith column of $\boldsymbol{C}(\boldsymbol{\sigma})$ (Jin et al., 2021). Then, we obtain $\sqrt{n}\hat{\boldsymbol{Q}}\hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{h}(\boldsymbol{s}) = \sqrt{n}\hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{h}(\boldsymbol{s})$ and

$$\boldsymbol{Q}\boldsymbol{\Omega}^{-1/2}\frac{\partial \boldsymbol{h}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}^T}\sqrt{n}(\boldsymbol{s} - \boldsymbol{\sigma}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{Q}\boldsymbol{Q}^T),$$

since $\boldsymbol{Q}$ is idempotent, $\boldsymbol{Q}\boldsymbol{\Omega}^{-1/2}\boldsymbol{\Sigma}_{vz} = \boldsymbol{0}$, and

$$\boldsymbol{\Omega}^{-1/2}\boldsymbol{P}(\boldsymbol{\sigma}) = \boldsymbol{Q}\boldsymbol{\Omega}^{-1/2}. \qquad (11)$$

If $\boldsymbol{Q}\boldsymbol{\Omega}^{-1/2}\boldsymbol{h}(\boldsymbol{\sigma}) = \boldsymbol{0}$, then

$$\sqrt{n}\hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{h}(\boldsymbol{s}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{Q}\boldsymbol{Q}^T),$$

where the rank of $\boldsymbol{Q}\boldsymbol{Q}^T$ is $L - K$. If $\boldsymbol{Q}\boldsymbol{\Omega}^{-1/2}\boldsymbol{h}(\boldsymbol{\sigma}) \neq \boldsymbol{0}$, the expectation of the asymptotic distribution is nonzero. Newey (1985) assumed that $\boldsymbol{h}(\boldsymbol{\sigma}) = n^{-1/2}\boldsymbol{\delta}$, for some vector $\boldsymbol{\delta}$. Hence, the asymptotic distribution of F is a non-central chi-square distribution with $L - K$ degrees of freedom and non-central parameter $\boldsymbol{\delta}^T\boldsymbol{\Omega}^{-1/2}\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{\Omega}^{-1/2}\boldsymbol{\delta}$. The non-central parameter is zero if and only if $\boldsymbol{Q}\boldsymbol{\Omega}^{-1/2}\boldsymbol{\delta} = \boldsymbol{0}$. In other words,

$$\boldsymbol{Q}\boldsymbol{\Omega}^{-1/2}\boldsymbol{h}(\boldsymbol{\sigma}) = \boldsymbol{0} \qquad (12)$$

is a sufficient and necessary condition for the asymptotic distribution of F to be chi-square. As explained by Newey (1985), the consequence is that F is not a consistent test if the IVs are invalid but Equation (12) holds.

4.2. Revised Theorems for MIIV

Because of Equation (11) and $\boldsymbol{P}(\boldsymbol{\sigma})\boldsymbol{h}(\boldsymbol{\sigma}) = \boldsymbol{h}(\boldsymbol{\sigma})$, the condition (Equation (12)) is the same as $\boldsymbol{h}(\boldsymbol{\sigma}) = \boldsymbol{0}$ in the MIIV approach. Hence, we can revise the conditions in Theorem 1 of Jin et al. (2021) and reach the following theorem. For ease of presentation, all mathematical proofs are placed in the Appendix.

Theorem 1. *Regardless of the validity of the MIIVs, F converges in distribution to a chi-square distribution with L – K degrees of freedom as $n \to \infty$, provided that $\boldsymbol{h}(\boldsymbol{\sigma}) = \boldsymbol{0}$ and $L - K > 0$.*

Both Jin and Cao (2018) and Jin et al. (2021) assumed that the model is correctly specified and the MIIVs are valid, which are sufficient conditions of $\boldsymbol{h}(\boldsymbol{\sigma}) = \boldsymbol{0}$. However, they are not necessary conditions. If a misspecified equation and invalid MIIVs still fulfill the assumptions in Theorem 1, the power of F converges to the size, yielding an inconsistent test. The basic idea behind the overidentification test is that the MIIVs that are uncorrelated with the composite error term should not explain any variation in the residual $\boldsymbol{y}^* - \boldsymbol{Z}^*\hat{\boldsymbol{\theta}}_1$ (Kirby & Bollen, 2009). Theorem 1 shows that it is possible that invalid MIIVs do not explain any variation in the residual $\boldsymbol{y}^* - \boldsymbol{Z}^*\hat{\boldsymbol{\theta}}_1$, even though $\hat{\boldsymbol{\theta}}_1$ is not a consistent estimator of the true value.

Similar to the revised conditions in Theorem 1, we can revise the conditions in Theorem 2 of Jin et al. (2021) to attain the asymptotic distribution of F_{Sargan}, which is as follows.

Theorem 2. *The asymptotic distribution of F_{Sargan} is a weighted sum of independent chi-square random variables with 1 degrees of freedom, if $\boldsymbol{h}(\boldsymbol{\sigma}) = \boldsymbol{0}$. The weights are the eigenvalues of $\boldsymbol{\Pi} = \varphi^{-2}\boldsymbol{Q}^T\boldsymbol{\Omega}^{1/2}\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Omega}^{1/2}\boldsymbol{Q}$.*

If the degrees of freedom of the overidentification test is $L - K = 1$, the asymptotic distribution of F_{Sargan} is tr$\{\boldsymbol{\Pi}\}$ times a chi-square distribution with 1 degrees of freedom. Hence, we expect F and F_m to be asymptotically equivalent. The following theorem shows that F, F_m, and F_{mv} are equivalent even at the finite sample size when $L - K = 1$.

Theorem 3. *For any n, $F = F_m = F_{mv}$, provided that $L - K = 1$.*

Theorem 3 holds regardless of the value of $\boldsymbol{h}(\boldsymbol{\sigma})$. In the case where F is inconsistent against model misspecification when $L - K = 1$, F_m and F_{mv} are not consistent either. On the other hand, if F is consistent against model misspecification when $L - K = 1$, F_m and F_{mv} are also consistent.

The theorems above imply that the overidentification tests for MIIV are not useful when $\boldsymbol{h}(\boldsymbol{\sigma}) = \boldsymbol{0}$. We have also

Table 1. Percentage of sets of MIIVs that yield $h(\sigma) = 0$.

	Overidentification degrees of freedom ($L - K$)						
Omitted parameter	1	2	3	4	5	6	7
λ_{81}	20.88	3.30	0.30	–	–	–	–
b_2	28.57	5.71	–	–	–	–	–
$cor(\varepsilon_2, \varepsilon_3)$	85.71	78.57	71.43	64.29	57.14	50.00	42.86
$cor(\zeta_1, \zeta_2)$	60.00	31.43	13.1	4.29	0.83	–	–

seen from Figure 2 that various sets of MIIVs lead to inconsistent tests. In fact, all those sets of MIIVs result in $h(\sigma) = 0$, which can be easily checked in software. We can also enumerate all sets of MIIVs for the model in Figure 2, and check whether $h(\sigma) = 0$ holds. It is seen from Table 1 that it is common to have $h(\sigma) = 0$ in our SEM model. However, it is worth mentioning that when $cor(\varepsilon_2, \varepsilon_3)$ or $cor(\zeta_3, \zeta_4)$ is omitted, there still exist many sets of valid MIIVs such that θ_1 can be consistently estimated.

5. Implications

De Blander (2008), Guggenberger (2012), and Parente and Santos Silva (2012) tried to formulate the hypothesis under which the overidentification tests remain consistent. In particular, Parente and Santos Silva (2012) and Guggenberger (2012) pointed out that the overidentification test actually checks whether we can find a pseudo-true value θ_1^* of θ_1 such that

$$E[v^*(y^* - z^{*T}\theta_1^*)] = 0. \quad (13)$$

When the MIIV approach is used to fit a SEM model, the pseudo-true value θ_1^* corresponds to $\gamma(\sigma)$ and the moment condition (Equation (13)) corresponds to $h(\sigma)$. Furthermore, the condition (Equation (13)) is equivalent to the condition that we can find a vector a such that the linear system

$$\Sigma_{vz}a = E(v^*e) \quad (14)$$

is consistent (Parente & Santos Silva, 2012). The studies in the econometrics literature often focus on the case where the IVs are external. In contrast, the IVs in the MIIV approach are obtained from the indicators of the SEM model, thus in an internal manner. The structure of the SEM model allows us to decompose Σ_{vz} and $E(v^*e)$ in Equation (14). Consequently, we can better understand the implications of Equation (14) on the omitted parameters in SEM. In this section, we will present examples where $h(\sigma) = 0$ even in misspecified models. The scenarios where $h(\sigma) = 0$ only in the correctly specified model with valid MIIVs will also be presented. Similar to the previous section, the technical details are placed in the Appendix. Throughout the section, we assume that the elements in η are mutually correlated and that $var(\eta)$ is an invertible matrix.

5.1. Measurement Model

Suppose that $y_4^* = \lambda_1^T\eta_1 + \lambda_2^T\eta_2 + \varepsilon_4$ in the true model. The hypothesized model assumes that y_4^* is loaded only on η_2,

yielding a misspecified equation if $\lambda_1 \neq 0$. The scaling indicator of η_2 is hypothesized to satisfy $y_2^* = \eta_2 + \varepsilon_2$, but the correct equation is $y_2^* = \Lambda_1\eta_1 + \eta_2 + \Lambda_3\eta_3 + \varepsilon_2$. Then, our hypothesized model to estimate λ_2 is in fact

$$y_4^* = \lambda_2^T y_2^* + \varepsilon_4 - \lambda_2^T\varepsilon_2 + (\lambda_1^T - \lambda_2^T\Lambda_1)\eta_1 - \lambda_2^T\Lambda_3\eta_3. \quad (15)$$

We do not exclude the possibility that some MIIVs may be correlated with a subset of $\{\varepsilon_2, \varepsilon_4\}$ but the correlations are misspecified. Hence, without loss of generality, we partition the MIIVs v^* into v_1^* and v_2^* such that

$$v_1^* = \tilde{\Lambda}_1\tilde{\eta}_1 + \varepsilon_{v,1} \quad \text{and} \quad v_2^* = \tilde{\Lambda}_2\tilde{\eta}_2 + \varepsilon_{v,2}, \quad (16)$$

where $\tilde{\eta}_1$ ($\tilde{k}_1 \times 1$) and $\tilde{\eta}_2$ ($\tilde{k}_2 \times 1$) are the factors, $\tilde{\Lambda}_1$ and $\tilde{\Lambda}_2$ are of full column ranks, $\varepsilon_{v,1}$ is uncorrelated with $\{\varepsilon_2, \varepsilon_4\}$, and $\varepsilon_{v,2}$ is correlated with a subset of $\{\varepsilon_2, \varepsilon_4\}$.

Proposition 1. *Consider estimation of Equation (15) and the MIIVs given by Equation (16). Let $\eta_M^T = (\eta_1^T \quad \eta_3^T \quad \eta_1^T\Lambda_1^T + \eta_2^T + \eta_3^T\Lambda_3^T)$ when $\Lambda_3 \neq 0$, or $\eta_M^T = (\eta_1^T \quad \eta_1^T\Lambda_1^T + \eta_2^T)$ when $\Lambda_3 = 0$. Suppose that $\tilde{\Lambda}_1$ is of full column rank, that Σ_{vz} is of full column rank, and that $cov(\tilde{\eta}_1, \eta_M^T)$ is of full column rank. Then, $h(\sigma) = 0$ if and only if $\lambda_1 - \Lambda_1^T\lambda_2 = 0$, $\Lambda_3^T\lambda_2 = 0$, and $cov(\varepsilon_{v,2}, \varepsilon_4) = cov(\varepsilon_{v,2}, \varepsilon_2^T)\lambda_2$.*

Proposition 1 shows that what assumptions are needed in order for the overidentification tests to be consistent against a nonzero λ_1. Besides the full column rank assumptions, if the hypothesized equation $y_2^* = \eta_2 + \varepsilon_2$ for the scaling indicator is correctly specified, then we must have $\lambda_1 = 0$ in order to have $h(\sigma) = 0$. It can be easily checked that the sets of MIIVs in Figure 3a and c all satisfy the sufficient condition in Proposition 1. As expected, the powers of the overidentification tests all converge to 1, despite that their small sample powers can differ. It is worth mentioning that we do not necessarily need the MIIVs to be loaded on the omitted η_1 to achieve a consistent test. The covariance matrix $cov(\tilde{\eta}_1, \eta_M^T)$ can still have full column rank even though η_1 is not included in $\tilde{\eta}_1$ (e.g., using IVs (y_6^*, y_9^*) in Figure 3a).

Because of the condition $cov(\varepsilon_{v,2}, \varepsilon_4) = cov(\varepsilon_{v,2}, \varepsilon_2^T)\lambda_2$, Proposition 1 also sheds some lights on testing the measurement error correlation. If $cov(\varepsilon_{v,2}, \varepsilon_2^T) = 0$, then $h(\sigma) = 0$ if and only if $cov(\varepsilon_{v,2}, \varepsilon_4) = 0$. This means that misspecification can be detected as long as some MIIVs are correlated with ε_4. In contrast, if $cov(\varepsilon_{v,2}, \varepsilon_2^T) \neq 0$, we can always find $cov(\varepsilon_{v,2}, \varepsilon_4)$ and λ_2 such that the linear system $cov(\varepsilon_{v,2}, \varepsilon_4) = cov(\varepsilon_{v,2}, \varepsilon_2^T)\lambda_2$ holds. In other words, there always exist "lucky" numbers such that $h(\sigma) = 0$ but $cov(\varepsilon_{v,2}, \varepsilon_2^T) \neq 0$ is omitted in the hypothesized model. Hence, the overidentification tests can be inconsistent if the error term of the scaling indicator is correlated with some other error terms but ignored in the hypothesized model.

5.2. Misspecified Latent Regression Model

Suppose that in the true model $\eta_3 = b_1^T\eta_1 + b_2^T\eta_2 + \zeta_3$, but the hypothesized model ignores $b_1^T\eta_1$ yielding a misspecified model if $b_1 \neq 0$. The scaling indicators are hypothesized to satisfy $y_2^* = \eta_2 + \varepsilon_2$ and $y_3^* = \eta_3 + \varepsilon_3$, but the correct

equations are $y_2^* = \Lambda_{21}\eta_1 + \eta_2 + \lambda_{23}\eta_3 + \Lambda_{24}\eta_4 + \Lambda_{25}\eta_5 + \varepsilon_2$ and $y_3^* = \lambda_{31}^T\eta_1 + \lambda_{32}^T\eta_2 + \eta_3 + \lambda_{34}^T\eta_4 + \lambda_{36}^T\eta_6 + \varepsilon_3$. Then, our hypothesized model to estimate b_2 is

$$y_3^* = b_2^T y_2^* + \zeta_3 + \varepsilon_3 - b_2^T\varepsilon_2 + (b_1^T + \lambda_{31}^T - b_2^T\Lambda_{21})\eta_1$$
$$+ \lambda_{32}^T\eta_2 - b_2^T\lambda_{23}\eta_3 + (\Lambda_{34} - b_2^T\Lambda_{24})\eta_4 - b_2^T\Lambda_{25}\eta_5 + \Lambda_{36}\eta_6.$$
$$(17)$$

Because of ζ_3 in the composite error, any items loaded on η_3 should be excluded as MIIVs. Similar to omitted factor loadings, we do not exclude the possibility that some MIIVs may be correlated with a subset of $\{\zeta_3, \varepsilon_2, \varepsilon_3\}$ but the correlations are misspecified. Hence, without loss of generality, we again partition the MIIVs into Equation (16), where $\tilde{\Lambda}_1$ and $\tilde{\Lambda}_2$ are of full column ranks, v_1^* is uncorrelated with $\{\zeta_3, \varepsilon_2, \varepsilon_3\}$, and v_2^* is correlated with $\{\zeta_3, \varepsilon_2, \varepsilon_3\}$. A nonzero correlation between some η and ζ_3 can arise from, for example, omitted error correlations among ζ or another misspecified latent regression equation. See the misspecified model in the simulation study of Jin et al. (2021) for the latter case.

Proposition 2. *Consider estimation of Equation (17)) and the MIIVs given by Equation (16). Let η_M be the rows of $(\eta_1^T \quad \eta_4^T \quad \eta_5^T \quad \eta_6^T \quad \eta_1^T(\Lambda_{21} + b_1\lambda_{23}^T) + \eta_2^T(I + b_2\lambda_{23}^T) + \eta_4^T\Lambda_{24} + \eta_5^T\Lambda_{25})^T$ that corresponds to nonzero rows in*

$$\begin{pmatrix} \mathbf{1}_1 \\ \lambda_{34} - \Lambda_{24}^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2) \\ -\Lambda_{25}^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2) \\ \lambda_{36} \\ \mathbf{1}_2 \end{pmatrix},$$

where $\mathbf{1}_1$ and $\mathbf{1}_2$ are vectors of 1's that have the same dimensions as η_1 and $(\Lambda_{21} + \lambda_{23}b_1^T)\eta_1 + (I + \lambda_{23}b_2^T)\eta_2 + \Lambda_{24}\eta_4 + \Lambda_{25}\eta_5$, respectively. Suppose that $\tilde{\Lambda}_1$ is of full column rank, that Σ_{vz} is of full column rank, and that $cov(\tilde{\eta}_1, \eta_M^T)$ is of full column rank. Then, $h(\sigma) = \mathbf{0}$ if and only if $\lambda_{31} + b_1 = (\Lambda_{21}^T + b_1\lambda_{23}^T)(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2)$, $\lambda_{34} = \Lambda_{24}^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2)$, $\Lambda_{25}^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2) = \mathbf{0}$, $\lambda_{36} = \mathbf{0}$, and

$$cov\{\tilde{\Lambda}_2\tilde{\eta}_2, \zeta_3[1 - \lambda_{23}^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2)]\}$$
$$= -cov\{\varepsilon_{v,2}, \varepsilon_3 - \varepsilon_2^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2)\}. \quad (18)$$

Similar to Proposition 1 for the measurement model, Proposition 2 shows that further assumptions are needed in order for the overidentification tests to be consistent against a misspecified b_1. Besides the full column rank assumptions, if we further assume the hypothesized equations for the scaling indicators are correctly specified, then we must have $b_1 = \mathbf{0}$ in order for $h(\sigma) = \mathbf{0}$. The sets of MIIVs in Figure 3 all satisfy the conditions in Proposition 2, yielding the expected high power at large samples. Similar to the test for omitted factor loadings, we do not necessarily need the MIIVs to be loaded on η_1 to achieve a consistent test, even though η_1 is mistakenly omitted from the model. For example, the MIIVs in Figure 3b are not loaded on the omitted η_3, but their powers approach one.

Because of the inclusion of Equation (18), Proposition 2 also sheds light on testing omitted latent regression error

correlation. However, similar to the previous discussion for the omitted measurement error correlations, Equation (18) indicates that, without further restrictions, there always exist error correlations such that Equation (18) holds but the overidentification tests are inconsistent. If we further assume $\lambda_{23} = \mathbf{0}$, $\lambda_{32} = \mathbf{0}$, $cov(\varepsilon_{v,2}, \varepsilon_3) = cov(\varepsilon_{v,2}, \varepsilon_2^T)b_2$, and $\tilde{\Lambda}_2$ is of full column rank, then Equation (18) reduces to $\mathbf{0} = cov\{\tilde{\eta}_2, \zeta_3\}$, indicating that the overidentification tests are consistent.

5.3. Assumptions in Propositions

In practice, the conditions in Propositions 1 and 2 are often difficult to verify, since we do not know $\tilde{\eta}_1$ and $\tilde{\eta}_2$ in practice. Both Jin and Cao (2018) and Kirby and Bollen (2009) suggested to use all available MIIVs from the hypothesized model for the overidentification tests. However, their suggestion does not guarantee the consistency of the tests. Take the model in Figure 5 as an example. Even though Θ is correctly specified, $\tilde{\eta}_1 = (\eta_1, \eta_2)^T$ and $\eta_M = (\eta_1, \eta_2, \eta_3)^T$ violate the full column rank condition of $cov(\tilde{\eta}_1, \eta_M^T)$ in Proposition 1. Hence, it is possible that the overidentification tests are inconsistent. As we can see from Figure 5, the tests cannot detect the missing factor loadings. In fact the populations values given by Equations (6) and (7) are chosen based on Lemma 2 in the Appendix, which is needed to prove Proposition 1. In particular, Equation (6) corresponds to $2^{-1}cov(\tilde{\eta}_1, \eta_2^T) = cov(\tilde{\eta}_1, \eta_1^T)\lambda_1$ and Equation (7) corresponds to $2^{-1}\tilde{\Lambda}_2 cov(\tilde{\eta}_2, \eta_2^T) = \tilde{\Lambda}_2 cov(\tilde{\eta}_2, \eta_1^T)\lambda_1 + cov(\varepsilon_{v,2}, \varepsilon_4)$, where $\eta_1^T = (\eta_1, \eta_2)^T$, $\eta_2 = \eta_3$, $\lambda_1 = (\lambda_{12,1} \quad \lambda_{12,2})^T$, and $\tilde{\Lambda}_2 = (\lambda_{10,3} \quad \lambda_{11,3})^T$.

The full column rank assumptions in Propositions 1 and 2 are merely sufficient conditions. Violating the sufficient conditions do not necessarily yield inconsistent overidentification tests. However, there is a scenario that is worth mentioning. Suppose that hypothesized equations for the scaling indicators are correctly specified, that all MIIVs belong to v_1^*, and that $\tilde{\eta}_1 = \eta_2$. Then, $cov(\tilde{\eta}_1, \eta_M^T)$ is not of full column rank, and $\Sigma_{vz} = \tilde{\Lambda}_1 cov(\tilde{\eta}_1, \eta_2^T)$. Hence,

$$h(\sigma) = P(\sigma)\tilde{\Lambda}_1 cov(\tilde{\eta}_1, \eta_1^T)\lambda_1$$
$$= P(\sigma)\Sigma_{vz}[cov(\tilde{\eta}_1, \eta_2^T)]^{-1}cov(\tilde{\eta}_1, \eta_1^T)\lambda_1$$

in Proposition 1, or

$$h(\sigma) = P(\sigma)\tilde{\Lambda}_1 cov(\tilde{\eta}_1, \eta_1^T)b_1$$
$$= P(\sigma)\Sigma_{vz}[cov(\tilde{\eta}_1, \eta_2^T)]^{-1}cov(\tilde{\eta}_1, \eta_1^T)b_1,$$

in Proposition 2. Since $P(\sigma)\Sigma_{vz} = \mathbf{0}$, we get $h(\sigma) = \mathbf{0}$ regardless of the values of λ_1 or b_1. The MIIVs that lead to inconsistent tests in Figure 2a, b, and d belong to this particular scenario. Hence, it is always better to avoid the factors that the MIIVs loaded on having the same dimension as η_2.

6. Conclusion and Discussion

In this paper, we studied the consistency of the overidentification tests for the MIIV approach. In line with the

knowledge in the econometrics literature, we showed that the overidentification tests are generally inconsistent against model misspecification also in SEM. In other words, there always exist "lucky" numbers such that the overidentification tests are not consistent. However, if we are willing to impose various restrictions that are hard to verify in practice, the overidentification tests can be consistent against omitted nonzero parameters in SEM.

The MIIV overidentification tests are actually testing $h(\sigma) = 0$, which is only a necessarily condition for valid MIIVs and correct model specification. If we have strong reasons to believe that the scaling items are correctly specified, then, under the assumptions in Proposition 1, testing $h(\sigma) = 0$ is equivalent to testing $\lambda_1 = 0$ and $\text{cov}(\varepsilon_{v,2}, \varepsilon_4) = \text{cov}(\varepsilon_{v,2}, \varepsilon_2^T)\lambda_2$ for the measurement model. Likewise, under the assumptions in Proposition 2, testing $h(\sigma) = 0$ is equivalent to testing $b_1 = 0$, and $\text{cov}\{\tilde{\Lambda}_2\tilde{\eta}_2, \zeta_3\} = -\text{cov}\{\varepsilon_{v,2}, \varepsilon_3 - \varepsilon_2^T b_2\}$ for the latent regression model. Consequently, the overidentification tests are consistent against missing factor loadings and latent regression paths. However, the condition of $\text{cov}(\tilde{\eta}_1, \eta_M^T)$ being of full column rank is hard to verify in practice.

Both Jin and Cao (2018) and Kirby and Bollen (2009) recommended to use all available MIIVs from the hypothesized model for the overidentification tests. We have showed that their recommendations do not necessarily lead to consistent overidentification tests. However, their recommendations can avoid certain scenarios that definitely lead to inconsistent tests. For example, we have showed that it is always better to avoid factors that the MIIVs are loaded on having the same dimension of factors in the hypothesized model. Using all available MIIVs is more likely to avoid such a pitfall, even though it is not guaranteed so. As we can see from our second numerical illustration, using all MIIVs can still miss the omitted parameters, but the goodness-of-fit tests for the entire model are able to detect misspecified models.

From our numerical illustrations and theoretical results, we suggest that the overidentification tests should not be used alone. They are better used as preliminary diagnostic tools before fitting the whole model. Failing the overidentification tests strongly suggest that some assumptions can be wrong. However, not failing those tests is not a green light for a correct model. This suggestion is in line with Guggenberger (2012) and Guggenberger and Kumar (2012), who also suggested to use the overidentification test as a pre-test. It is also worth mentioning that a significant overidentification test does not necessarily mean that the hypothesized model is misspecified. It can also be caused by heterogeneity in the data (Angrist et al., 2000), e.g., multiple group data treated as one group. After preliminary screen of the overidentification tests, the whole model should be fitted and the goodness-of-fit tests should be carried out. It is also worth pointing out that, even though the overidentification tests are inconsistent against some "lucky" parameters, the MIIV estimator is still more robust than systemwide estimators. Hence, it is still more robust to interpret the MIIV estimator in a misspecified model, since the MIIV estimator is less likely to spread the bias due to one misspecification to other parts of the model.

References

, J. D., Graddy, K., & Imbens, G. W. (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *Review of Economic Studies*, 67, 499–527. https://doi.org/10.1111/1467-937X.00141

Bapat, R. B. (2012). *Linear algebra and linear models* (3rd ed.). Springer-Verlag.

Bollen, K. A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, 61, 109–121. https://doi.org/10.1007/BF02296961

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

en, K. A. (2020). When good loadings go bad: Robustness in factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 27, 515–524. https://doi.org/10.1080/10705511.2019.1691005

Bollen, K. A., Gates, K. M., & Fisher, Z. (2018). 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., & Maydeu-Olivares, A. (2007). A polychoric instrumental variable (PIV) estimator for structural equation models with categorical variables. *Psychometrika*, 72, 309–326. https://doi.org/10.1007/s11336-007-9006-3

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.

De Blander, R. (2008). Which null hypothesis do overidentification restrictions actually test? *Economics Bulletin*, 3, 1–9.

Fisher, Z. F., 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, 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, 246–263. https://doi.org/10.1080/00273171.2018.1519406

Guggenberger, P. (2012). A note on the (in)consistency of the test of overidentifying restrictions and the concepts of true and pseudo-true parameters. *Economics Letters*, 117, 901–904. https://doi.org/10.1016/j.econlet.2012.07.008

Guggenberger, P., & Kumar, G. (2012). On the size distortion of tests after an overidentifying restrictions pretest. *Journal of Applied Econometrics*, 27, 1138–1160. https://doi.org/10.1002/jae.1251

Jin, S., & Cao, C. (2018). Selecting polychoric instrumental variables in confirmatory factor analysis: An alternative specification test and effects of instrumental variables. *The 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. https://doi.org/10.1080/10705511.2016.1189334

Jin, S., Wallentin, F., & Bollen, K. A. (2021). A unified model-implied instrumental variable approach for structural equation modeling with mixed variables. *Psychometrika*, 86, 564–594.

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.

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.

Nestler, S. (2013). A Monte Carlo study comparing PIV, ULS and DWLS in the estimation of dichotomous confirmatory factor analysis. *The 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. *The 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. https://doi.org/10.1080/10705511.2014.994744

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. https://doi.org/10.1080/10705511.2014.934948

Newey, W. K. (1985). Generalized method of moments specification testing. *Journal of Econometrics, 29,* 229–256.

Parente, P. M., & Santos Silva, J. (2012). A cautionary note on tests of overidentifying restrictions. *Economics Letters, 115,* 314–317. https://doi.org/10.1016/j.econlet.2011.12.047

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

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). SAGE Publications.

Stock, J. H. (2015). Instrumental variables in statistics and econometrics. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (pp. 205–209, 2nd ed.). Elsevier.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data* (5th ed.). The MIT Press.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). South-Western Cengage Learning.

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. https://doi.org/10.1080/10705511.2010.489003

Appendix

Mathematical Proofs

In this section, the proofs of the theorems and propositions are provided. Before we prove the propositions, we first state a result that will be repeatedly used.

Lemma 1. *Suppose that* $\tilde{\Lambda}$ *is of full column rank. Then for any matrix* G*, the rank of* $\tilde{\Lambda}G$ *is the same as the rank of* G *(Bapat, 2012, p. 129).*

Proof of Theorem 1. Despite that the conditions in Theorem 1 of Jin et al. (2021) require correctly specified equations and valid IVs, its proof only relies on the assumption $h(\sigma) = 0$. Hence, their proof is applicable to the theorem with revised conditions. □

Proof of Theorem 2. Despite that the conditions in Theorem 2 of Jin et al. (2021) require correctly specified equations and valid IVs, its proof only relies on the assumption $h(\sigma) = 0$. Hence, their proof is applicable to the theorem with revised conditions. □

Proof of Theorem 3. Since Q is idempotent, the rank of Q equals its trace, which is $L-K = 1$. Hence, the singular value decomposition of Q is $Q = d^{1/2}pu^T$, where $d > 0$ is the eigenvalue of QQ^T, p is a $L \times 1$ vector, and u is also a $L \times 1$ vector. Likewise, the spectral decomposition of QQ^T is $QQ^T = dpp^T$, since $u^Tu = 1$. It can be easily verified that the Moore–Penrose inverse QQ^T is

$$G = d^{-1}pp^T. \tag{19}$$

When $L-K = 1$, the mean-scaled statistic reduces to

$$F_m = \frac{n(S_{vy} - S_{vz}\hat{\theta})^T S_{vv}^{-1}(S_{vy} - S_{vz}\hat{\theta})}{\operatorname{tr}\{\hat{Q}^T\hat{\Omega}^{1/2}S_{vv}^{-1}\hat{\Omega}^{1/2}\hat{Q}\}}.$$

Then, $F_m - F$ becomes

$$n(S_{vy} - S_{vz}\hat{\theta})^T \left[\frac{S_{vv}^{-1}}{\operatorname{tr}\{\hat{Q}^T\hat{\Omega}^{1/2}S_{vv}^{-1}\hat{\Omega}^{1/2}\hat{Q}\}} - \hat{\Omega}^{-1/2}(d^{-1}\hat{p}\hat{p}^T)\hat{\Omega}^{-1/2} \right]$$

$$(S_{vy} - S_{vz}\hat{\theta}) = nS_{vy}^T\hat{\Omega}^{-1/2} \left[\frac{\hat{Q}^T\hat{\Omega}^{1/2}S_{vv}^{-1}\hat{\Omega}^{1/2}\hat{Q}}{\operatorname{tr}\{\hat{\Omega}^{1/2}S_{vv}^{-1}\hat{\Omega}^{1/2}\hat{Q}\}} - d^{-1}\hat{Q}^T\hat{p}\hat{p}^T\hat{Q} \right] \hat{\Omega}^{-1/2}S_{vy},$$

where the equality holds since

$$\hat{\Omega}^{1/2}\hat{Q}\hat{\Omega}^{-1/2}S_{vy} = S_{vy} - S_{vz}\hat{\theta}.$$

Hence, the $F = F_m$ holds if

$$\frac{\hat{Q}^T\hat{\Omega}^{1/2}S_{vv}^{-1}\hat{\Omega}^{1/2}\hat{Q}}{\operatorname{tr}\{\hat{\Omega}^{1/2}S_{vv}^{-1}\hat{\Omega}^{1/2}\hat{Q}\}} = d^{-1}\hat{Q}^T\hat{p}\hat{p}^T\hat{Q}. \tag{20}$$

By the singular decomposition of Q, we can obtain

$$\begin{aligned}\hat{Q}^T\hat{\Omega}^{1/2}S_{vv}^{-1}\hat{\Omega}^{1/2}\hat{Q} &= \hat{d}\hat{u}\hat{p}^T\hat{\Omega}^{1/2}S_{vv}^{-1}\hat{\Omega}^{1/2}\hat{p}\hat{u}^T \\ &= \hat{d}(\hat{p}^T\hat{\Omega}^{1/2}S_{vv}^{-1}\hat{\Omega}^{1/2}\hat{p})\hat{u}\hat{u}^T,\end{aligned} \tag{21}$$

where the second equality holds since $\hat{p}^T\hat{\Omega}^{1/2}S_{vv}^{-1}\hat{\Omega}^{1/2}\hat{p}$ is a scalar when $L-K = 1$. Then

$$\operatorname{tr}\{\hat{Q}^T\hat{\Omega}^{1/2}S_{vv}^{-1}\hat{\Omega}^{1/2}\hat{Q}\} = \hat{d}(\hat{p}^T\hat{\Omega}^{1/2}S_{vv}^{-1}\hat{\Omega}^{1/2}\hat{p}), \tag{22}$$

since $\operatorname{tr}\{\hat{u}\hat{u}^T\} = \operatorname{tr}\{\hat{u}^T\hat{u}\} = 1$. Hence, the left-hand side of Equation (20) reduces to $\hat{u}\hat{u}^T$. Likewise, by replacing all $\hat{Q}$ by its singular value decomposition, the right-hand size of Equation (20) also reduces to $\hat{u}\hat{u}^T$. This proves the first half of the theorem.

To show the second half of the theorem, it suffices to show $[\operatorname{tr}\{\hat{\Pi}\}]^2 = \operatorname{tr}\{\hat{\Pi}^2\}$. By Equation (21), we obtain

$$\operatorname{tr}\{\hat{\Pi}^2\} = \hat{d}^2(\hat{p}^T\hat{\Omega}^{1/2}S_{vv}^{-1}\hat{\Omega}^{1/2}\hat{p})^2$$

since $\hat{u}^T\hat{u} = 1$ and $\operatorname{tr}\{\hat{u}\hat{u}^T\} = 1$. We have already derived $\operatorname{tr}\{\hat{\Pi}\}$ in Equation (22). This completes the proof of the theorem. □

Lemma 2. *Consider estimation of Equation (15) and the MIIVs given by Equation (16). Suppose that* Σ_{vz} *is of full column rank. Then,* $h(\sigma) = 0$ *if and only if there exists a vector* a *such that*

$$\operatorname{cov}(\tilde{\Lambda}_1\tilde{\eta}_1, \eta_1^T\Lambda_1^T + \eta_2^T + \eta_3^T\Lambda_3^T)a$$

$$= \operatorname{cov}(\tilde{\Lambda}_1\tilde{\eta}_1, \eta_1^T\lambda_1 - \eta_1^T\Lambda_1^T\lambda_2 - \eta_3^T\Lambda_3^T\lambda_2),$$

$$\operatorname{cov}(\tilde{\Lambda}_2\tilde{\eta}_2, \eta_1^T\Lambda_1^T + \eta_2^T + \eta_3^T\Lambda_3^T)a + \operatorname{cov}(\varepsilon_{v,2}, \varepsilon_2^T)a$$

$$= \operatorname{cov}(\tilde{\Lambda}_2\tilde{\eta}_2, \eta_1^T\lambda_1 - \eta_1^T\Lambda_1^T\lambda_2 - \eta_3^T\Lambda_3^T\lambda_2) + \operatorname{cov}(\varepsilon_{v,2}, \varepsilon_4) - \operatorname{cov}(\varepsilon_{v,2}, \varepsilon_2^T)\lambda_2.$$

Proof of Lemma 2. It is easily seen that $P(\sigma) = I - \Sigma_{vz}(\Sigma_{vz}^T\Sigma_{vv}^{-1}\Sigma_{vz})^{-1}\Sigma_{vz}^T\Sigma_{vv}^{-1}$ is idempotent. Then, the rank of $P(\sigma)$ equals its trace, which is $L - K$. Then, the column vectors in Σ_{vz} form a basis for the null space of $P(\sigma)$, since Σ_{vz} is of full column rank.

It can be easily shown that

$$\Sigma_{vz} = \begin{pmatrix} \text{cov}(\tilde{\Lambda}_1\tilde{\eta}_1, \eta_1^T\Lambda_1^T + \eta_2^T + \eta_3^T\Lambda_3^T) \\ \text{cov}(\tilde{\Lambda}_2\tilde{\eta}_2, \eta_1^T\Lambda_1^T + \eta_2^T + \eta_3^T\Lambda_3^T) \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \text{cov}(\varepsilon_{v,2}, \varepsilon_2^T) \end{pmatrix},$$

$$\Sigma_{vy} = \begin{pmatrix} \text{cov}(\tilde{\Lambda}_1\tilde{\eta}_1, \eta_1^T\lambda_1 + \eta_2^T\lambda_2) \\ \text{cov}(\tilde{\Lambda}_2\tilde{\eta}_2, \eta_1^T\lambda_1 + \eta_2^T\lambda_2) \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \text{cov}(\varepsilon_{v,2}, \varepsilon_4) \end{pmatrix}.$$

Since $P(\sigma)\Sigma_{vz} = \mathbf{0}$, we obtain

$$P(\sigma)\begin{pmatrix} \text{cov}(\tilde{\Lambda}_1\tilde{\eta}_1, \eta_2^T) \\ \text{cov}(\tilde{\Lambda}_2\tilde{\eta}_2, \eta_2^T) \end{pmatrix} = -P(\sigma)\begin{pmatrix} \text{cov}(\tilde{\Lambda}_1\tilde{\eta}_1, \eta_1^T\Lambda_1^T + \eta_3^T\Lambda_3^T) \\ \text{cov}(\tilde{\Lambda}_2\tilde{\eta}_2, \eta_1^T\Lambda_1^T + \eta_3^T\Lambda_3^T) \end{pmatrix}$$
$$-P(\sigma)\begin{pmatrix} \mathbf{0} \\ \text{cov}(\varepsilon_{v,2}, \varepsilon_2^T) \end{pmatrix}.$$

Hence, $h(\sigma) = P(\sigma)\Sigma_{vy}$ can be expressed as

$$h(\sigma) = P(\sigma)$$
$$\begin{pmatrix} \tilde{\Lambda}_1\text{cov}(\tilde{\eta}_1, \eta_1^T\lambda_1 - \eta_1^T\Lambda_1^T\lambda_2 - \eta_3^T\Lambda_3^T\lambda_2) \\ \tilde{\Lambda}_2\text{cov}(\tilde{\eta}_2, \eta_1^T\lambda_1 - \eta_1^T\Lambda_1^T\lambda_2 - \eta_3^T\Lambda_3^T\lambda_2) + \text{cov}(\varepsilon_{v,2}, \varepsilon_4) - \text{cov}(\varepsilon_{v,2}, \varepsilon_2^T)\lambda_2 \end{pmatrix}.$$

Because of $P(\sigma)\Sigma_{vz} = \mathbf{0}$, then $h(\sigma) = \mathbf{0}$ means that the column space of Σ_{vy} belongs to the null space of $P(\sigma)$, that is, the column space of Σ_{vz}. The proof is completed. □

Proof of Proposition 1. By Lemma 2, there exists a vector a such that

$$\tilde{\Lambda}_1\text{cov}(\tilde{\eta}_1, \eta_1^T\Lambda_1^T + \eta_2^T + \eta_3^T\Lambda_3^T)a = \tilde{\Lambda}_1\text{cov}(\tilde{\eta}_1, \eta_1^T\lambda_1 - \eta_1^T\Lambda_1^T\lambda_2 \\ - \eta_3^T\Lambda_3^T\lambda_2). \quad (23)$$

Since $\tilde{\Lambda}_1$ is of full column rank, the rank of $\tilde{\Lambda}_1\text{cov}(\tilde{\eta}_1, \eta_M^T)$ is the same as the rank of $\text{cov}(\tilde{\eta}_1, \eta_M^T)$ (Lemma 1). When $\Lambda_3 \neq \mathbf{0}$, Equation (23) can be expressed as

$$\tilde{\Lambda}_1\text{cov}(\tilde{\eta}_1, \eta_1^T\Lambda_1^T + \eta_2^T + \eta_3^T\Lambda_3^T)a = \tilde{\Lambda}_1\text{cov}[\tilde{\eta}_1, (\eta_1^T \quad \eta_3^T)](\begin{smallmatrix} \lambda_1 - \Lambda_1^T\lambda_2 \\ -\Lambda_3^T\lambda_2 \end{smallmatrix}).$$

By the assumption that $\text{cov}(\tilde{\eta}_1, \eta_M^T)$ is of full column rank, we must have $\lambda_1 - \Lambda_1^T\lambda_2 = \mathbf{0}$, $\Lambda_3^T\lambda_2 = \mathbf{0}$, and $a = \mathbf{0}$. Consequently, Lemma 2 indicates that $\text{cov}(\varepsilon_{v,2}, \varepsilon_4) = \text{cov}(\varepsilon_{v,2}, \varepsilon_2^T)\lambda_2$. When $\Lambda_3 = \mathbf{0}$, Equation (23) can be expressed as

$$\tilde{\Lambda}_1\text{cov}(\tilde{\eta}_1, \eta_1^T\Lambda_1^T + \eta_2^T + \eta_3^T\Lambda_3^T)a = \tilde{\Lambda}_1\text{cov}(\tilde{\eta}_1, \eta_1^T)(\lambda_1 - \Lambda_1^T\lambda_2).$$

By the assumption that $\text{cov}(\tilde{\eta}_1, \eta_M^T)$ is of full column rank, we must have $\lambda_1 - \Lambda_1^T\lambda_2 = \mathbf{0}$ and $a = \mathbf{0}$. Consequently, Lemma 2 further indicates that $\text{cov}(\varepsilon_{v,2}, \varepsilon_4) = \text{cov}(\varepsilon_{v,2}, \varepsilon_2^T)\lambda_2$. This completes the proof. □

Lemma 3. *Consider estimation of Equation (17) and the MIIVs given by Equation (16). Suppose that Σ_{vz} is of full column rank. Then, $h(\sigma) = \mathbf{0}$ if and only if there exists a vector a such that*

$$\text{cov}\left(\tilde{\Lambda}_1\tilde{\eta}_1, \eta_1^T(\Lambda_{21}^T + b_1\lambda_{23}^T) + \eta_2^T(I + b_2\lambda_{23}^T) + \eta_4^T\Lambda_{24}^T + \eta_5^T\Lambda_{25}^T\right)a$$
$$= \text{cov}\{\tilde{\Lambda}_1\tilde{\eta}_1, \eta_1^T[\lambda_{31} + b_1 - (\Lambda_{21}^T + b_1\lambda_{23}^T)(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2)]\}$$
$$+ \text{cov}\{\tilde{\Lambda}_1\tilde{\eta}_1, \eta_4^T[\lambda_{34} - \Lambda_{24}^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2)]$$
$$- \eta_5^T\Lambda_{25}^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2) + \eta_6^T\lambda_{36}\}$$

and

$$\text{cov}\left(\tilde{\Lambda}_2\tilde{\eta}_2, \eta_1^T(\Lambda_{21}^T + b_1\lambda_{23}^T) + \eta_2^T(I + b_2\lambda_{23}^T) + \eta_4^T\Lambda_{24}^T + \eta_5^T\Lambda_{25}^T + \zeta_3\lambda_{23}^T\right)$$
$$a + \text{cov}(\varepsilon_{v,2}, \varepsilon_2^T)a = \text{cov}\{\tilde{\Lambda}_2\tilde{\eta}_2, \eta_1^T[\lambda_{31} + b_1 - (\Lambda_{21}^T + b_1\lambda_{23}^T)(I + b_2\lambda_{23}^T)^{-1}$$
$$(\lambda_{32} + b_2)]\} + \text{cov}\{\tilde{\Lambda}_2\tilde{\eta}_2, \eta_4^T[\lambda_{34} - \Lambda_{24}^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2)]$$

$$- \eta_5^T\Lambda_{25}^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2) + \eta_6^T\lambda_{36}\}$$

$$+ \text{cov}\{\tilde{\Lambda}_2\tilde{\eta}_2, \zeta_3[1 - \lambda_{23}^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2)]\}$$

$$+ \text{cov}\{\varepsilon_{v,2}, \varepsilon_3 - \varepsilon_2^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2)\}.$$

Proof of Lemma 3. Similar to Lemma 2, the column vectors in Σ_{vz} form a basis for the null space of $P(\sigma)$, since Σ_{vz} is of full column rank. Furthermore, it can be easily shown that

$$\Sigma_{vz} =$$
$$\begin{pmatrix} \text{cov}\left(\tilde{\Lambda}_1\tilde{\eta}_1, \eta_1^T(\Lambda_{21}^T + b_1\lambda_{23}^T) + \eta_2^T(I + b_2\lambda_{23}^T) + \eta_4^T\Lambda_{24}^T + \eta_5^T\Lambda_{25}^T\right) \\ \text{cov}\left(\tilde{\Lambda}_2\tilde{\eta}_2, \eta_1^T(\Lambda_{21}^T + b_1\lambda_{23}^T) + \eta_2^T(I + b_2\lambda_{23}^T) + \eta_4^T\Lambda_{24}^T + \eta_5^T\Lambda_{25}^T + \zeta_3\lambda_{23}^T\right) + \text{cov}(\varepsilon_{v,2}, \varepsilon_2^T) \end{pmatrix},$$

$$\Sigma_{vy} =$$
$$\begin{pmatrix} \text{cov}\left(\tilde{\Lambda}_1\tilde{\eta}_1, \eta_1^T(\lambda_{31} + b_1) + \eta_2^T(\lambda_{32} + b_2) + \eta_4^T\lambda_{34} + \eta_6^T\lambda_{36}\right) \\ \text{cov}\left(\tilde{\Lambda}_2\tilde{\eta}_2, \eta_1^T(\lambda_{31} + b_1) + \eta_2^T(\lambda_{32} + b_2) + \eta_4^T\lambda_{34} + \eta_6^T\lambda_{36}\right) + \text{cov}(\tilde{\Lambda}_2\tilde{\eta}_2, \zeta_3) + \text{cov}(\varepsilon_{v,2}, \varepsilon_3) \end{pmatrix}.$$

Since $P(\sigma)\Sigma_{vz} = \mathbf{0}$, we obtain

$$P(\sigma)\begin{pmatrix} \text{cov}(\tilde{\Lambda}_1\tilde{\eta}_1, \eta_2^T) \\ \text{cov}(\tilde{\Lambda}_2\tilde{\eta}_2, \eta_2^T) \end{pmatrix}(I + b_2\lambda_{23}^T)$$

$$= -P(\sigma)\begin{pmatrix} \text{cov}\left(\tilde{\Lambda}_1\tilde{\eta}_1, \eta_1^T(\Lambda_{21}^T + b_1\lambda_{23}^T) + \eta_4^T\Lambda_{24}^T + \eta_5^T\Lambda_{25}^T\right) \\ \text{cov}\left(\tilde{\Lambda}_2\tilde{\eta}_2, \eta_1^T(\Lambda_{21}^T + b_1\lambda_{23}^T) + \eta_4^T\Lambda_{24}^T + \eta_5^T\Lambda_{25}^T\right) \end{pmatrix}$$

$$- P(\sigma)\begin{pmatrix} \mathbf{0} \\ \text{cov}(\tilde{\Lambda}_2\tilde{\eta}_2, \zeta_3)\lambda_{23}^T + \text{cov}(\varepsilon_{v,2}, \varepsilon_2^T) \end{pmatrix}.$$

Hence, $h(\sigma) = P(\sigma)\Sigma_{vy}$ can be expressed as

$$h(\sigma) = P(\sigma)$$

$$\begin{pmatrix} \text{cov}\{\tilde{\Lambda}_1\tilde{\eta}_1, \eta_1^T[\lambda_{31} + b_1 - (\Lambda_{21}^T + b_1\lambda_{23}^T)(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2)]\} \\ \text{cov}\{\tilde{\Lambda}_2\tilde{\eta}_2, \eta_1^T[\lambda_{31} + b_1 - (\Lambda_{21}^T + b_1\lambda_{23}^T)(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2)]\} \end{pmatrix}$$

$$+ P(\sigma)\begin{pmatrix} \text{cov}\{\tilde{\Lambda}_1\tilde{\eta}_1, \eta_4^T[\lambda_{34} - \Lambda_{24}^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2)]\} \\ \text{cov}\{\tilde{\Lambda}_2\tilde{\eta}_2, \eta_4^T[\lambda_{34} - \Lambda_{24}^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2)]\} \end{pmatrix}$$

$$+ P(\sigma)\begin{pmatrix} \text{cov}\{\tilde{\Lambda}_1\tilde{\eta}_1, - \eta_5^T\Lambda_{25}^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2) + \eta_6^T\lambda_{36}\} \\ \text{cov}\{\tilde{\Lambda}_2\tilde{\eta}_2, - \eta_5^T\Lambda_{25}^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2) + \eta_6^T\lambda_{36}\} \end{pmatrix}$$

$$+ P(\sigma)\begin{pmatrix} \mathbf{0} \\ \text{cov}\{\tilde{\Lambda}_2\tilde{\eta}_2, \zeta_3[1 - \lambda_{23}^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2)]\} \end{pmatrix}$$

$$+ P(\sigma)\begin{pmatrix} \mathbf{0} \\ \text{cov}\{\varepsilon_{v,2}, \varepsilon_3 - \varepsilon_2^T(I + b_2\lambda_{23}^T)^{-1}(\lambda_{32} + b_2)\} \end{pmatrix}.$$

Because of $P(\sigma)\Sigma_{vz} = \mathbf{0}$, then $h(\sigma) = \mathbf{0}$ means that the column space of Σ_{vy} belongs to the null space of $P(\sigma)$, that is, the column space of Σ_{vz}. The proof is completed. □

Proof of Proposition 2. By Lemma 3, there exists a vector a such that

$$\tilde{\Lambda}_1\text{cov}\left(\tilde{\eta}_1, \eta_1^T(\Lambda_{21}^T + b_1\lambda_{23}^T) + \eta_2^T(I + b_2\lambda_{23}^T) + \eta_4^T\Lambda_{24}^T + \eta_5^T\Lambda_{25}^T\right)a$$
$$= \tilde{\Lambda}_1\text{cov}\{\tilde{\eta}_1, (\eta_1^T, \eta_4^T, \eta_5^T, \eta_6^T)\}$$

$$\begin{pmatrix} \boldsymbol{\lambda}_{31} + \boldsymbol{b}_1 - (\boldsymbol{\Lambda}_{21}^T + \boldsymbol{b}_1 \boldsymbol{\lambda}_{23}^T)(\boldsymbol{I} + \boldsymbol{b}_2 \boldsymbol{\lambda}_{23}^T)^{-1}(\boldsymbol{\lambda}_{32} + \boldsymbol{b}_2) \\ \boldsymbol{\lambda}_{34} - \boldsymbol{\Lambda}_{24}^T(\boldsymbol{I} + \boldsymbol{b}_2 \boldsymbol{\lambda}_{23}^T)^{-1}(\boldsymbol{\lambda}_{32} + \boldsymbol{b}_2) \\ -\boldsymbol{\Lambda}_{25}^T(\boldsymbol{I} + \boldsymbol{b}_2 \boldsymbol{\lambda}_{23}^T)^{-1}(\boldsymbol{\lambda}_{32} + \boldsymbol{b}_2) \\ \boldsymbol{\lambda}_{36} \end{pmatrix}.$$

Since $\tilde{\boldsymbol{\Lambda}}_1$ is of full column rank, the rank of $\tilde{\boldsymbol{\Lambda}}_1 \text{cov}(\tilde{\boldsymbol{\eta}}_1, \boldsymbol{\eta}_M^T)$ is the same as the rank of $\text{cov}(\tilde{\boldsymbol{\eta}}_1, \boldsymbol{\eta}_M^T)$ (Lemma 1). By the assumption that $\text{cov}(\tilde{\boldsymbol{\eta}}_1, \boldsymbol{\eta}_M^T)$ is of full column rank, we must have $\boldsymbol{\lambda}_{31} + \boldsymbol{b}_1 =$

$(\boldsymbol{\Lambda}_{21}^T + \boldsymbol{b}_1 \boldsymbol{\lambda}_{23}^T)(\boldsymbol{I} + \boldsymbol{b}_2 \boldsymbol{\lambda}_{23}^T)^{-1}(\boldsymbol{\lambda}_{32} + \boldsymbol{b}_2)$, $\quad \boldsymbol{\lambda}_{34} = \boldsymbol{\Lambda}_{24}^T(\boldsymbol{I} + \boldsymbol{b}_2 \boldsymbol{\lambda}_{23}^T)^{-1}(\boldsymbol{\lambda}_{32} + \boldsymbol{b}_2)$, $\quad \boldsymbol{\Lambda}_{25}^T(\boldsymbol{I} + \boldsymbol{b}_2 \boldsymbol{\lambda}_{23}^T)^{-1}(\boldsymbol{\lambda}_{32} + \boldsymbol{b}_2) = \boldsymbol{0}$, $\quad \boldsymbol{\lambda}_{36} = \boldsymbol{0}$, $\quad$ and $\quad \boldsymbol{a} = \boldsymbol{0}$. Consequently, Lemma 3 indicates that

$$\text{cov}\{\tilde{\boldsymbol{\Lambda}}_2 \tilde{\boldsymbol{\eta}}_2, \zeta_3[1 - \boldsymbol{\lambda}_{23}^T(\boldsymbol{I} + \boldsymbol{b}_2 \boldsymbol{\lambda}_{23}^T)^{-1}(\boldsymbol{\lambda}_{32} + \boldsymbol{b}_2)]\}$$
$$= -\text{cov}\{\boldsymbol{\varepsilon}_{v,2}, \varepsilon_3 - \boldsymbol{\varepsilon}_2^T(\boldsymbol{I} + \boldsymbol{b}_2 \boldsymbol{\lambda}_{23}^T)^{-1}(\boldsymbol{\lambda}_{32} + \boldsymbol{b}_2)\}.$$

This completes the proof. $\qquad \square$