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

Shaobo Jin

To cite this article: Shaobo Jin (2023) On Inconsistency of the Overidentification Test for the Model-implied Instrumental Variable Approach, Structural Equation Modeling: A Multidisciplinary Journal, 30:2, 245-257, DOI: 10.1080/10705511.2022.2122978

To link to this article: https://doi.org/10.1080/10705511.2022.2122978

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

Published online: 04 Oct 2022.

Submit your article to this journal

Article views: 307

View related articles

View Crossmark data
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 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

\[ y^* = \Lambda \eta + \varepsilon, \]

\[ \eta = B \theta + \zeta, \]

where \( y^* \) is the vector of continuous variables, \( \Lambda \) is the matrix of factor loadings, \( \eta \) is the vector of latent variables, \( B \) is the matrix of latent regression coefficients, \( \varepsilon \) is the disturbance of the measurement model with var(\( \varepsilon \)) = \( \Theta \), and \( \zeta \) is disturbance of the latent regression model with var(\( \zeta \)) = \( \Psi \). Both \( \varepsilon \) and \( \zeta \) are homoscedastic in the model. We follow Jin et al. (2021) and assume that \( E(\varepsilon^*) = 0 \). Instead of observing \( y^* \), the observed vector is denoted by \( y \). If the \( j \)th element of \( y \), denoted by \( y_j \), is continuous, then \( y_j = y^*_j \). 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 \( \eta \) is independent of \( \varepsilon \) and that \( \zeta \) is independent of \( \varepsilon \). 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 \( \eta \) 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 \( \eta \) in the SEM model, we obtain a regression system with observed variables

\[
\begin{pmatrix}
    y^*_2 \\
    y^*_1
\end{pmatrix}
=
\begin{pmatrix}
    \Lambda_2 \\
    B
\end{pmatrix}
\begin{pmatrix}
    \eta \\
    \zeta + (I - B)e_1
\end{pmatrix},
\]

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 \( \theta \) into two vectors: regression coefficient \( \theta_1 \) (free parameters in \( \Lambda_2 \) and \( B \)) and dispersion parameter \( \theta_2 \) (free parameters in \( \Psi \) and \( \Theta \)).

In most cases, \( y^*_1 \) is correlated with the composite error term. Hence, IV regression is used to estimate the parameters of \( \theta_1 \) in a row-by-row manner (Bollen, 1996; Bollen & Maydeu-Olivares, 2007). Suppose that the \( j \)th row of the system (Equation (3)) is

\[ y^*_j = z_j^T \theta_1^{(j)} + \varepsilon_j, \]

where \( \theta_1^{(j)} \) is the vector of unknown parameters and \( z_j^* \) is subset of \( y^*_j \) 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 \( \varepsilon \). Second, there exists a set of IVs, denoted by \( v^* \), which is correlated with \( z^* \) and uncorrelated with \( \varepsilon \). 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 \( \Sigma \) be the population covariance matrix of \( y^* \) and \( S \) be its sample counterpart. We always assume that \( s \) is a consistent estimator of \( \sigma \) such that

\[ \sqrt{n}(s - \sigma) \xrightarrow{d} N(0, \Theta), \]

where \( n \) is the sample size, \( s \) is the vector of free elements in \( \Sigma \), \( \sigma \) is the vector of free elements in \( \Sigma \), and \( \Theta \) is the asymptotic covariance matrix. Let
\[
\gamma(\sigma) = \left( \Sigma_{\epsilon_z}^{-1} \Sigma_{v_z}^{-1} \Sigma_{w_z} \right)^{-1} \Sigma_{v_y}^{-1} \Sigma_{w_y},
\]
where \( \Sigma_{\epsilon_z} = \text{cov}(\epsilon^*, \epsilon^*) \), \( \Sigma_{v_z} = \text{var}(\nu^*) \), and \( \Sigma_{w_y} = \text{cov}(\nu^*, \nu^*) \). Then, the MIIV estimator of \( \theta \) is \( \hat{\theta} = \gamma(s) \).

Because of the consistency of \( s \), \( \gamma(s) \) is a consistent estimator of \( \gamma(\sigma) \). It is also a consistent estimator of the true \( \theta \), if Equation (4) is correctly specified, the MIIVs are valid, and both \( \Sigma_{\epsilon_z}^{-1} \Sigma_{v_z}^{-1} \Sigma_{w_z} \) and \( \Sigma_{v_y}^{-1} \Sigma_{w_y} \) are invertible. However, if some MIIVs are invalid, then \( \gamma(\sigma) \) is not necessarily equal to \( \theta \).

When all observed variables are continuous, Kirby and Bollen (2009) proposed to use the Sargan (1958) chi-square statistic
\[
F_{\text{Sargan}} = \frac{n}{\hat{\theta}^2} (S_{v_y} - S_{v_z} \hat{\theta})^T S_{v_y}^{-1} (S_{v_y} - S_{v_z} \hat{\theta}),
\]
to test the specification of the \( j \)th equation, where \( \hat{\theta}^2 = S_{v_y} - 2\theta^T S_{v_z} + \theta^T S_{v_z} \theta \). When all observed variables are ordinal, Jin and Cao (2018) showed that the asymptotic distribution of \( F_{\text{Sargan}} \) is a weighted sum of chi-square distributions with one degree of freedom and proposed to apply the Satorra and Bentler (1994) adjustments to \( F_{\text{Sargan}} \). They also proposed an asymptotic chi-square distributed statistic.

Recently, Jin et al. (2021) proposed the overidentification tests in Jin and Cao (2018). 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 (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
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 $\lambda_{81}$ is omitted when estimating $\lambda_{82}$ (Figure 2a), the equation to be estimated can be expressed as

$$y_8^* = \lambda_{81} \eta_1 + \lambda_{82} \varepsilon_2 + \varepsilon_8.$$

In the hypothesized model where $\lambda_{81} = 0$, all indicators expect $y_8^*$ 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 $\eta$ 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

$$y_8^* = \lambda_{81} \eta_1 + \lambda_{82} \varepsilon_2 + \varepsilon_8 - \lambda_{82} \varepsilon_5 + \lambda_{81} \eta_1.$$

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), \( \eta_3 \) 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 \( \lambda_{21} \) (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), \( y_3^* \) to \( y_8^* \) and \( y_9^* \) to \( y_{12}^* \) are selected as MIIVs. However, \( y_9^* \) to \( y_{12}^* \) are invalid, since they rely on \( \eta_4 \) that is correlated with \( \zeta_4 \) through \( \zeta_5 \).

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 guarantee 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

\[(a) \; y_8^* = \lambda_{81} \eta_1 + \lambda_{82} \eta_2 + \varepsilon_8 \text{ in the true model but } \lambda_{81} = 0 \text{ in the hypothesized model} \]

\[(b) \; \eta_4 = b_1 \eta_1 + b_2 \eta_3 + \zeta_4 \text{ in the true model but } b_2 = 0 \text{ in the hypothesized model} \]

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

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

**Figure 3.** Empirical rejection rate of the consistent overidentification tests. The MIIVs used for the overidentification tests are stated in the figure title.
The population values of the cross loadings ($k_{21,1}$ and $k_{12,2}$) are set to
\[
\frac{1}{2} \left( \text{var}(\eta_1) \quad \text{cov}(\eta_1, \eta_2) \right) \left( \text{cov}(\eta_1, \eta_1) \quad \text{cov}(\eta_2, \eta_2) \right). \tag{6}
\]

The population values of the error covariances ($\text{cov}(e_{10}, e_{12})$ and $\text{cov}(e_{11}, e_{12})$) are set to
\[
\frac{1}{2} \left( \text{var}(\eta_3) - \hat{\lambda}_{12,1} \text{cov}(\eta_1, \eta_3) - \hat{\lambda}_{12,2} \text{cov}(\eta_2, \eta_3) \right) \left( \hat{\lambda}_{12,1} \quad \hat{\lambda}_{12,2} \right), \tag{7}
\]
which depend on the values of $\hat{\lambda}_{21,1}$ and $\hat{\lambda}_{12,2}$ from Equation (6). The reasons of choosing such values will be explained in a later section.

We let $y_2$, $y_4$, $y_6$, $y_8$, $y_{10}$, and $y_{12}$ be ordinal, whereas other indicators are continuous indicators. Two misspecified models are considered: one model omits two cross loadings and another model omits both cross loadings and error correlations. In the first model $y_1$ to $y_9$ will be selected as MIIVs, and in the second model all indicators except $y_9$ and $y_{12}$ will be selected as MIIVs. Similar to Example 1 when some cross loadings are omitted, all indicators are invalid MIIVs, due to the correlations among factors.

In this example, we focus on the equation for $y_{12}$. It is seen from Figure 5 that the overidentification tests are inconsistent in both models even when all MIIVs are used. Alternatively, we can estimate the whole model using MIIV (including the covariance parameters), and apply the MIIV goodness-of-fit tests developed in Jin et al. (2021). It is seen from Figure 6 that the goodness-of-fit tests are consistent.
4. Inconsistency of Overidentification Tests

The starting point of IV regression to estimate \( \theta_1 \) in Equation (4) is the moment condition

\[
E[v^*(y^* - z^T \theta_1)] = 0. \tag{8}
\]

This condition holds if \( v^* \) 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 \( h(\sigma) = \Sigma_{G\gamma} - \Sigma_{w\gamma}(\sigma) \). The delta method to Equation (5) indicates that

\[
\sqrt{n}h(s) = \sqrt{n}h(\sigma) + \frac{\partial h(\sigma)}{\partial \sigma} \cdot \sqrt{n}(s - \sigma) + o_p(1), \tag{9}
\]

where \( h(s) = S_{G\gamma} - S_{w\gamma}(s) \). By the chain rule, the ith column in the partial derivative \( \frac{\partial h(\sigma)}{\partial \sigma_i} \) can be computed by

\[
\frac{\partial h(\sigma)}{\partial \sigma_i} = P(\sigma)c_i(\sigma) - \Sigma_{w\gamma}(\Sigma_{G\gamma}^{-1} - \Sigma_{w\gamma}^{-1})^{-1} \Sigma_{w\gamma}^{-1}\Sigma_{w\gamma}^{-1}h(\sigma), \tag{10}
\]

where \( P(\sigma) = I - \Sigma_{w\gamma}(\Sigma_{G\gamma}^{-1} - \Sigma_{w\gamma}^{-1})^{-1} \Sigma_{w\gamma}^{-1}\Sigma_{w\gamma}^{-1} \) and \( c_i(\sigma) \) is the ith column of \( C(\sigma) \) (Jin et al., 2021). Then, we obtain

\[
\sqrt{n}Q^{-1/2}h(s) = \sqrt{n}Q^{-1/2}h(\sigma) = N(0, Q^TQ),
\]

since \( Q \) is idempotent, \( Q^{-1/2} \Sigma_{w\gamma} = 0 \), and

\[
Q^{-1/2}P(\sigma) = Q^{-1/2}. \tag{11}
\]

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

\[
\sqrt{n}Q^{-1/2}h(s) \overset{d}{\rightarrow} N(0, Q^TQ),
\]

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

\[
Q^{-1/2}h(\sigma) = 0 \tag{12}
\]

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 \( P(\sigma)h(\sigma) = h(\sigma) \), the condition (Equation (12)) is the same as \( h(\sigma) = 0 \) 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 \( h(\sigma) = 0 \) 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 \( h(\sigma) = 0 \). 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 \( y^* - Z^* \hat{\theta}_1 \) (Kirby & Bollen, 2009). Theorem 1 shows that it is possible that invalid MIIVs do not explain any variation in the residual \( y^* - Z^* \hat{\theta}_1 \), even though \( \hat{\theta}_1 \) 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 \( F_{\text{Sargan}} \) which is as follows.

Theorem 2. The asymptotic distribution of \( F_{\text{Sargan}} \) is a weighted sum of independent chi-square random variables with 1 degrees of freedom, if \( h(\sigma) = 0 \). The weights are the eigenvalues of \( \Pi = \varphi^{-2}Q^{-1/2}\Sigma_{w\gamma}^{-1}Q^{-1/2}Q \).

If the degrees of freedom of the overidentification test is \( L - K = 1 \), the asymptotic distribution of \( F_{\text{Sargan}} \) is \( \text{tr}[\Pi] \) 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 \( h(\sigma) \). 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.

Theorems above imply that the overidentification tests for MIIV are not useful when \( h(\sigma) = 0 \). We have also
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 \( \text{cor}(\epsilon_2, \epsilon_3) \) or \( \text{cor}(\epsilon_3, \epsilon_4) \) is omitted, there still exist many sets of valid MIIVs such that \( \theta_1 \) 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 \( \theta'_1 \) of \( \theta_1 \) such that

\[
E[\nu^* (\nu^* - z^T \theta'_1)] = 0.
\]

When the MIIV approach is used to fit a SEM model, the pseudo-true value \( \theta'_1 \) corresponds to \( y(\sigma) \) 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

\[
\Sigma_{yz} a = E(\nu^* e)
\]

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 \( \Sigma_{yz} \) and \( E(\nu^* e) \) 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 the correctly specified model. 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 \( \eta \) are mutually correlated and that \( \text{var}(\eta) \) is an invertible matrix.

5.1. Measurement Model

Suppose that \( y_1^* = \lambda_1^T \eta_1 + \lambda_2^T \eta_2 + \epsilon_1 \) in the true model. The hypothesized model assumes that \( y_1^* \) is loaded only on \( \eta_2 \), yielding a misspecified equation if \( \lambda_1 = 0 \). The scaling indicator of \( \eta_1 \) is hypothesized to satisfy \( y_2^* = \eta_2 + \epsilon_2 \), but the correct equation is \( y_2^* = \lambda_1 \eta_1 + \lambda_2 \eta_2 + \epsilon_2 \). Then, our hypothesized model to estimate \( \lambda_2 \) is in fact

\[
y_2^* = \lambda_2^T \eta_2 + \epsilon_4 - \lambda_1^T \epsilon_2 + \left( \lambda_1^T - \lambda_2^T \right) \eta_1 - \lambda_2^T \lambda_3 \eta_3.
\]

We do not exclude the possibility that some MIIVs may be correlated with a subset of \( \{\epsilon_2, \epsilon_4\} \) but the correlations are misspecified. Hence, without loss of generality, we partition the MIIVs \( \nu^* \) into \( \nu_1^* \) and \( \nu_2^* \) such that

\[
\nu_1^* = \Lambda_1 \eta_1 + \epsilon_{1,1} \quad \text{and} \quad \nu_2^* = \Lambda_2 \eta_2 + \epsilon_{1,2},
\]

where \( \eta_1 \) and \( \eta_2 \) are in fact \( \Lambda_1^T \eta_1 + \epsilon_{1,1} \) and \( \Lambda_2^T \eta_2 + \epsilon_{1,2} \) for the scaling indicator.

Proof 1. Consider estimation of Equation (15) and the MIIVs given by Equation (16). Let \( \eta_M^T = (\eta_1^T \eta_2^T \eta_3^T \eta_4^T \Lambda_1^T + \eta_2^T + \eta_3^T + \eta_4^T \Lambda_3^T) \) when \( \Lambda_3 \neq 0 \), or \( \eta_M^T = (\eta_1^T \eta_2^T \eta_4^T \Lambda_1^T + \eta_2^T) \) when \( \Lambda_3 = 0 \). Suppose that \( \Lambda_1 \) is of full column rank, that \( \Sigma_{yz} \) is of full column rank, and that \( \text{cov}(\eta_{1,2}) \) is full column rank. Then, \( h(\sigma) = 0 \) if and only if \( \lambda_1 - \Lambda_1^T \lambda_2 = 0 \), \( \Lambda_1^T \lambda_2 = 0 \), and \( \text{cov}(\epsilon_{1,1}, \epsilon_{1,2}) = \text{cov}(\epsilon_{2,1}, \epsilon_{2,2}) \lambda_2 \).

Proof 1 shows that what assumptions are needed in order for the overidentification tests to be consistent against a nonzero \( \lambda_1 \). Besides the full column rank assumptions, if the hypothesized equation \( y_2^* = \eta_2 + \epsilon_2 \) 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 \( \eta_1 \) to achieve a consistent test. The covariance matrix \( \text{cov}(\eta_{1,2}) \) can still have full column rank even though \( \eta_1 \) is not included in \( \eta_1 \) (e.g., using IVs \( y_1^*, y_2^* \)) in Figure 3a).

Because of the condition \( \text{cov}(\epsilon_{1,2}, \epsilon_{1,4}) = \text{cov}(\epsilon_{2,1}, \epsilon_{2,2}) \lambda_2 \), Proposition 1 also sheds some lights on testing the measurement error correlation. If \( \text{cov}(\epsilon_{1,2}, \epsilon_{1,4}) = \text{cov}(\epsilon_{2,1}, \epsilon_{2,2}) \lambda_2 \), then \( h(\sigma) = 0 \) if and only if \( \text{cov}(\epsilon_{1,2}, \epsilon_{1,4}) = \text{cov}(\epsilon_{2,1}, \epsilon_{2,2}) \lambda_2 \). This means that misspecification can be detected as long as some MIIVs are correlated with \( \epsilon_4 \). In contrast, if \( \text{cov}(\epsilon_{1,2}, \epsilon_{1,4}) \neq 0 \), we can always find \( \text{cov}(\epsilon_{1,1}, \epsilon_{1,4}) = \lambda_2 \) such that the linear system \( \text{cov}(\epsilon_{1,2}, \epsilon_{1,4}) = \text{cov}(\epsilon_{2,1}, \epsilon_{2,2}) \lambda_2 \) holds. In other words, there always exist “lucky” numbers such that \( h(\sigma) = 0 \) but \( \text{cov}(\epsilon_{1,2}, \epsilon_{1,4}) \neq 0 \) 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 \( y_3 = b_1 \eta_1 + b_2 \eta_2 + \epsilon_3 \), but the hypothesized model ignores \( b_1 \eta_1 \), yielding a misspecified model if \( b_1 \neq 0 \). The scaling indicators are hypothesized to satisfy \( y_2^* = \eta_2 + \epsilon_2 \) and \( y_3^* = \eta_3 + \epsilon_3 \), but the correct

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Overidentification degrees of freedom (L - K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_k )</td>
<td>20.88</td>
</tr>
<tr>
<td>( b_2 )</td>
<td>28.57</td>
</tr>
<tr>
<td>( \text{cor}(\epsilon_2, \epsilon_3) )</td>
<td>85.71</td>
</tr>
<tr>
<td>( \text{cor}(\epsilon_1, \epsilon_2) )</td>
<td>60.00</td>
</tr>
</tbody>
</table>

Table 1. Percentage of sets of MIIVs that yield \( h(\sigma) = 0 \).
correlation between some scaling indicators are correctly specified, then we must have all satisfy the conditions in Proposition 2, yielding the MIIVs may be correlated with a subset of the omitted measurement error correlations such that Equation (18) holds but the overidentification tests are inconsistent. If we further assume $\lambda_{23} = 0$, $\lambda_{32} = 0$, $\text{cov}(\varepsilon_2, \varepsilon_3) = \text{cov}(\varepsilon_1, \varepsilon_2, \varepsilon_5)$, and $\hat{A}_2$ is of full column rank, then Equation (18) reduces to $0 = \text{cov}(\hat{\eta}_2, \hat{\zeta}_3)$, 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 $\hat{\eta}_1$ and $\hat{\eta}_2$ 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 $\Theta$ is correctly specified, $\hat{\eta}_1 = (\eta_1, \eta_2, \eta_3)^T$ and $\eta_M = (\eta_1, \eta_2, \eta_3)^T$ violate the full column rank condition of $\text{cov}(\hat{\eta}_1, \eta_M^T)$ in Proposition 1. Hence, it is possible that the overidentification tests are inconsistent. 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 $2^{-1}\text{cov}(\hat{\eta}_1, \eta_M^T) = \lambda_1$ and Equation (7) corresponds to $2^{-1}\lambda_2\text{cov}(\hat{\eta}_2, \eta_M^T) = \lambda_1\text{cov}(\hat{\eta}_2, \eta_M^T)\lambda_1 + \text{cov}(\varepsilon_2, \varepsilon_4)$, where $\eta_M^T = (\eta_1, \eta_2, \eta_3)^T$, $\eta_2 = \eta_3$, $\lambda_1 = (\lambda_{12,1}, \lambda_{12,3})^T$, and $\hat{A}_2 = (\lambda_{10,3}, \lambda_{11,3})^T$.

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 $v_1^T$, and that $\hat{\eta}_1 = \eta_1$. Then, $\text{cov}(\hat{\eta}_1, \eta_M^T)$ is not of full column rank, and $\Sigma_{\varepsilon} = \lambda_1$, $\text{cov}(\hat{\eta}_1, \eta_M^T)$. Hence, $\Sigma_{\varepsilon} = \lambda_1\text{cov}(\hat{\eta}_1, \eta_M^T)\lambda_1 + \text{cov}(\varepsilon_2, \varepsilon_4)$, where $\Sigma_{\varepsilon} = (\hat{A}_2, \lambda_1)$ and $\hat{A}_2 = (\lambda_{10,3}, \lambda_{11,3})^T$.

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 \( h(\beta) = 0 \), 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 \( h(\beta) = 0 \) is equivalent to testing \( \lambda_1 = 0 \) and \( \text{cov}(e_{n,2}, e_4) = \text{cov}(e_{n,2}, e_1^2) \lambda_2 \) for the measurement model. Likewise, under the assumptions in Proposition 2, testing \( h(\beta) = 0 \) is equivalent to testing \( b_1 = 0 \), and \( \text{cov}(\Lambda_2, \eta_2, \xi_3) = \text{cov}(e_{n,2}, e_3 - e_1^2 b_1) \) for the latent regression model. Consequently, the overidentification tests are consistent against missing factor loadings and latent regression paths. However, the condition of \( \text{cov}(\eta_1, \eta_M) \) 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 pretest. 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.

Acknowledgments

We are grateful to the reviewers for their comments to improve the study.

ORCID

Shaobo Jin http://orcid.org/0000-0001-6538-3477

References


Jin et al. (2021) require correctly specified equations and valid IVs, its proof only relies on the assumption of its identity, which is $L - K = 1$. Hence, the singular value decomposition of $Q$ is $Q = d^T p u^*$, where $d > 0$ is the eigenvalue of $Q Q^T$, $p$ is a $L \times 1$ vector, and $u$ is also a $L \times 1$ vector. Likewise, the spectral decomposition of $Q Q^T$ is $Q Q^T = d d^T$, since $u u^* = 1$. It can be easily verified that the Moore–Penrose inverse $Q Q^T$ is

$$G = d^{-1} p u^*.$$  

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

$$F_m = n (S_{y y} - S_{a a} \hat{\theta})^T \Omega^{-1} (S_{y y} - S_{a a} \hat{\theta})$$

Then, $F_m - F$ becomes

$$n (S_{y y} - S_{a a} \hat{\theta})^T \frac{S_{a a}^{-1}}{\text{tr}(\Omega^{-1/2} S_{a a}^{-1/2} \Omega^{-1/2})} - \frac{-1/2}{\text{tr}(\Omega^{-1/2} p p^T \Omega^{-1/2})}$$

where the equality holds since $\Omega^{-1/2} Q Q^T \Omega^{-1/2} S_{y y} = S_{y y} - S_{a a} \hat{\theta}$.

Hence, the $F_m$ holds if

$$\Omega^{-1/2} Q Q^T \Omega^{-1/2} S_{y y} = S_{y y} - S_{a a} \hat{\theta}.$$  

By the singular decomposition of $Q$, we can obtain

$$\Omega^{-1/2} S_{y y}^{-1/2} Q = d u^T \Omega^{-1/2} S_{y y}^{-1/2} \Omega^{-1/2} p u^* = \frac{d (p^T \Omega^{-1/2} S_{y y}^{-1/2} \Omega^{-1/2} p) u u^*}{\text{tr}(\Omega^{-1/2} p p^T \Omega^{-1/2})},$$

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

$$\text{tr}(\Omega^{-1/2} S_{y y}^{-1/2} Q) = \frac{d (p^T \Omega^{-1/2} S_{y y}^{-1/2} \Omega^{-1/2} p)}{\text{tr}(\Omega^{-1/2} p p^T \Omega^{-1/2})},$$

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

To show the second half of the theorem, it suffices to show $\text{tr}(\Omega^2) = \text{tr}(\Omega^2)$. By Equation (21), we obtain

$$\text{tr}(\Omega^2) = d^2 (p^T \Omega^{-1/2} S_{y y}^{-1/2} \Omega^{-1/2} p)^2$$

since $u u^* = 1$ and $\text{tr}(u u^*) = 1$. We have already derived $\text{tr}(\Omega^2)$ in Equation (22). This completes the proof of the theorem.

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

$$\text{cov}(A_i, A_j) = \text{cov}(A_i, A_j) = 0.$$  

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

It can be easily shown that

\[
\Sigma_{\sigma} = \begin{pmatrix}
(\sigma(1,1), \sigma(1,2), \sigma(2,1), \sigma(2,2)) \\
(\sigma(1,2), \sigma(2,1))
\end{pmatrix}.
\]

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

\[
P(\sigma) = -P(\sigma) = \begin{pmatrix}
(\sigma(1,1), \sigma(1,2), \sigma(2,1), \sigma(2,2)) \\
(\sigma(1,2), \sigma(2,1))
\end{pmatrix}.
\]

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

\[
h(\sigma) = P(\sigma) = \begin{pmatrix}
(\sigma(1,1), \sigma(1,2), \sigma(2,1), \sigma(2,2)) \\
(\sigma(1,2), \sigma(2,1))
\end{pmatrix}.
\]

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

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

\[
\Lambda_1 \text{cov}(\sigma(1,1), \sigma(1,2), \sigma(2,1), \sigma(2,2)) a = \Lambda_1 \text{cov}(\sigma(1,1), \sigma(1,2)) (\Lambda_1 - \Lambda_2 \sigma).
\]

(23)

Since \( \Lambda_1 \) is full column rank, the rank of \( \Lambda_1 \text{cov}(\sigma(1,1), \sigma(1,2)) \) is the same as the rank of \( \text{cov}(\sigma(1,1), \sigma(1,2)) \) (Lemma 1). When \( \Lambda_3 \neq 0 \), Equation (23) can be expressed as

\[
\Lambda_1 \text{cov}(\sigma(1,1), \sigma(1,2), \sigma(2,1), \sigma(2,2)) a = \Lambda_1 \text{cov}(\sigma(1,1), \sigma(1,2)) (\Lambda_1 - \Lambda_2 \sigma).
\]

By the assumption that \( \text{cov}(\sigma(1,1), \sigma(1,2)) \) is of full rank, we must have \( \Lambda_1 - \Lambda_2 \sigma = 0 \), \( \Lambda_1 \sigma = 0 \), and \( a = 0 \). Consequently, Lemma 2 indicates that \( \text{cov}(\sigma(1,1), \sigma(1,2), \sigma(2,1), \sigma(2,2)) \) is of full rank, \( \Lambda_3 \neq 0 \), Equation (23) can be expressed as

\[
\Lambda_1 \text{cov}(\sigma(1,1), \sigma(1,2), \sigma(2,1), \sigma(2,2)) a = \Lambda_1 \text{cov}(\sigma(1,1), \sigma(1,2)) (\Lambda_1 - \Lambda_2 \sigma).
\]

By the assumption that \( \text{cov}(\sigma(1,1), \sigma(1,2)) \) is of full rank, we must have \( \Lambda_1 - \Lambda_2 \sigma = 0 \) and \( a = 0 \). Consequently, Lemma 2 further indicates that \( \text{cov}(\sigma(1,1), \sigma(1,2)) \) is of full rank. The proof is completed.

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

\[
\begin{align*}
\text{cov}(\Lambda_1, \sigma(1,1), \sigma(1,2), \sigma(2,1), \sigma(2,2)) a &= \text{cov}(\Lambda_1, \sigma(1,1), \sigma(1,2), \sigma(2,1), \sigma(2,2)) (\Lambda_1 - \Lambda_2 \sigma) \\
&= -\Lambda_1 \text{cov}(\sigma(1,1), \sigma(1,2)) (\Lambda_1 - \Lambda_2 \sigma)
\end{align*}
\]

and

\[
\begin{align*}
\text{cov}(\Lambda_2, \sigma(1,1), \sigma(1,2), \sigma(2,1), \sigma(2,2)) a &= \text{cov}(\Lambda_2, \sigma(1,1), \sigma(1,2), \sigma(2,1), \sigma(2,2)) (\Lambda_1 - \Lambda_2 \sigma) \\
&= -\Lambda_2 \text{cov}(\sigma(1,1), \sigma(1,2)) (\Lambda_1 - \Lambda_2 \sigma)
\end{align*}
\]

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

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

\[
\Lambda_1 \text{cov}(\sigma(1,1), \sigma(1,2), \sigma(2,1), \sigma(2,2)) a = \Lambda_1 \text{cov}(\sigma(1,1), \sigma(1,2), \sigma(2,1), \sigma(2,2)) (\Lambda_1 - \Lambda_2 \sigma)
\]
\[
\begin{pmatrix}
\hat{\lambda}_{31} + b_1 - (\Lambda^T_{31} + b_1 \lambda^T_{33}) (I + b_2 \lambda^T_{33})^{-1} (\lambda_{32} + b_2)
\hat{\lambda}_{34} - \Lambda^T_{33} (I + b_2 \lambda^T_{33})^{-1} (\lambda_{32} + b_2)
- \Lambda^T_{33} (I + b_2 \lambda^T_{33})^{-1} (\lambda_{32} + b_2)
\hat{\lambda}_{36}
\end{pmatrix}.
\]

Since \(\hat{\Lambda}_1\) is of full column rank, the rank of \(\hat{\Lambda}_1 \text{cov}(\hat{\eta}_1, \eta^T_{id})\) is the same as the rank of \(\text{cov}(\hat{\eta}_1, \eta^T_{id})\) (Lemma 1). By the assumption that \(\text{cov}(\hat{\eta}_1, \eta^T_{id})\) is of full column rank, we must have \(\hat{\lambda}_{31} + b_1 = (\Lambda^T_{31} + b_1 \lambda^T_{33}) (I + b_2 \lambda^T_{33})^{-1} (\lambda_{32} + b_2), \quad \hat{\lambda}_{34} = \Lambda^T_{33} (I + b_2 \lambda^T_{33})^{-1} (\lambda_{32} + b_2), \quad \hat{\lambda}_{36} = 0, \quad \text{and} \quad a = 0\).

Consequently, Lemma 3 indicates that
\[
\text{cov}\left\{ \hat{\lambda}_2 \hat{\eta}_2, \zeta_3 \mid 1 - \lambda^T_{33} (I + b_2 \lambda^T_{33})^{-1} (\lambda_{32} + b_2) \right\}
= -\text{cov}\{e_{3,1}, e_3 - \epsilon^T (I + b_2 \lambda^T_{33})^{-1} (\lambda_{32} + b_2)\}.
\]

This completes the proof. \(\square\)