Testing for Identification in Potentially Misspecified Linear GMM*

Frank Kleibergen† Zhaoguo Zhan‡

April 17, 2025

Abstract

The difference between the population analogs of the traditional Jacobian rank statistic and the J-statistic provides an appropriate identification measure for the pseudo-true values of the continuous updating and two-stage estimators in the misspecified linear generalized method of moments, whilst the population analog of the traditional Jacobian rank statistic alone does not. The sample analog of the identification measure leads to a likelihood ratio no-identification test. We construct a conditional critical value function for the test in the homoskedastic setting which controls the size. Applying the test to empirical asset pricing where misspecification is common, we find that no-identification is often not rejected at the 5% significance level while the Jacobian rank test wrongly signals the opposite.

JEL Classification: C1, C3

Keywords: misspecification, weak identification, conditional likelihood ratio test

*We thank particpants of the 34-th EC^2 meeting in Manchester, the 2024 IAAE meeting in Thessaloniki, the workshop "Econometrics in Rio 2024: Integrating Theory and Practice in Econometrics", the Aarhus Workshop in Econometrics 1, and seminar participants at Brown, Cornell, LSE, PennState, Penn, Princeton, Tilburg, UCL and Warwick for comments and discussion.

†Email: f.r.kleibergen@uva.nl. Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands.

‡Email: zzhan@kennesaw.edu. Department of Economics, Finance and Quantitative Analysis, Coles College of Business, Kennesaw State University, GA 30144, USA.

1 Introduction

Many widely used econometric models, such as the linear instrumental variables (IV) regression, linear asset pricing, dynamic panel data and New-Keynesian Phillips curve models, are analyzed using the generalized method of moments (GMM) of Hansen (1982). Recently awareness has risen that structural parameters in popular models estimated using GMM might be weakly identified, which implies that traditional inference methods are unreliable; see, *e.g.*, Staiger and Stock (1997), Dufour (1997), and Stock and Wright (2000). Inference methods have therefore been developed which remain reliable under weak identification; see, *e.g.*, Kleibergen (2002, 2005), Moreira (2003), Andrews and Cheng (2012), and Andrews and Mikusheva (2016). Weak identification in correctly specified GMM occurs when the Jacobian is relatively close to a reduced rank value. Tests for a reduced rank value of the Jacobian are therefore commonly employed to determine the strength of identification of the structural parameters of interest; see, *e.g.*, Cragg and Donald (1997), Robin and Smith (2000), and Kleibergen and Paap (2006).

The GMM toolkit of Hansen (1982) has foremost been developed for analyzing correctly specified models, *i.e.*, models for which there is a, so-called, true value of the structural parameters at which the population moments are exactly zero. Many empirical models estimated using GMM, or perhaps even every (over-identified) model with more moment equations than structural parameters, are yet to some extent misspecified. For these models, there is no longer a true value of the structural parameters at which the population moment conditions hold exactly. The earlier literature on misspecified models primarily focusses on the consequences of the inconsistency of estimators for the true value of the structural parameters; see, *e.g.*, Maasoumi (1990), and Maasoumi and Phillips (1982). Applied researchers, however, mostly just proceed with interpretating the estimated structural parameters that result from misspecified models. The population analogs of these estimators are then referred to as pseudo-true values which are defined as the minimizers of the population analogs of the sample objective functions; see, *e.g.*, Kan, Robotti, and Shanken (2013). Different objective functions lead to distinct pseudo-true values under misspecification, while for correctly specified models, true and pseudo-true values coincide. Inference methods for analyzing the pseudo-true values have therefore been developed by, amongst others: Hall and Inoue (2003), Gospodinov, Kan, and Robotti (2014), Lee (2018), and Hansen and Lee (2021).

Identification issues for the pseudo-true values similarly play out in misspecified models. We show that the identification condition in misspecified over-identified models differs from the one in correctly specified models. For linear moment conditions, identification of the pseudo-true value of the structural parameters is reflected by the difference between the population analogs of the traditional rank statistic for identification and the over-identification J-statistic. Because the former results from constrained optimization of the objective function involved in the latter, their difference, which is our proposed identification measure, is non-negative by construction. Identification of the pseudo-true value fails when this identification measure equals zero. For correctly specified models, the population analog of the J-statistic equals zero, so the identification measure reduces to the traditional one which just relies on the population analog of the Jacobian rank statistic.

Because the traditional identification measure based on the Jacobian does not properly reflect identification in misspecified linear models estimated by GMM, the common practice of using its sample analog to test for (no) identification is inappropriate, *i.e.*, it does not test the relevant no-identification hypothesis. In contrast, we develop an appropriate (quasi) likelihood ratio (LR) no-identification test for misspecified linear models. Its test statistic equals the difference between the sample analogs of the traditional rank statistic and the J-statistic. For a boundary setting of no-identification, we construct a conditional critical value function based on homoskedasticity which implies a Kronecker product structure (KPS) of the joint covariance matrix of the sample moment vector and its Jacobian.

We develop the conditional critical value function for a sequence of gradually more challenging (no) identification testing problems. For ease of exposition, we start from the setting of one structural parameter and a known covariance matrix. It allows us to specify the LR test statistic as a function of two of the three elements of an appropriate specification of the maximal invariant while its third element provides an approximately independent conditioning statistic. This is similar in spirit to the conditional LR test of Moreira (2003) albeit that the conditioning statistic and the null distribution under which the conditional critical value function is computed differ. Hereafter, we provide empirically important generalizations which incorporate covariance matrix estimators and multiple structural parameters. Because misspecification allows for population moments which are non-zero at the pseudo-true value of the structural parameters, an accurate approximation of the conditional distribution of the LR statistic has to take the estimation error resulting from the covariance matrix estimators into account; see, *e.g.*, Maasoumi and Phillips (1982), Hall and Inoue (2003), Gospodinov, Kan, and Robotti (2014), Lee (2018), and Hansen and Lee (2021). Alongside the dependence on the conditioning statistic, the conditional critical value function that we provide for empirical settings with multiple structural parameters therefore depends on the sample size at hand and a few consistently estimable nuisance parameters.

The remaining part of the paper is organized as follows. Section 2 introduces the appropriate identification measure for possibly misspecified over-identified linear GMM. Section 3 emphasizes its empirical importance by showcasing eight well known studies from the asset pricing literature. Section 4 develops the conditional LR no-identification test along the lines alluded to previously. We show that the test has good size and power properties, and also compare it with existing tests that test part of the no-identification hypothesis or just use elements of the LR statistic. While these existing tests can have superior power for specific settings, we show that they are inadequate in other settings which overall renders them inappropriate for testing the no-identification hypothesis. Section 4 also applies the LR test to the Fama-French (1993) three-factor model using data from Lettau, Ludvigson, and Ma (2019). The misspecification J-test signals that the three-factor model is misspecified, because the J-test rejects correct specification at tiny significance levels for both the specification that incorporates the zero-beta return and the one without. The traditional identification rank test indicates strong identification for both of these specifications. On the other hand, the appropriate LR no-identification test just rejects no-identification with 5% significance when the zero-beta return is not incorporated while it does not when the zero-beta return is incorporated. Given the importance of the Fama-French

(1993) three-factor model, our empirical application illustrates the relevance of using the appropriate identification test. The fifth section draws some conclusions and provides extensions for future work. Technical details and additional empirical findings are relegated to the Appendix.

2 Identification in misspecified linear GMM

2.1 Over-identified linear moment equations

We conduct inference on a k_f-dimensional moment vector $\mu_f(\theta)$, which is a continuous function of the m-dimensional parameter vector θ. The parameter vector θ is over-identified by the linear moment equations, so $k_f > m$. The linear moment equations for GMM are specified accordingly:

$$\begin{aligned} E_X(f(\theta, X_i)) &= \mu_f(\theta) \\ &= \mu_f(0) + J(0)\theta, \end{aligned} \tag{1}$$

with $J(0) = \frac{\partial}{\partial\theta'}\mu_f(\theta)$. Many widely used econometric models, like, for example, the linear IV regression model, the linear factor model, and the linear dynamic panel data model, accord with this setting. The sample moment vector $\hat{\mu}_f(\theta)$ then just depends on the estimators of $\mu_f(0)$, $\hat{\mu}_f(0)$, and the Jacobian $J(0)$, $\hat{J}(0)$, whose joint convergence when the sample size N increases results from Assumption 1.

Assumption 1: *The joint limit behavior of $\hat{\mu}_f(0)$ and $\hat{J}(0)$ is described by:*

$$\begin{aligned} \sqrt{N}\left(\begin{pmatrix} \hat{\mu}_f(0) \\ \text{vec}(\hat{J}(0)) \end{pmatrix} - \begin{pmatrix} \mu_f(0) \\ \text{vec}(J(0)) \end{pmatrix}\right) &\underset{d}{\rightarrow} \begin{pmatrix} \psi_\mu \\ \psi_J \end{pmatrix} \\ \begin{pmatrix} \psi_\mu \\ \psi_J \end{pmatrix} &\sim \mathbf{N}(0, V), \end{aligned} \tag{2}$$

where the covariance matrix V of the limit behavior of $(\hat{\mu}_f(0)' \vdots$ vec$(\hat{J}(0))')'$ reads:

$$
\begin{aligned}
V &= \lim_{N\to\infty} E\left(N\begin{pmatrix}\hat{\mu}_f(0)-\mu_f(0)\\ \mathrm{vec}(\hat{J}(0))-\mathrm{vec}(J(0))\end{pmatrix}\begin{pmatrix}\hat{\mu}_f(0)-\mu_f(0)\\ \mathrm{vec}(\hat{J}(0))-\mathrm{vec}(J(0))\end{pmatrix}'\right)\\
&= \begin{pmatrix}V_{\mu\mu} & V_{\mu J}\\ V_{J\mu} & V_{JJ}\end{pmatrix}, \qquad (3)
\end{aligned}
$$

with $V_{\mu\mu}$, $V_{\mu J} = V'_{J\mu}$ and V_{JJ} resp. $k_f \times k_f$, $k_f \times mk_f$ and $mk_f \times mk_f$ dimensional matrices.[1]

Assumption 1 is satisfied under mild conditions; see, *e.g.*, White (1984).[2]

2.2 Identification and misspecification measures

When the moment equations are correctly specified, there is a unique, so-called, true value of θ, say θ_0, at which the population moment conditions exactly hold:

$$\mu_f(\theta_0) = \mu_f(0) + J(0)\theta_0 = 0. \qquad (4)$$

The moment vector $\mu_f(0)$ is then spanned by its Jacobian $J(0)$, so $(\mu_f(0) \vdots J(0))$ is a $k_f \times (m+1)$ dimensional matrix which is at most of rank m :

$$(\mu_f(0) \vdots J(0)) = J(0)(-\theta_0 \vdots I_m) \Leftrightarrow (\mu_f(0) \vdots J(0))\binom{1}{\theta_0} = 0. \qquad (5)$$

The true value of θ, θ_0, is identified when $J(0)$ is a full rank matrix, so the traditional identification measure is based on a rank test which tests for a lower rank value of $J(0)$, like, for example, the Cragg and Donald (1997) rank test. We therefore define the identification measure denoted by IS (identification strength) accordingly.

[1] The covariance matrix V is not required to be of full rank. Positive semi-definite values of V can, for example, occur for the moment equations resulting from linear dynamic panel data models.

[2] For example, consider the linear IV regression model with $\mu_f(\theta) = E(Z_iY_i) - E(Z_iX_i')\theta$, where Z_i is the k_f-dimensional vector of instrumental variables, X_i is the m-dimensional endogenous variables, Y_i is the scalar endogenous variable. In this case, we have $\mu_f(0) = E(Z_iY_i)$, $J(0) = -E(Z_iX_i')$. Assumption 1 then just imposes that a central limit theorem applies to $\hat{\mu}_f(0) = \frac{1}{N}\sum_{i=1}^N Z_iY_i$ and the vectorization of $\hat{J}(0) = -\frac{1}{N}\sum_{i=1}^N Z_iX_i'$.

Definition 1: *The IS identification measure equals the population value of the Cragg and Donald (1997) rank statistic which tests for a reduced rank value of* $J(0)$*, see also Kleibergen (2007) and Kleibergen and Mavroeidis (2009):*

$$\begin{aligned} \text{IS} :=& \ \min_{\varphi \in \mathbb{R}^{m-1}} Q_{IS}(\varphi) \\ Q_{IS}(\varphi) =& \ \binom{1}{\varphi}' J(0)' \left[\left(\binom{1}{\varphi} \otimes I_{k_f} \right)' V_{JJ} \left(\binom{1}{\varphi} \otimes I_{k_f} \right) \right]^{-1} J(0) \binom{1}{\varphi} \\ =& \ \min_{A \in \mathbb{R}^{k_f \times (m-1)}} \left[\text{vec} \left(J(0) - A \left(I_{m-1} \vdots - \varphi \right) \right) \right]' V_{JJ}^{-1} \\ & \left[\text{vec} \left(J(0) - A \left(I_{m-1} \vdots - \varphi \right) \right) \right]. \end{aligned} \tag{6}$$

When the moment equations are misspecified, there is no longer a true value at which the moment equations hold exactly and interest is on the, so-called, pseudo-true value which is the minimizer of the population objective function. Different population objective functions lead to distinct pseudo-true values, while we focus on the pseudo-true value of the continuous updating estimator (CUE). We do so for the invariance properties of CUE, because of which inference that results from weak identification robust procedures is centered around it. Later we also discuss the popular two-stage estimator and show that its identification condition corresponds with the one for the CUE, which is counter to the condition one obtains at casual inspection.

The pseudo-true value of the CUE, θ^*, is defined as the minimizer of the continuous updating population objective function:

$$Q_{CUE}(\theta) = \ \mu_f(\theta)' \left[\binom{I_{k_f}}{(\theta \otimes I_{k_f})}' V \binom{I_{k_f}}{(\theta \otimes I_{k_f})} \right]^{-1} \mu_f(\theta), \tag{7}$$

so

$$\theta^* := \ \arg \ \min_{\theta \in \mathbb{R}^m} Q_{CUE}(\theta). \tag{8}$$

Next, we define the misspecification measure as the minimum of the CUE population objective function, which is the population analog of the J-statistic.

Definition 2: *The MISS misspecification measure equals the minimal value of the CUE population objective function (7):*

$$\text{MISS} := \min_{\theta \in \mathbb{R}^m} Q_{CUE}(\theta). \tag{9}$$

When the moment equations are correctly specified, MISS equals zero but not so in case of misspecified moment equations.[3]

Theorem 1: *MISS is at most as large as IS:*

$$\text{MISS} \leq \text{IS}. \tag{10}$$

Proof. MISS is invariant with respect to the specification of the moment equation, so the same value results when we use the moment equation:

$$\eta_f(\alpha, \varphi) = J(0)_1 + \mu_f(0)\alpha + J(0)_2\varphi, \tag{11}$$

with $J(0) = (J(0)_1 \vdots J(0)_2)$, $J(0)_1 : k_f \times 1$, $J(0)_2 : k_f \times (m-1)$, which results from transforming the parameters:

$$\alpha = \theta_1^{-1}, \ \varphi = \begin{pmatrix} \theta_2 \\ \vdots \\ \theta_m \end{pmatrix} \theta_1^{-1}, \tag{12}$$

with $\theta = (\theta_1, \theta_2, ..., \theta_m)'$. The transformation assumes that $\theta_1 \neq 0$ which holds in sample with probability one, so

$$\begin{aligned} \text{MISS} = & \ \min_{\theta \in \mathbb{R}^m} Q_{CUE}(\theta) \\ = & \ \min_{\alpha \in \mathbb{R}, \ \varphi \in \mathbb{R}^{m-1}} Q_{CUE}(\alpha, \varphi) \\ \leq & \ \min_{\varphi \in \mathbb{R}^{m-1}} Q_{CUE}(\alpha = 0, \varphi) \\ = & \ \text{IS}. \end{aligned} \tag{13}$$

[3] As follows from Theorem 1, MISS can, however, still be equal to zero when IS is zero despite that there is no value of θ where the moment equation holds exactly.

Note that (11) nests the $m = 1$ case, for which it reduces to

$$\eta_f(\alpha) = J(0) + \mu_f(0)\alpha, \tag{14}$$

and the proof above similarly holds. ■

The MISS at most as large as IS inequality in Theorem 1 implies that the identification condition of the pseudo-true value of the CUE in misspecified GMM is more stringent than the identification condition of the true value in correctly specified GMM.

Corollary 1: *When MISS equals IS:*

$$\begin{aligned} \text{MISS} = &\ \ \text{IS} && \Leftrightarrow \\ \alpha^* = &\ \ \arg\ \min_{\alpha\in\mathbb{R}} \min_{\varphi\in\mathbb{R}^{m-1}} Q_{CUE}(\alpha,\varphi) = 0 && \Leftrightarrow \\ \theta_1^* = &\ \ \tfrac{1}{\alpha^*} = \infty, \end{aligned} \tag{15}$$

which implies that the pseudo-true value of the CUE, θ^*, is not identified.

Corollary 1 shows that IS has to exceed MISS for the pseudo-true value of the CUE to be identified. When the moment equations are misspecified and

$$\text{MISS} > 0, \tag{16}$$

we therefore need

$$\text{IS} > \text{MISS} > 0 \tag{17}$$

for the pseudo-true value of the CUE to be identified.

Corollary 2: *Because MISS is zero when the moment equations are correctly specified, an appropriate GMM identification measure which applies to both misspecified and correctly specified linear moment equations is:*

$$\text{IS} - \text{MISS} > 0. \tag{18}$$

Since the CUE of α equal to zero is a measure zero event, in sample the inequality in Theorem 1 is a strict one, as stated in Corollary 3 below.

Corollary 3: *The sample value of MISS, $\widehat{MISS}$, is always strictly smaller than the sample value of IS, $\widehat{IS}$, so:*

$$\widehat{\mathrm{IS}} - \widehat{\mathrm{MISS}} > 0 \text{ with probability one.} \tag{19}$$

Corollary 3 shows that just using $\widehat{\mathrm{IS}}$ to determine the identification strength overstates the identification strength when the linear moment equations are misspecified. Many commonly used empirical models have misspecified moment equations, so their identification strength is overstated when using $\widehat{\mathrm{IS}}$ as the measure of identification.

3 Empirical $\widehat{\mathbf{IS}} - \widehat{\mathbf{MISS}}$ identification measures

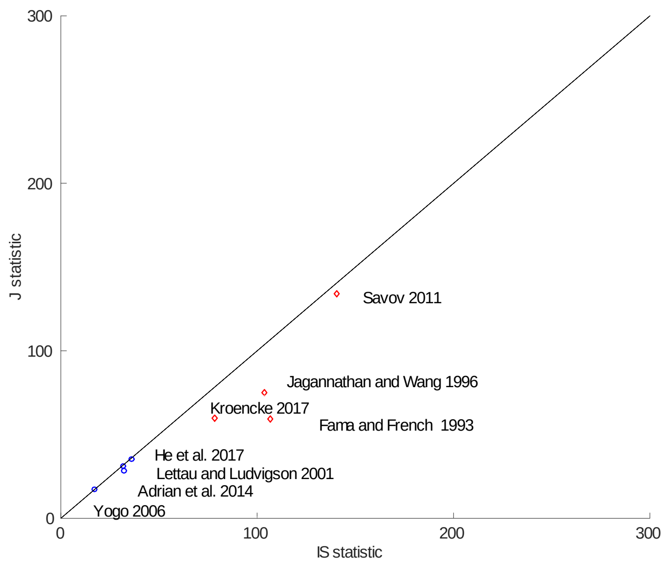
To emphasize the importance of $\widehat{\mathrm{IS}} - \widehat{\mathrm{MISS}}$ to signal identification issues instead of just $\widehat{\mathrm{IS}}$, Figure 1 shows a scatter plot of $\widehat{\mathrm{MISS}}$ (=J-statistic) and $\widehat{\mathrm{IS}}$ for eight well known specifications of linear asset pricing models: Fama and French (1993), Jagannathan and Wang (1996), Lettau and Ludvigson (2001), Yogo (2006), Savov (2011), Adrian, Etula, and Muir (2014), Kroencke (2017), and He, Kelly, and Manela (2017).[4]

The Appendix provides the specifications of the above linear asset pricing models, as well as the detailed expressions of $\widehat{\mathrm{MISS}}$ and $\widehat{\mathrm{IS}}$ for gauging the identification of risk premia (*i.e.*, the parameters of interest).

Because $\widehat{\mathrm{MISS}} < \widehat{\mathrm{IS}}$, all scatter points in Figure 1 are below the 45-degree line, but their proximity to it is striking. For all points, the distance to the 45-degree line measured by $\widehat{\mathrm{IS}} - \widehat{\mathrm{MISS}}$, is much smaller than the value of $\widehat{\mathrm{IS}}$. Figure 1 overall shows that $\widehat{\mathrm{IS}}$ overstates the identification strength, so it is important to use $\widehat{\mathrm{IS}} - \widehat{\mathrm{MISS}}$ as the identification measure.

[4] We thank the authors of Jagannathan and Wang (1996), Yogo (2006), Lettau and Ludvigson (2001), Savov (2011), and Kroencke (2017) for sharing their data. For the models of Fama and French (1993), Adrian, Etula, and Muir (2014), and He, Kelly, and Manela (2017), we use the extended data of risk factors and test assets as in Lettau, Ludvigson, and Ma (2019).

Figure 1: Scatter plot of MISS (=J) and IS statistics for different specifications.



Notes: Figure 1 shows MISS and IS statistics for eight specifications of linear asset pricing models. Their associated factors include Fama and French (1993): market, SMB, and HML; Jagannathan and Wang (1996): market, corporate bond yield spread, and per capita labor income growth; Lettau and Ludvigson (2001): consumption growth, (lagged) consumption wealth ratio and their interaction; Yogo (2006): market, durable and nondurable consumption growth; Savov (2011): garbage growth; Adrian, Etula, and Muir(2014): leverage; Kroencke (2017): unfiltered consumption growth; He, Kelly, and Manela (2017): market and the banking equity-capital ratio. All specifications incorporate the zero-beta return. For detailed descriptions of the risk factors and test assets, we refer to the published articles.

Table 1 illustrates the effect on identification and misspecification from imposing further restrictions. The left hand side of Table 1 contains the estimates of MISS and IS shown in Figure 1, which all incorporate the, so-called, zero-β return, λ_0, while the right hand side shows the counterparts when we impose the restriction that the zero-β return, λ_0, is zero, as described in the Appendix. Imposing restrictions leads to more misspecification, so it increases $\widehat{\text{MISS}}$, but also improves identification, so $\widehat{\text{IS}}$ rises as well. The overall net effect on the identification strength is, however, reflected by $\widehat{\text{IS}} - \widehat{\text{MISS}}$, which therefore incorporates the tradeoff of sacrificing misspecification to

Table 1: **MISS and IS statistics**

Notes: Panel A contains the MISS and IS statistics from Figure 1, for which the zero-beta return, indicated by λ_0, is incorporated. In Panel B, the zero-beta return is removed so $\lambda_0 = 0$. Significance at 1%, ***; 5%, **; 10%, *.

	A. Impose $\lambda_0 = 0$: No		B. Impose $\lambda_0 = 0$: Yes	
	$\widehat{\text{MISS}}$	$\widehat{\text{IS}}$	$\widehat{\text{MISS}}$	$\widehat{\text{IS}}$
Fama and French (1993)	59.34***	106.81***	87.47***	974.39***
Jagannathan and Wang (1996)	75.07	103.54	86.46	103.56
Lettau and Ludvigson (2001)	31.11*	31.75*	37.15**	40.90**
Yogo (2006)	17.14	17.34	19.42	19.60
Savov (2011)	134.27***	140.68***	268.60***	296.78***
Adrian, Etula, and Muir (2014)	28.42	31.97	30.41	42.03**
Kroencke (2017)	59.84***	78.47***	60.03***	102.77***
He, Kelly, and Manela (2017)	35.32**	35.88**	44.44***	59.74***

improve identification. The right hand side of Table 1 shows that this clearly occurs for the Fama and French (1993) application. The net effect on $\widehat{\text{IS}} - \widehat{\text{MISS}}$ for Fama and French (1993) shows that the additional misspecification is outweighted by the improved identification, but this is much less so for the other specifications presented in Table 1. For these other specifications, removing the zero-β return has little effect on the identification of the pseudo-true value of the risk premia in the studied linear asset pricing models.

4 Conditional LR no-identification test

In this section, we propose $\widehat{\text{IS}} - \widehat{\text{MISS}}$ as a conditional (quasi-) likelihood ratio (LR) test statistic for no-identification in linear GMM. We do so in a sequence of five steps, which correspond to Subsections 4.1-4.5, respectively:

1. We characterize the data generating process (DGP) parameter setting which leads to equality of IS and MISS.

2. We show the implications of the no-identification DGP parameter setting for the pseudo-true values of the CUE and two-stage estimator.

3. For a single structural parameter, $\widehat{\text{IS}} - \widehat{\text{MISS}}$ is the (quasi-) LR statistic to test for a zero value of α in (14). Using the maximal invariant, we obtain a conditioning statistic which is approximately independent of the LR statistic. It enables us to compute a conditional critical value function for the LR test of no-identification with correct size.

4. We incorporate the randomness of the covariance matrix estimator in the conditional critical value function.

5. We extend to multiple structural parameters in linear moment equations.

4.1 DGP parameter setting for IS = MISS

To obtain the parameter setting of the DGP for equality of IS and MISS, we assume a Kronecker product structure (KPS) of the joint covariance matrix V used in Assumption 1. We later discuss why it is complicated to do so for more general covariance matrix structures.

Assumption 2: *The covariance matrix V has a Kronecker product structure:*

$$V = \Omega \otimes \Sigma, \tag{20}$$

with $\Omega = \begin{pmatrix} \omega_{\mu\mu} & \vdots & \omega_{\mu J} \\ \omega_{J\mu} & \vdots & \Omega_{JJ} \end{pmatrix}$, $\omega_{\mu\mu}$, $\omega_{J\mu} = \omega'_{\mu J}$, Ω_{JJ}, Σ *resp.* 1×1, $m \times 1$, $m \times m$ *and* $k_f \times k_f$ *dimensional matrices.*

The KPS of the covariance matrix results, for example, under homoskedasticity of the errors in linear IV regression and linear factor models.

Under the KPS covariance matrix from Assumption 2, the CUE population objective function (7) is a ratio of quadratic forms:

$$Q_{CUE}(\theta) = \frac{\binom{1}{\theta}'(\mu_f(0) \vdots J(0))'\Sigma^{-1}(\mu_f(0) \vdots J(0))\binom{1}{\theta}}{\binom{1}{\theta}'\Omega\binom{1}{\theta}}, \tag{21}$$

so its minimal value, MISS, equals the smallest root of the characteristic polynomial:

$$\left|\lambda\Omega - (\mu_f(0) \vdots J(0))'\Sigma^{-1}(\mu_f(0) \vdots J(0))\right| = 0. \tag{22}$$

Similary, IS is the smallest root of the characteristic polynomial:

$$|\tau\Omega_{JJ} - J(0)'\Sigma^{-1}J(0)| = 0. \tag{23}$$

It is convenient to analyze these roots using the orthogonalized and normalized components a and C defined below.

Definition 3: *Under Assumptions 1 and 2, let* $\mu_J(0) = \mu_f(0) - J(0)\Omega_{JJ}^{-1}\omega_{J\mu}$, *and* $\omega_{\mu\mu.J} = \omega_{\mu\mu} - \omega_{\mu J}\Omega_{JJ}^{-1}\omega_{J\mu}$. *Define* a *and* C *as the normalized counterparts of* $\mu_J(0),\ J(0)$, *respectively:*

$$\begin{gathered} a := \sqrt{N}\Sigma^{-\frac{1}{2}}\mu_J(0)\omega_{\mu\mu.J}^{-\frac{1}{2}},\ C := \sqrt{N}\Sigma^{-\frac{1}{2}}J(0)\Omega_{JJ}^{-\frac{1}{2}}, \\ \Updownarrow \\ \mu_J(0) = \tfrac{1}{\sqrt{N}}\Sigma^{\frac{1}{2}}a\omega_{\mu\mu.J}^{\frac{1}{2}},\ J(0) = \tfrac{1}{\sqrt{N}}\Sigma^{\frac{1}{2}}C\Omega_{JJ}^{\frac{1}{2}}. \end{gathered} \tag{24}$$

The expressions above imply a drifting sequence of the parameters akin to the ones used in weak instrument asymptotics; see, *e.g.*, Staiger and Stock (1997). The specifications for MISS and IS then simplify to the following:

- MISS is the smallest characteristic root of

$$\left|\lambda I_{m+1} - \tfrac{1}{N}(a \vdots C)'(a \vdots C)\right| = 0 \ \Leftrightarrow\ \left|\lambda I_{m+1} - \tfrac{1}{N}\begin{pmatrix} a'a & a'C \\ C'a & C'C \end{pmatrix}\right| = 0. \tag{25}$$

- IS is the smallest characteristic root of

$$\left|\tau I_m - \tfrac{1}{N}C'C\right| \ = 0. \tag{26}$$

For the single structural parameter setting with $m = 1$, it is straightforward to derive the explicit expressions for MISS and IS. Corollary 4 thus relates the equality of IS and MISS to the expressions based on a and C.

Corollary 4: *For the single structural parameter setting, $m = 1$, equality of IS and MISS is equivalent with:*

$$\text{IS} = \text{MISS} \quad \Leftrightarrow \quad \begin{cases} C'C \leq a'a \\ C'a = 0 \end{cases}. \tag{27}$$

4.2 No-identification DGP and pseudo-true values

Next, we illustrate the implications of the no-identification DGP parameter setting ($C'C \leq a'a$, $C'a = 0$) for the population objective functions and pseudo-true values of the CUE and two-stage estimators for a single structural parameter, $m = 1$.

Figure 2 shows the population objective function and accompanying contour lines for the CUE in a setting with $a'a = 10$ and $C'a = 0$ for varying values of $C'C$, while Figure 3 does so for the two-stage estimator. Under Assumptions 1 and 2, the population objective function of the two-stage estimator is:

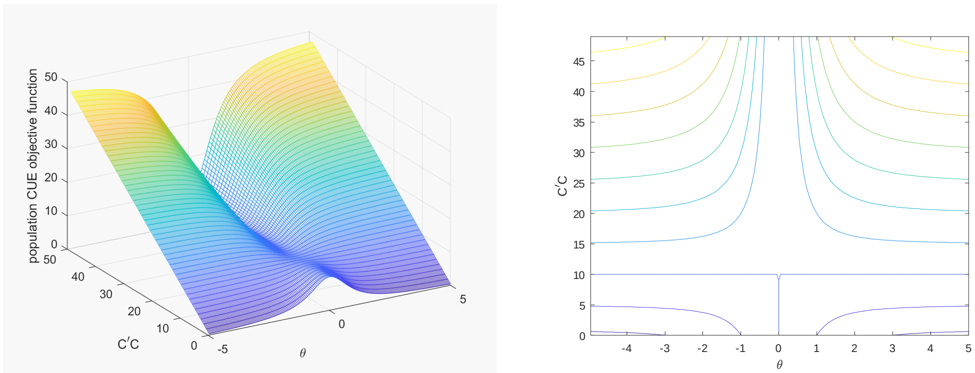
$$Q_{2s}(\theta) = \binom{1}{\theta}'(\mu_f(0)\vdots J(0))'\Sigma^{-1}(\mu_f(0)\vdots J(0))\binom{1}{\theta} \tag{28}$$

with θ^*_{2s} the resulting pseudo-true value of the two-stage estimator:

$$\begin{aligned} \theta^*_{2s} &= \arg\min_{\theta\in\mathbb{R}^m} Q_{2s}(\theta) \\ &= -(J(0)'\Sigma^{-1}J(0))^{-1}J(0)'\Sigma^{-1}\mu_f(0). \end{aligned} \tag{29}$$

The shape of the CUE population objective function in Figure 2 changes dramatically around $C'C = a'a = 10$. At the value of ten, the contour lines indicate that the objective function is flat; while above ten, it is increasing away from zero, and below ten it is decreasing away from zero. It all shows that for $C'a = 0$, the pseudo-true value of the CUE is identified when $C'C$ exceeds $a'a$, so IS > MISS, and not identified when $C'C$ is smaller than or equal to $a'a$, so IS = MISS. All these findings are thus consistent with Corollary 4.

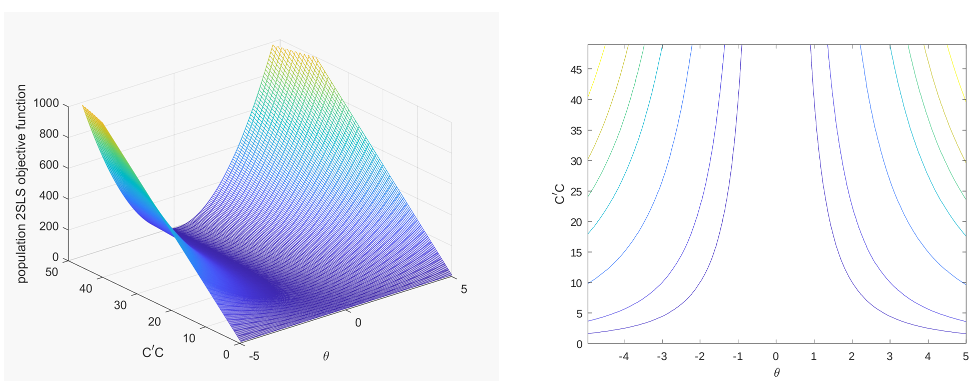
Figure 2: CUE population objective function $Q_{CUE}(\theta)$ for $a'a = 10$ and varying values of $C'C$, $C'a = 0$.



Panel 2.1. $Q_{CUE}(\theta)$ Panel 2.2. Contour lines of $Q_{CUE}(\theta)$

In contrast, the population objective function of the two-stage estimator in Figure 3 has a well defined minimum for all values of $C'C$. It thus does not show any effect of the misspecification on the identification of the pseudo-true value of the two-stage estimator, which is only not identified when $C'C = 0$, *i.e.*, the population objective function is flat at $C'C = 0$.

Figure 3: Two-stage population objective function $Q_{2S}(\theta)$ for $a'a = 10$ and varying values of $C'C$, $C'a = 0$.



Panel 3.1: $Q_{2s}(\theta)$ Panel 3.2: Contour lines of $Q_{2s}(\theta)$

The closed-form expressions for the pseudo-true values of the two-stage estimator and CUE when expressed as functions of a and C similarly show the identification issues (see also Andrews (2019)).

- Two-stage estimator:

$$\begin{aligned}
\theta_{2s}^{*} &= -(J(0)'\Sigma^{-1}J(0))^{-1}J(0)'\Sigma^{-1}\mu_f(0) \\
&= -\Omega_{JJ}^{-1}\omega_{J\mu} - (J(0)'\Sigma^{-1}J(0))^{-1}J(0)'\Sigma^{-1}\mu_J(0) \\
&= -\Omega_{JJ}^{-1}\omega_{J\mu} - \Omega_{JJ}^{-\frac{1}{2}}(C'C)^{-1}C'a\omega_{\mu\mu.J}^{\frac{1}{2}} \\
&= -\Omega_{JJ}^{-1}\omega_{J\mu} \qquad \text{when IS} = \text{MISS} \ > 0.
\end{aligned} \tag{30}$$

- CUE (k-class notation), see, *e.g.*, Hausman (1983):

$$\begin{aligned}
\theta_{CUE}^{*} &= -(J(0)'\Sigma^{-1}J(0) - \text{MISS} \times \Omega_{JJ})^{-1}(J(0)'\Sigma^{-1}\mu_f(0) - \text{MISS} \times \omega_{J\mu}) \\
&= -\Omega_{JJ}^{-1}\omega_{J\mu} - \Omega_{JJ}^{-\frac{1}{2}}(C'C - N \times \text{MISS} \times I_m)^{-1}C'a\omega_{\mu\mu.J}^{\frac{1}{2}} \\
&= \text{not identified when IS} = \text{MISS} = \tfrac{1}{N}C'C \text{ (zero divided by zero).}
\end{aligned} \tag{31}$$

While the pseudo-true value of the two-stage estimator under IS $=$ MISS > 0 seems to indicate that it is identified, (30) also shows that all values of a and C with $C'a = 0$ and $a'a \geq C'C$ lead to the same pseudo-true value $-\Omega_{JJ}^{-1}\omega_{J\mu}$. This hints that there are also identification issues for the pseudo-true value of the two-stage estimator. These identification issues become more apparent from the limit behavior of the sample analog of $J(0)'\Sigma^{-1}\mu_J(0)$, which is an important component of the two-stage estimator as shown by (30). Theorem 2 therefore provides the limit behavior of $\hat{J}(0)'\Sigma^{-1}\hat{\mu}_J(0)$, the sample analog of $J(0)'\Sigma^{-1}\mu_J(0)$.

Theorem 2: *For θ^* the pseudo-true value of the CUE, $D(\theta^*)'\Sigma^{-1}\mu_f(\theta^*) = 0$ with $D(\theta) = J(0) - \mu_f(\theta)\frac{\omega_{\mu J}(\theta)}{\omega_{\mu\mu}(\theta)}$, $\omega_{\mu J}(\theta) = \omega_{\mu J} + \theta'\Omega_{JJ}$, $\omega_{\mu\mu}(\theta) = \omega_{\mu\mu} + 2\omega_{\mu J}\theta + \theta'\Omega_{JJ}\theta$, under (24) and $m = 1$, the limit behavior of the normalized sample analog of $J(0)'\Sigma^{-1}\mu_J(0)$, $\hat{J}(0)'\Sigma^{-1}\hat{\mu}_J(0)$, is:*

$$
\begin{aligned}
& N\Omega_{JJ}^{-\frac{1}{2}}\hat{J}(0)'\Sigma^{-1}\hat{\mu}_J(0)\omega_{\mu\mu.J}^{-\frac{1}{2}} \\
\underset{d}{\rightarrow}\; & \underbrace{\left[\bar{\mu}(\theta^*)'\bar{\mu}(\theta^*) - \bar{D}(\theta^*)'\bar{D}(\theta^*)\right]}_{\text{misspecification - identification}} \tfrac{\omega_{\mu\mu.J}^{\frac{1}{2}}\Omega_{JJ}^{\frac{1}{2}}}{\omega_{\mu\mu}(\theta^*)}(\theta^* + \Omega_{JJ}^{-1}\omega_{J\mu}) + \\
& \tfrac{1}{\sqrt{\omega_{\mu\mu.J}}}\left[\bar{D}(\theta^*)\sqrt{\omega_{\mu\mu.J}} + \bar{\mu}(\theta^*)(\theta^* + \Omega_{JJ}^{-1}\omega_{J\mu})\Omega_{JJ}^{\frac{1}{2}}\right]'\psi_\mu^*(\theta^*) + \\
& \tfrac{\sqrt{\omega_{\mu\mu.J}}}{\omega_{\mu\mu}(\theta^*)}\left[\bar{\mu}(\theta^*)\sqrt{\omega_{\mu\mu.J}} - \bar{D}(\theta^*)(\theta^* + \Omega_{JJ}^{-1}\omega_{J\mu})\Omega_{JJ}^{\frac{1}{2}}\right]'\psi_\theta^*(\theta^*) + \psi_\theta^*(\theta^*)'\psi_\mu^*(\theta^*)
\end{aligned}
\tag{32}
$$

where $\psi_\mu^*(\theta^*)$ and $\psi_\theta^*(\theta^*)$ are independent k_f-dimensional standard normal random vectors, $\Omega_{JJ.\mu}(\theta) = \Omega_{JJ} - \omega_{\mu J}(\theta)'\omega_{\mu\mu}(\theta)^{-1}\omega_{\mu J}(\theta)$, $\bar{D}(\theta) = \sqrt{N}\Sigma^{-\frac{1}{2}}D(\theta)\Omega_{JJ.\mu}(\theta)^{-\frac{1}{2}}$, $\bar{\mu}(\theta) = \sqrt{N}\Sigma^{-\frac{1}{2}}\mu_f(\theta)\omega_{\mu\mu}(\theta)^{-\frac{1}{2}}$.

Proof. See the Appendix. ■

The expansion of $\hat{J}(0)'\Sigma^{-1}\hat{\mu}_J(0)$ in Theorem 2 results from the decomposition of $\hat{\mu}_f(0)$ and $\hat{J}(0)$ into the independent building blocks of the weak instrument robust statistics, *i.e.*, the sample analogs of $\mu_f(\theta)$ and $D(\theta)$; see also Kleibergen and Zhan (2025). We do so because this independence allows for unambiguous normalized definitions for the strength of identification, $\bar{D}(\theta^*)'\bar{D}(\theta^*)$, and misspecification, $\bar{\mu}(\theta^*)'\bar{\mu}(\theta^*)$. It also leads to the decomposition of $\hat{J}(0)'\Sigma^{-1}\hat{\mu}_J(0)$ into deterministic and independent random components.

The first component of the expansion of $\hat{J}(0)'\Sigma^{-1}\hat{\mu}_J(0)$ in Theorem 2 is the only deterministic one. It equals zero when $\theta^* = -\Omega_{JJ}^{-1}\omega_{J\mu}$ and when the normalized misspecification and identification strengths at the pseudo-true value of the CUE are identical. Tests based on $\hat{J}(0)'\Sigma^{-1}\hat{\mu}_J(0)$, like, for example, the two-stage t-statistic,[5] can therefore not have power for discriminating values of θ^* from $-\Omega_{JJ}^{-1}\omega_{J\mu}$ if $\bar{\mu}(\theta^*)'\bar{\mu}(\theta^*)$ and $\bar{D}(\theta^*)'\bar{D}(\theta^*)$ are identical. The value of θ^* is thus not distinguishable from $-\Omega_{JJ}^{-1}\omega_{J\mu}$ when $\bar{\mu}(\theta^*)'\bar{\mu}(\theta^*)$ equals $\bar{D}(\theta^*)'\bar{D}(\theta^*)$, so it is not identified.

The expansion in Theorem 2 implies for the two-stage estimator that distinct CUE pseudo-true values θ^* for which the difference between the misspecification and identifications strengths is the same, lead to two-stage estimators that only differ by a random amount. This all results from the property that for different values of a

[5] The large sample distribution of the t-statistic is also non-standard for this setting.

and C with $C'a = 0$ and $a'a \geq C'C$, the pseudo-true value of the two-stage estimator is identical as shown by (30). The identification condition for the pseudo-true value of the two-stage estimator is thus the same as that for the CUE, although this is not obvious at face value.

When the model is correctly specified, $\bar{\mu}(\theta^*) = 0$, so the first component of the expansion in Theorem 2 is only equal to zero when either $\theta^* = -\Omega_{JJ}^{-1}\omega_{J\mu}$ or $\bar{D}(\theta^*) = 0$. Since $D(\theta^*) = J(0)$ under correct specification, the identification condition is just the traditional one that the Jacobian, $J(0)$, is non-zero. Thus, there is an additional component in the deterministic part of the expansion of $\hat{J}(0)'\Sigma^{-1}\hat{\mu}_J(0)$ in case of misspecification, because of which the identification condition of the pseudo-true value changes for both the CUE and the two-stage estimator when moving from correct specification to misspecification. Theorem 2 therefore shows how the identification condition changes when misspecification is present compared to the correctly specified setting. Existing misspecification-robust test procedures, like, for example, Hansen and Lee (2021), Lee (2018), and Kan, Robotti, and Shanken (2013), all require that the identification strength as reflected by $\bar{D}(\theta^*)'\bar{D}(\theta^*)$ is larger than the misspecification reflected by $\bar{\mu}(\theta^*)'\bar{\mu}(\theta^*)$.

The difference between IS and MISS is thus as indicative for interpreting the pseudo-true values of the CUE as for the two-stage estimator. We therefore next provide a generalization of the traditional test for identification in correctly specified GMM, *i.e.*, the sample analog of IS, $\widehat{\text{IS}}$, towards a test for identification in possibly misspecified GMM based on $\widehat{\text{IS}} - \widehat{\text{MISS}}$.

4.3 LR no-identification test for $m = 1$

For ease of exposition, we first assume that the KPS covariance matrix from Assumption 2 is known and there is a single structural parameter, so $m = 1$. We generalize to a to-be-estimated covariance matrix and more structural parameters thereafter in the next subsections. The sample analog of the CUE population objective function (21) then equals (twice) the concentrated log-likelihood for θ that results under normal errors in linear models. Because IS (6) and MISS (9) equal the CUE objective function at set or minimized values respectively, $\widehat{\text{IS}} - \widehat{\text{MISS}}$ equals the (quasi-) likelihood ratio (LR) statistic that tests $\text{H}_0 : \alpha = 0$ in (14):

$$\text{LR}(\alpha = 0) = \widehat{\text{IS}} - \widehat{\text{MISS}} = N \times \hat{Q}_{CUE}(\alpha = 0) - N \times \min_{\alpha \in \mathbb{R}} \hat{Q}_{CUE}(\alpha), \qquad (33)$$

with $\widehat{\text{IS}}$, $\widehat{\text{MISS}}$ and $\hat{Q}_{CUE}(\alpha)$ the sample analogs of IS, MISS and $Q_{CUE}(\alpha)$.[6]

When $m = 1$, the characteristic polynomial in (25) is quadratic so its smallest root, which equals $\widehat{\text{MISS}}$ in sample, has a closed-form expression, and we have an analytical expression for the LR statistic (see Moreira (2003)):

$$\text{LR}(\alpha = 0) = \tfrac{1}{2}\left[\hat{C}'\hat{C} - \hat{a}'\hat{a} + \sqrt{\left(\hat{C}'\hat{C} - \hat{a}'\hat{a}\right)^2 + 4(\hat{C}'\hat{a})^2}\right], \qquad (34)$$

for $\hat{a} = \sqrt{N}\Sigma^{-\frac{1}{2}}\hat{\mu}_J(0)\omega_{\mu\mu.J}^{-\frac{1}{2}}$, $\hat{C} = \sqrt{N}\Sigma^{-\frac{1}{2}}\hat{J}(0)\Omega_{JJ}^{-\frac{1}{2}}$ and $\hat{\mu}_J(0) = \hat{\mu}_f(0) - \hat{J}(0)\Omega_{JJ}^{-1}\omega_{J\mu}$. While the expression of the LR statistic (34) is for testing $\text{H}_0 : \alpha = 0$, our no-identification hypothesis of interest concerns a specific setting of (a, C) that implies $\alpha = 0$ as provided in Corollary 4:

$$\text{H}_{\text{non-ind}} : \text{MISS} = \text{IS} \Longleftrightarrow C'C - a'a \leq 0 \text{ and } C'a = 0. \qquad (35)$$

The expression of the LR statistic (34) contains the sample analogs of all elements of $\text{H}_{\text{non-ind}}$.

To obtain a conditional critical value function for the LR test that controls the size over all parameter configurations of $\text{H}_{\text{non-ind}}$ (35), we use the boundary non-identified setting:

$$\text{H}^*_{\text{non-ind}} : C'C - a'a = 0,\ C'a = 0. \qquad (36)$$

A conditional critical value function that controls the size of the LR test for $\text{H}^*_{\text{non-ind}}$ also does so for $\text{H}_{\text{non-ind}}$, because the derivative of $\text{LR}(\alpha = 0)$ with respect to $(\hat{C}'\hat{C} - \hat{a}'\hat{a})$ is strictly positive at $\hat{C}'\hat{C} - \hat{a}'\hat{a} = 0$. Negative values of $\hat{C}'\hat{C} - \hat{a}'\hat{a}$ for the same value of $\hat{C}'\hat{a}$ and conditioning statistic thus only decrease the value of $\text{LR}(\alpha = 0)$ and will not lead to size distortion.

[6] Based on (14), $Q_{CUE}(\alpha) = \eta_f(\alpha)'\left[\binom{(\alpha\otimes I_{k_f})}{I_{k_f}}' V \binom{(\alpha\otimes I_{k_f})}{I_{k_f}}\right]^{-1}\eta_f(\alpha)$. However, the expressions of IS, MISS remain unaffected.

To construct a conditional critical value function of the LR statistic (34), we provide the limit behavior for $\hat{a}$ and $\hat{C}$.

Corollary 5: *When Assumptions 1 and 2, $m = 1$, and the drifting sequence in (24) apply, the limit behavior of $\hat{a} = \sqrt{N}\Sigma^{-\frac{1}{2}}\hat{\mu}_J(0)\omega_{\mu\mu.J}^{-\frac{1}{2}}$, $\hat{C} = \sqrt{N}\Sigma^{-\frac{1}{2}}\hat{J}(0)\Omega_{JJ}^{-\frac{1}{2}}$ with $\hat{\mu}_J(0) = \hat{\mu}_f(0) - \hat{J}(0)\Omega_{JJ}^{-1}\omega_{J\mu}$, is:*

$$\begin{pmatrix} \hat{a} \\ \hat{C} \end{pmatrix} \underset{d}{\rightarrow} \begin{pmatrix} a \\ C \end{pmatrix} + \begin{pmatrix} \psi_{\mu.J}^{*} \\ \psi_{J}^{*} \end{pmatrix}, \tag{37}$$

where $\psi_{\mu.J}^{*}$ and ψ_{J}^{*} are independent k_f-dimensional standard normal random vectors.

The LR statistic (34) is an invariant statistic and therefore, as all invariant statistics are, a function of the maximal invariant which equals the normalized quadratic form of $(\hat{\mu}_f(0) \vdots \hat{J}(0))$; see, *e.g.*, Andrews, Moreira and Stock (2006):

$$\begin{aligned} \text{MAXINV} = & \; N \times \Omega^{-\frac{1}{2}\prime}(\hat{\mu}_f(0) \vdots \hat{J}(0))'\Sigma^{-1}(\hat{\mu}_f(0) \vdots \hat{J}(0))\Omega^{-\frac{1}{2}} \\ = & \; (\hat{a} \vdots \hat{C})'(\hat{a} \vdots \hat{C}) \\ = & \; \begin{pmatrix} \hat{a}'\hat{a} & \vdots & \hat{a}'\hat{C} \\ \hat{C}'\hat{a} & \vdots & \hat{C}'\hat{C} \end{pmatrix}. \end{aligned} \tag{38}$$

The limiting distribution of the maximal invariant therefore only depends on three population parameters: $a'a$, $C'a$, and $C'C$.

Any invertible function of the maximal invariant is also a maximal invariant. To obtain a convenient conditioning statistic for the LR test of no-identification, we transform the components of the maximal invariant:

$$\text{MAXINV} = (\hat{C}'\hat{C} - \hat{a}'\hat{a},\; \hat{C}'\hat{a},\; \hat{C}'\hat{C} + \hat{a}'\hat{a}), \tag{39}$$

whose limiting distribution only depends on the population parameters $C'C - a'a$, $C'a$, and $C'C + a'a$. The transformation is motivated by:

1. The LR statistic (34) is only a function of the first two elements of the maximal invariant (39): $\hat{C}'\hat{C} - \hat{a}'\hat{a}$, $\hat{C}'\hat{a}$, and not of the third element, $\hat{C}'\hat{C} + \hat{a}'\hat{a}$.

2. Under $\text{H}^*_{\text{non-ind}}$, the third element of (39) is approximately independent of the first two, as shown by Theorem 3 below.

Theorem 3: *Under the conditions of Corollary 5 and* $\text{H}^*_{\text{non-ind}}$*, the limit behavior of* $\hat{C}'\hat{C} - \hat{a}'\hat{a}$, $\hat{C}'\hat{a}$, $\hat{C}'\hat{C} + \hat{a}'\hat{a}$ *can be written as:*

$$\begin{aligned} \hat{C}'\hat{C} - \hat{a}'\hat{a} &\underset{d}{\to} 2\binom{C}{-a}'\binom{\psi_J^*}{\psi_{\mu.J}^*} + \binom{\psi_J^*}{-\psi_{\mu.J}^*}'\binom{\psi_J^*}{\psi_{\mu.J}^*} \\ \hat{C}'\hat{a} &\underset{d}{\to} \binom{a}{C}'\binom{\psi_J^*}{\psi_{\mu.J}^*} + \psi_J^{*\prime}\psi_{\mu.J}^* \\ \hat{C}'\hat{C} + \hat{a}'\hat{a} &\underset{d}{\to} C'C + a'a + 2\binom{C}{a}'\binom{\psi_J^*}{\psi_{\mu.J}^*} + \binom{\psi_J^*}{\psi_{\mu.J}^*}'\binom{\psi_J^*}{\psi_{\mu.J}^*}. \end{aligned} \tag{40}$$

Under $\text{H}^*_{\text{non-ind}}$*, the dependence between the three components in (40) only results from the last elements in their expressions which are uncorrelated.*

Proof. Results from Corollary 5 and applying $\text{H}^*_{\text{non-ind}}$ (36), so $\binom{C}{-a}'\binom{a}{C} = 0$, $\binom{C}{-a}'\binom{C}{a} = 0$, and $\binom{a}{C}'\binom{C}{a} = 0$, because of which only the last elements are dependent. ■

Because the LR statistic (34) is a function of the maximal invariant (39), it depends on the same population parameters. Under Assumptions 1, 2, the drifting sequence in (24) and $\text{H}^*_{\text{non-ind}}$, the only parameter where the maximal invariant (39) depends on is $C'C+a'a$ because all the other parameters which it depends on, $C'C-a'a$ and $C'a$, are equal to zero. We therefore use:

$$\text{rk} = \hat{C}'\hat{C} + \hat{a}'\hat{a}, \tag{41}$$

as a conditioning statistic for the only parameter in the distribution of $\text{LR}(\alpha = 0)$ (34). The decomposition in (40) shows that this conditioning statistic is independently distributed of $\text{LR}(\alpha = 0)$ for larger values of $C'C + a'a$, because the multipliers of the random components $\binom{\psi_J^*}{\psi_{\mu.J}^*}$ are all orthogonal to each other under $\text{H}^*_{\text{non-ind}}$. It is approximately independently distributed for smaller values of $C'C + a'a$.

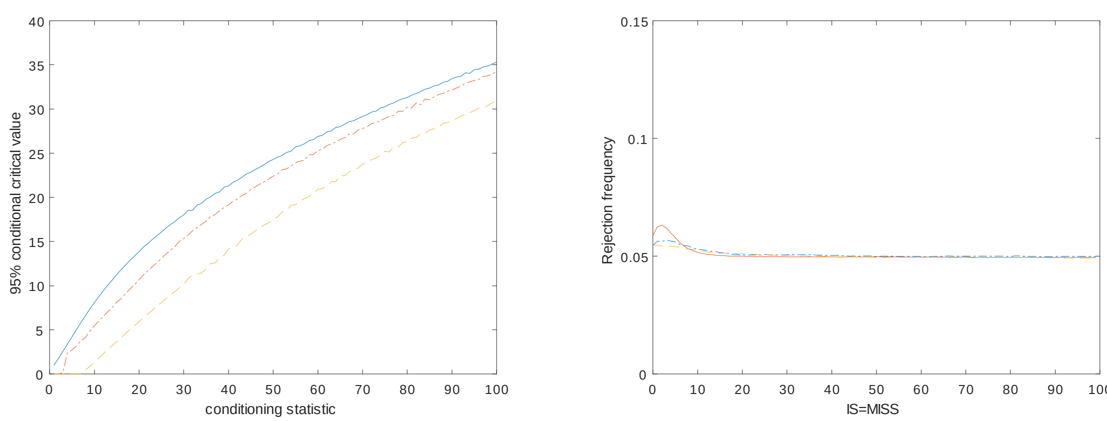
The conditioning statistic (41) differs from the conditioning statistic of the conditional LR test of Moreira (2003) for the linear IV regression model with one included endogenous variable. For our specification, that conditioning statistic would correspond to:

$$\text{rk}_{\text{Moreira (2003)}} = \hat{a}'\hat{a}. \tag{42}$$

Under the hypothesis of interest of Moreira (2003), (42) provides a sufficient statistic for the only parameter which the LR statistic then depends on and is also independently distributed of. Our conditioning statistic (41) provides an estimator for the only parameter the limiting distribution of the LR statistic (34) depends on under $H^*_{\text{non-ind}}$, *i.e.*, $C'C + a'a$, but it is not a sufficient statistic. It is also approximately independent of the LR statistic (34) because of which we use a different (simulation) algorithm to obtain conditional critical values than Moreira (2003).[7]

The algorithm for computing the conditional critical value function is stated in the Appendix. By sampling $\hat{a}$ and $\hat{C}$ using $\psi^*_{\mu.J}$ and ψ^*_J for a large range of values of a and C that satisfy $H^*_{\text{non-ind}}$, it computes the conditional distribution of LR$(\alpha = 0)$ given rk$= \hat{C}'\hat{C} + \hat{a}'\hat{a}$. The conditional critical value function for conducting a 5% significance test of $H^*_{\text{non-ind}}$ using LR$(\alpha = 0)$ then corresponds with the 95% percentiles of the computed conditional distribution of LR$(\alpha = 0)$ given rk$= \hat{C}'\hat{C} + \hat{a}'\hat{a}$.

Figure 4: Conditional critical value function for testing $H^*_{\text{non-ind}}$ at the 5% significance level using LR$(\alpha = 0)$ as a function of rk (41) and the resulting rejection frequencies for $k_f = 3$ (solid), 10 (dash-dotted) and 25 (dashed).



Panel 4.1. Conditional critical values

Panel 4.2. Rejection frequencies

[7] Hillier (2009) provides a recurrence algorithm to provide critical values for the Moreira (2003) conditional LR test that does not involve simulation.

4.3.1 Size and power of the LR test for $m = 1$

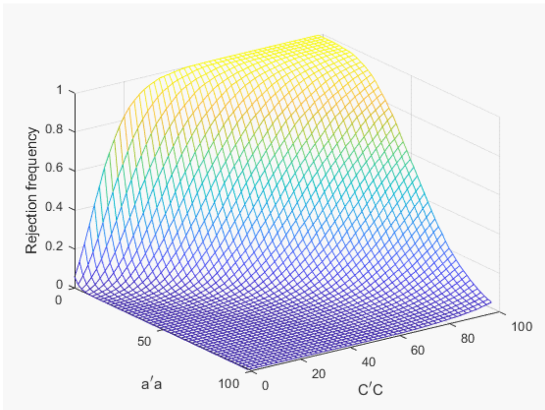
Figure 4 shows the 95% conditional critical value functions for different numbers of moment equations in Panel 4.1, and the resulting rejection frequencies when we conduct a 5% significance LR test of $\mathrm{H}^*_{\text{non-ind}}$ in Panel 4.2. It is striking how close the rejection frequencies in Panel 4.2 are to the nominal 5%. There is only some minor overrejection for small values of IS = MISS, which could be removed by further calibrating the conditional critical value function. Because of the computational ease of the algorithm and the just very small size distortions it leads to, we, for now, refrain from doing so. It all shows that the approximate independence of the conditioning statistic (41) allows us to compute a conditional critical value function for the LR test of no-identification with excellent size properties.

Figure 5 shows the power surfaces of the LR test of no-identification. The involved conditional critical value function is calibrated to the boundary setting of no-identification $\mathrm{H}^*_{\text{non-ind}}$ (36), but our main hypothesis of interest is $\mathrm{H}_{\text{non-ind}}$ (35). The power surfaces in Figure 5 show the rejection frequencies with respect to potential violation of one of the two components in $\mathrm{H}_{\text{non-ind}}$ (35) while the other one is kept at the hypothesized value. Panel 5.1, and its contour lines in Panel 5.3, therefore show the power surface of the LR test for violations of $C'C \leq a'a$ while $C'a = 0$, and Panel 5.2 shows it for violations of $C'a = 0$ while $C'C = a'a$.

Panels 5.1 and 5.3 show that the conditional LR test controls the size of testing $\mathrm{H}_{\text{non-ind}}$ (35) because, while the conditional critical values are calibrated to $\mathrm{H}^*_{\text{non-ind}}$ (36), the rejection frequency is at most 5% when $a'a \geq C'C$ as indicated by the contour lines in Panel 5.3. When $\mathrm{H}_{\text{non-ind}}$ does not hold, so $C'C$ exceeds $a'a$, Panels 5.1 and 5.3 show that the LR test has discriminatory power for rejecting $\mathrm{H}_{\text{non-ind}}$. Interestingly, the contour line at 5% in Panel 5.3 is on the 45^0 degree line. On lines orthogonal to the 45^0 degree, $C'C + a'a$ is constant and so is then, approximately, the conditioning statistic. The contour lines therefore show that for constant values of $C'C + a'a$, the LR test clearly discriminates between settings with $C'C > a'a$, for which it mostly rejects, or $C'C < a'a$, for which it does not reject.

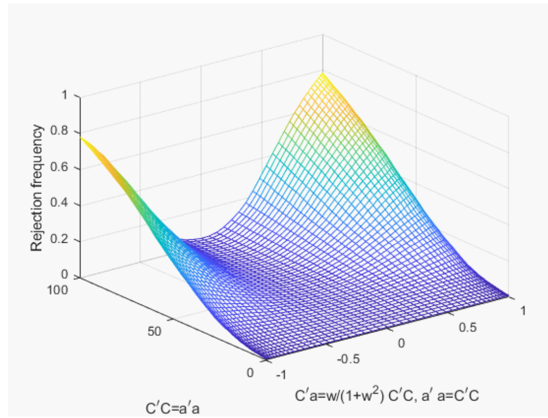
Panel 5.2 similarly shows that the LR test has discriminatory power for detecting non-zero values of $C'a$ when $C'C = a'a$.

Figure 5: Power of 5% significance conditional LR no-identification test, $k_f = 3$.



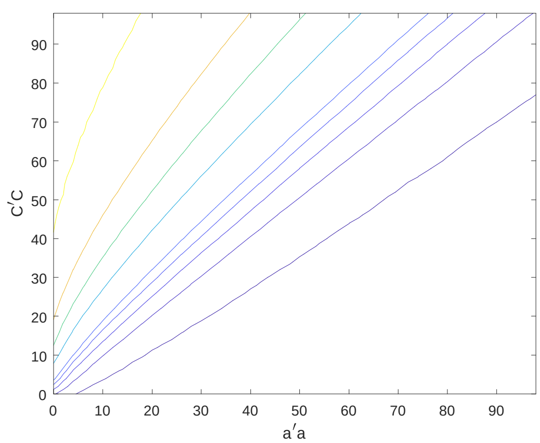
Panel 5.1

$C'a = 0$



Panel 5.2

$C'C = a'a,\ C'a = \frac{w}{1+w^2} C'C$



Panel 5.3. Contour lines of Panel 5.1 at 1, 5, 10, 15, 20, 40, 60, 80 and 99%.

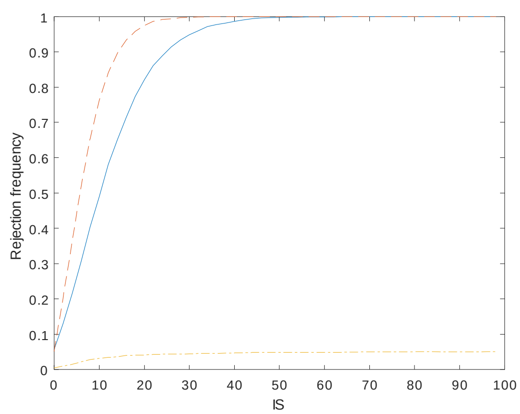
4.3.2 Comparison of LR with IS, MISS, and DRLM tests

It is interesting to compare the power of the LR no-identification test with other tests that test part of the composite null hypothesis of no-identification. Figure 6 therefore shows the power of: the LR test of no-identification, the IS test whose test statistic equals the F-statistic for testing $J(0) = 0$, and the MISS test whose test statistic equals the J-statistic. The level of significance is 5%.

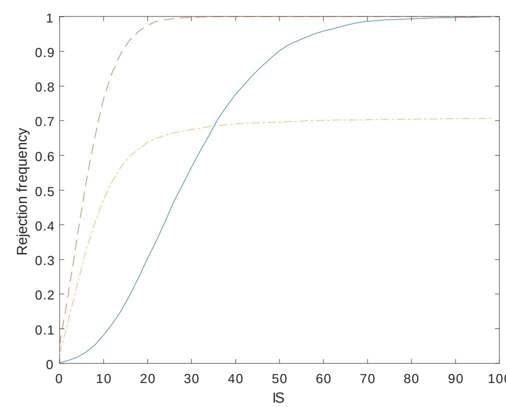
Panel 6.1 shows the power of the three tests when there is no misspecification.

The IS test is then more powerful than the LR test, while the MISS test rejects at most 5% because there is no misspecification. To facilitate comparison, Panel 6.2 has an increased level of misspecification of 10, for which the power curve of the 5% IS test has not changed compared to Panel 6.1. Up to IS equal to 10 there is, however, no identification. The rejection frequency of the MISS test at IS = 10, is around 45%. Panel 6.2 is interesting because the MISS test still mostly does not reject up to IS = 10, at which the IS test rejects around 80%. The combination of these two tests therefore overstates the identification strength and understates the misspecification, which shows the importance of a proper test for no-identification that allows for misspecification like the LR no-identification test. Note that Figure 6 also shows that the LR test is size-correct, *i.e.*, its rejection frequencies are near the nominal 5% at IS = MISS = 0 in Panel 6.1, and at IS = MISS = 10 in Panel 6.2, respectively.

Figure 6: Power of 5% significance conditional LR no-identification test (solid), IS (dashed) and MISS (dash-dotted) tests, $k_f = 3$.



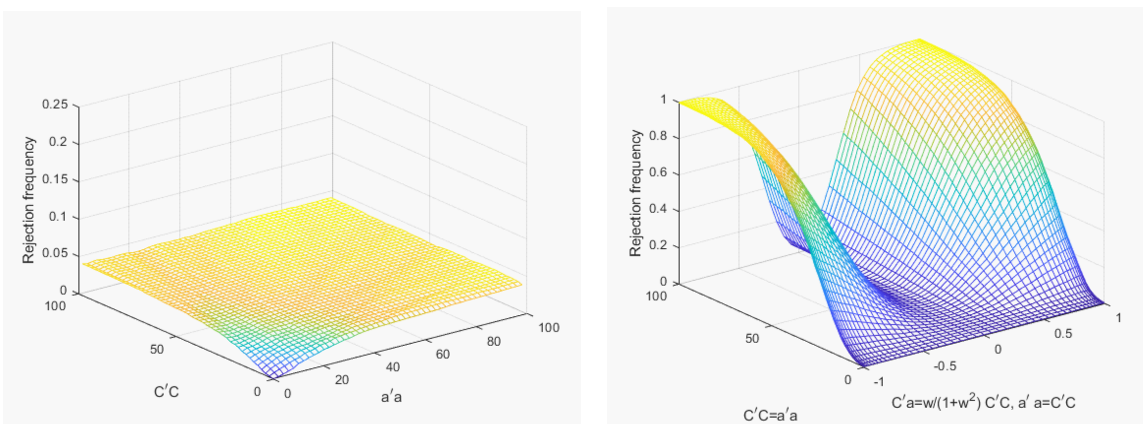
Panel 6.1. MISS = 0, $C'a = 0$



Panel 6.2. MISS = 10, $C'a = 0$

Figure 7 shows the power surface of the double robust Lagrange multiplier (DRLM) test proposed in Kleibergen and Zhan (2025), while the counterpart of the LR test is presented in Figure 5. The DRLM test is a size-correct test of $H_0 : \alpha = 0$ when using $\chi^2(m)$ critical values. It is based on the score of the population objective function, which equals zero when $C'a = 0$. Therefore, the power and size of the DRLM test coincide in Panel 7.1, for which $C'a = 0$ is imposed. In contrast, Panel 7.2 shows that the DRLM test has good power, exceeding that of the LR test in Panel 5.2. Similar to the IS test in Panel 6.1, Figure 7 shows that for specific settings, the power of tests of just one component of the composite hypothesis $\mathrm{H}_{\text{non-ind}}$ (35) can exceed that of the LR test, but these tests have misleading power or no power when the other component of the composite hypothesis $\mathrm{H}_{\text{non-ind}}$ (35) gets violated.

Figure 7: Power of 5% significance DRLM test of $\mathrm{H}_0 : \alpha = 0$, $k_f = 3$.



Panel 7.1

$C'a = 0$

Panel 7.2

$C'C = a'a, \ C'a = \frac{w}{1+w^2} C'C$

4.4 Homoskedasticity, unknown covariance, $m = 1$

The covariance matrices $\Omega = \left(\begin{smallmatrix} \omega_{\mu\mu} & \vdots & \omega_{\mu J} \\ \omega_{J\mu} & \vdots & \Omega_{JJ} \end{smallmatrix}\right)$ and Σ in Assumption 2 are considered as given for ease of exposition in the previous subsection. In empirical settings, these covariance matrices are typically unknown, so we use consistent estimators $\hat{\Omega}$ and $\hat{\Sigma}$ instead, which can be written as:

$$\begin{aligned} \hat{\Omega} = & \ \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \binom{u_i}{V_i} \binom{u_i}{V_i}' = \begin{pmatrix} \hat{\omega}_{\mu\mu} & \vdots & \hat{\omega}_{\mu J} \\ \hat{\omega}_{J\mu} & \vdots & \hat{\Omega}_{JJ} \end{pmatrix} \\ \hat{\Sigma} = & \ \tfrac{1}{N} \textstyle\sum_{i=1}^{N} Z_i Z_i', \end{aligned} \tag{43}$$

with $\binom{u_i}{V_i}$ and Z_i realizations of $m+1$ and k_f dimensional random vectors that lead to consistent esimation of Ω and Σ, respectively. We use the covariance matrix estimators (43) and adapt $\hat{a}$ and $\hat{C}$ defined below (34) accordingly. Therefore, in this subsection, we consider $\hat{a} = \sqrt{N}\hat{\Sigma}^{-\frac{1}{2}}\hat{\mu}_J(0)\hat{\omega}_{\mu\mu.J}^{-\frac{1}{2}}$, $\hat{C} = \sqrt{N}\hat{\Sigma}^{-\frac{1}{2}}\hat{J}(0)\hat{\Omega}_{JJ}^{-\frac{1}{2}}$ with $\hat{\mu}_J(0) = \hat{\mu}_f(0) - \hat{J}(0)\hat{\Omega}_{JJ}^{-1}\hat{\omega}_{J\mu}$, and $\hat{\omega}_{\mu\mu.J} = \hat{\omega}_{\mu\mu} - \hat{\omega}_{\mu J}\hat{\Omega}_{JJ}^{-1}\hat{\omega}_{J\mu}$. The convergence of the covariance estimators alters the behavior of the components of the LR statistic (34) stated in Theorem 3, so Theorem 4 states their higher order expansion.

Theorem 4: *When Assumptions 1 and 2, $m = 1$, the drifting sequence in (24) and* $\mathrm{H}^*_{\text{non-ind}} : C'C = a'a,\ C'a = 0$, *apply, the higher order expressions for the components of the LR statistic (34) are*:

$$
\begin{aligned}
\hat{C}'\hat{C} - \hat{a}'\hat{a} = \;& 2\binom{C}{-a}'\binom{\psi_J^*}{\psi_{\mu.J}^*} + \binom{\psi_J^*}{-\psi_{\mu.J}^*}'\binom{\psi_J^*}{\psi_{\mu.J}^*} + \tfrac{1}{\sqrt{N}}\binom{-C'C}{a'a}'\binom{\psi_{JJ}}{\psi_{\mu\mu}} + \\
& \tfrac{1}{\sqrt{N}}\left(2\binom{-\psi_{JJ}C}{\psi_{\mu\mu}a}'\binom{\psi_J^*}{\psi_{\mu.J}^*} + \binom{-\psi_{JJ}\psi_J^*}{\psi_{\mu\mu}\psi_{\mu.J}^*}'\binom{\psi_J^*}{\psi_{\mu.J}^*}\right) + \\
& \tfrac{1}{\sqrt{N}}\left(a \otimes a - C \otimes C + 2\left(\psi_{\mu.J}^* \otimes a - \psi_J^* \otimes C\right)\right)'\mathrm{vec}(\Psi_\Sigma) + \\
& \tfrac{1}{\sqrt{N}}\left(\psi_{\mu.J}^* \otimes \psi_{\mu.J}^* - \psi_J^* \otimes \psi_J^*\right)'\mathrm{vec}(\Psi_\Sigma) + \\
& \tfrac{2}{\sqrt{N}}\left(\binom{a}{C}'\binom{\psi_J^*}{\psi_{\mu.J}^*} + \psi_J^{*\prime}\psi_{\mu.J}^*\right)'\psi_{J\mu} + O_p(N^{-1}) \\
\hat{C}'\hat{a} = \;& \binom{a}{C}'\binom{\psi_J^*}{\psi_{\mu.J}^*} + \psi_J^{*\prime}\psi_{\mu.J}^* - \tfrac{1}{\sqrt{N}}(C + \psi_J^*)'(C + \psi_J^*)\psi_{J\mu} - \\
& \tfrac{1}{\sqrt{N}}\left(C \otimes a + \psi_J^* \otimes a + C \otimes \psi_{\mu.J}^* + \psi_J^* \otimes \psi_{\mu.J}^*\right)'\mathrm{vec}(\Psi_\Sigma) - \\
& \tfrac{1}{2\sqrt{N}}\left(\psi_{\mu\mu} + \psi_{JJ}\right)\left(\binom{a}{C}'\binom{\psi_J^*}{\psi_{\mu.J}^*} + \psi_J^{*\prime}\psi_{\mu.J}^*\right) + O_p(N^{-1}) \\
\hat{C}'\hat{C} + \hat{a}'\hat{a} = \;& C'C + a'a + 2\binom{C}{a}'\binom{\psi_J^*}{\psi_{\mu.J}^*} + \binom{\psi_J^*}{\psi_{\mu.J}^*}'\binom{\psi_J^*}{\psi_{\mu.J}^*} - \tfrac{1}{\sqrt{N}}\binom{C'C}{a'a}'\binom{\psi_{JJ}}{\psi_{\mu\mu}} - \\
& \tfrac{1}{\sqrt{N}}\left(2\binom{\psi_{JJ}C}{\psi_{\mu\mu}a}'\binom{\psi_J^*}{\psi_{\mu.J}^*} + \binom{\psi_{JJ}\psi_J^*}{\psi_{\mu\mu}\psi_{\mu.J}^*}'\binom{\psi_J^*}{\psi_{\mu.J}^*}\right) - \\
& \tfrac{1}{\sqrt{N}}\left(a \otimes a + C \otimes C + 2\left(\psi_{\mu.J}^* \otimes a + \psi_J^* \otimes C\right)\right)'\mathrm{vec}(\Psi_\Sigma) - \\
& \tfrac{1}{\sqrt{N}}\left(\psi_{\mu.J}^* \otimes \psi_{\mu.J}^* + \psi_J^* \otimes \psi_J^*\right)'\mathrm{vec}(\Psi_\Sigma) - \\
& \tfrac{2}{\sqrt{N}}\left(\binom{a}{C}'\binom{\psi_J^*}{\psi_{\mu.J}^*} + \psi_J^{*\prime}\psi_{\mu.J}^*\right)'\psi_{J\mu} + O_p(N^{-1}),
\end{aligned}
\tag{44}
$$

where $\psi_{\mu\mu}$, $\psi_{J\mu}$, ψ_{JJ} and ψ_Σ are 1, m, $\frac{1}{2}m(m+1)$ and $\frac{1}{2}k_f(k_f+1)$ dimensional mean zero standardized normal distributed random vectors further defined in the proof in the Appendix, and Ψ_Σ is a symmetric $k_f \times k_f$ dimensional matrix which is such that $\psi_\Sigma = vech(\Psi_\Sigma)$.

Proof. See the Appendix. ■

Compared to the known covariance matrix setting in Theorem 3, Theorem 4 shows two main consequences of the covariance matrix estimation on the components of the LR statistic, which are stated in Corollaries 6 and 7.

Corollary 6. *When $C'C$ ($= a'a$ under $\mathrm{H}^*_{\text{non-ind}}$) is no longer fixed but increases with the sample size at a rate of at least $\sqrt{N}$, the higher order elements affect the large sample behavior of each of the components in (44), which has to be accounted for in the conditional critical value function of the LR test.*

Corollary 7. *The approximate independence of $(\hat{C}'\hat{C} - \hat{a}'\hat{a},\, \hat{C}'\hat{a})$ from $\hat{C}'\hat{C} + \hat{a}'\hat{a}$ from Theorem 3 also applies to the higher order elements in (44). Firstly, it results from the orthogonality of the components by which the standardized random variables get multiplied.*[8] *Secondly, it results because dependent components are of a lower order in the sample size, which are dominated by the zero-th order components.*

Based on Corollary 7, we similarly use $\hat{C}'\hat{C} + \hat{a}'\hat{a}$ as a conditioning statistic for computing the conditional critical value function for $\mathrm{LR}(\alpha = 0)$ under $\mathrm{H}^*_{\text{non-ind}}$ when using the covariance matrix estimators (43).

Under $\mathrm{H}^*_{\text{non-ind}}$, Theorem 4 shows that the behavior of $\hat{C}'\hat{C} - \hat{a}'\hat{a}$, $\hat{C}'\hat{a}$ and $\hat{C}'\hat{C} + \hat{a}'\hat{a}$ only depends on: the standardized random variables $\psi_{\mu\mu}$, $\psi_{J\mu}$, ψ_{JJ}, ψ_{Σ}, $\psi^*_{\mu.J}$, ψ^*_J, the length of C (and a) and the sample size N. We therefore compute the conditional critical value function for $\mathrm{LR}(\alpha = 0)$ under $\mathrm{H}^*_{\text{non-ind}}$ by using the same conditioning statistic as for the known homoskedastic covariance setting:

$$\mathrm{rk} = \hat{C}'\hat{C} + \hat{a}'\hat{a}. \tag{45}$$

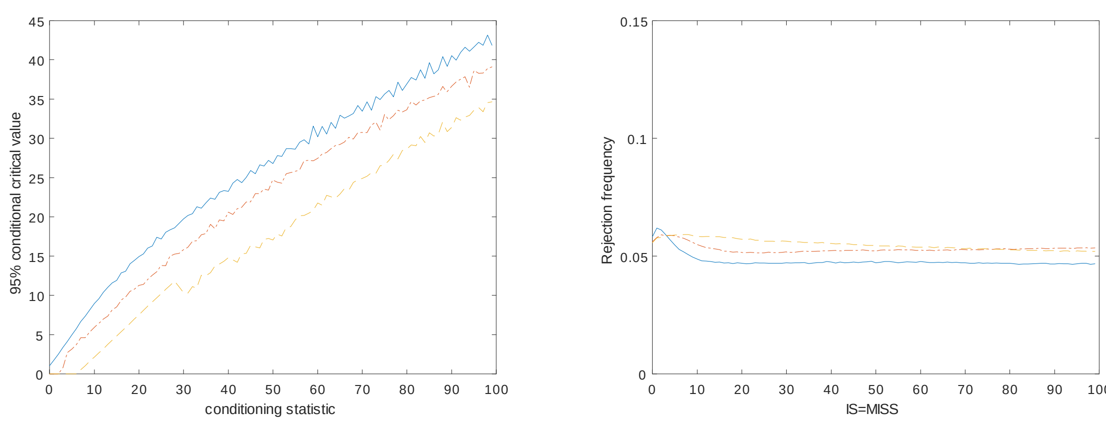
The conditional critical value function is computed specifically for the sample size under consideration, so all elements that contribute to $\mathrm{LR}(\alpha = 0)$ remain relevant. We calibrate a conditional critical value function for $\mathrm{LR}(\alpha = 0)$ by simulating $\hat{a}$ and $\hat{C}$ for a range of values of the identical lengths of a and C, while they are also orthogonal, and the respective sample size N for the data under consideration. Using simulated standardized i.i.d. random variables, we compute $\hat{a}$ and $\hat{C}$, using their expressions stated in the proof of Theorem 4 in the Appendix, and construct the critical value function given rk (45) using the algorithm discussed previously and stated in the

[8] For example, $\binom{\psi_{JJ}}{\psi_{\mu\mu}}$ is pre-multiplied by $\binom{-C'C}{a'a}$, $\binom{a}{C}'\binom{\psi^*_J}{\psi^*_{\mu.J}} + \psi^{*\prime}_J\psi^*_{\mu.J}$, and $\binom{C'C}{a'a}$ respectively, all of which are orthogonal to each other or independent under $\mathrm{H}^*_{\text{non-ind}}$. This similarly applies to the elements by which $\mathrm{vec}(\Psi_\Sigma)$ is pre-multiplied in (44).

Appendix.[9]

For the estimated covariance matrix setting, Panels 8.1 and 8.2 in Figure 8 show respectively: the 95% conditional critical value function of the conditional LR test of $\text{H}^*_{\text{non-ind}}$ (36) given rk (45) for $k_f = 3, 10, 25$ and $N = 250$, and the rejection frequencies of the resulting 5% significance LR test of $\text{H}^*_{\text{non-ind}}$. Compared to Panel 4.1, Panel 8.1 shows that the 95% conditional critical values have increased for larger values of the conditioning statistic. For these values, the assumption that $C'C/\sqrt{N} = a'a/\sqrt{N} \to 0$, which validates the critical value function in Panel 4.1, is no longer appropriate. The rejection frequencies in Panel 8.2 are all close to the nominal 5%, which shows that the computed conditional critical value function controls the size of the LR test. Power surfaces for the estimated covariance matrix setting are comparable to those for the known covariance matrix setting shown in Figure 5, so they are omitted for brevity.

Figure 8: Conditional critical value function for testing $\text{H}^*_{\text{non-ind}}$ at the 5% significance level using LR($\alpha = 0$) as a function of rk (45) and the resulting rejection frequencies for $k_f = 3$ (solid), 10 (dash-dotted) and 25 (dashed), $N = 250$.



Panel 8.1. Conditional critical values

Panel 8.2. Rejection frequencies

[9] The conditional critical value function can similarly be computed by resampling the normalized data using the bootstrap. It is also possible to construct the conditional critical value function using the higher order approximation from Theorem 4, which similarly depends on the sample size N.

4.5 Homoskedasticity, unknown covariance, $m > 1$

For more than one structural parameter, $m > 1$, the difference between the sample analogs of IS and MISS equals the LR statistic that tests for a zero value of α in the population moment equation:[10]

$$\eta_f(\alpha, \varphi) = J(0)_1 + \mu_f(0)\alpha + J(0)_2\varphi, \tag{46}$$

with $J(0) = (J(0)_1 \vdots J(0)_2)$, $J(0)_1 : k_f \times 1$, $J(0)_2 : k_f \times (m-1)$, so

$$\begin{aligned} \text{LR}(\alpha = 0) &= \widehat{\text{IS}} - \widehat{\text{MISS}} \\ &= N \times \min_{\varphi \in \mathbb{R}^{m-1}} \hat{Q}_{CUE}(\alpha = 0, \varphi) - N \times \min_{\alpha \in \mathbb{R}, \varphi \in \mathbb{R}^{m-1}} \hat{Q}_{CUE}(\alpha, \varphi), \end{aligned} \tag{47}$$

for

$$\hat{Q}_{CUE}(\alpha, \varphi) = \hat{\eta}_f(\alpha, \varphi)' \left[\begin{pmatrix} (\alpha \otimes I_{k_f}) \\ I_{k_f} \\ (\varphi \otimes I_{k_f}) \end{pmatrix}' \hat{V} \begin{pmatrix} (\alpha \otimes I_{k_f}) \\ I_{k_f} \\ (\varphi \otimes I_{k_f}) \end{pmatrix} \right]^{-1} \hat{\eta}_f(\alpha, \varphi), \tag{48}$$

with $\hat{\eta}_f(\alpha, \varphi) = \hat{J}(0)_1 + \hat{\mu}_f(0)\alpha + \hat{J}(0)_2\varphi$.

While the expression of the LR statistic in (47) is generic with respect to the specification of the covariance matrix estimator $\hat{V}$, we next pin down the parameter setting for a DGP with the KPS covariance matrix from Assumption 2 that implies equality of IS and MISS.

Theorem 5: *Under Assumptions 1, 2 and the specification from (24), IS equals $\frac{1}{N}$ times the squared smallest singular value that results from a singular value decomposition (SVD) of C :*

$$C = U_C S_C V_C', \tag{49}$$

[10] See also (11). The moment equation in (46) is normalized using the first column of $J(0)$. The LR statistic is invariant with respect to this normalization, so an identical value results when normalized using any other column of $J(0)$.

with $U_C = (U_{C,1} \vdots U_{C,m} \vdots U_{C,2})$ an orthonormal $k_f \times k_f$ dimensional matrix, $U_{C,1} : k_f \times (m-1)$, $U_{C,m} : k_f \times 1$, $U_{C,2} : k_f \times (k_f - m)$ dimensional matrices, $V_C = (V_{C,1} \ldots V_{C,m})$ an $m \times m$ dimensional orthonormal matrix, and S_C a $k_f \times m$ dimensional matrix with the singular values, $s_{C,11}, \ldots, s_{C,mm}$, in decreasing order on the main diagonal. For large values of $s_{C,11} \ldots s_{C,(m-1)(m-1)}$, a necessary and sufficient condition for no-identification of the pseudo-true value of the CUE, or equality of IS and MISS, is:

$$\mathrm{H}_{\text{non-ind}} : U_{C,m}' a = 0 \text{ and } a' M_{U_{C,1}} a \geq s_{C,mm}^2 \Leftrightarrow \mathrm{H}_{\text{non-ind}} : \mathrm{IS} = \mathrm{MISS}. \tag{50}$$

Proof. See the Appendix. ■

Theorem 5 states the identification condition for large values of $s_{C,11} \ldots s_{C,(m-1)(m-1)}$. The identification condition for $m > 1$ is then basically identical to the one for $m = 1$ when we pre-multiply $(a \vdots C)$ by the eigenvectors of the smallest and null singular values that result from the SVD (49):

$$\begin{array}{c} (a \vdots C)'(U_{C,m} \vdots U_{C,2})(U_{C,m} \vdots U_{C,2})'(a \vdots C) \\ = \begin{pmatrix} 1 & 0 \\ 0 & V_{C,m} \end{pmatrix} \begin{pmatrix} a' M_{U_{C,1}} a & a' U_{C,m} s_{C,mm} \\ s_{C,mm} U_{C,m}' a & s_{C,mm}^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & V_{C,m}' \end{pmatrix}, \end{array} \tag{51}$$

where we used that $a'(U_{C,m} \vdots U_{C,2})(U_{C,m} \vdots U_{C,2})' a = a' M_{U_{C,1}} a$. Since $\begin{pmatrix} 1 \vdots & 0 \\ 0 \vdots & V_{C,m} \end{pmatrix}$ is orthonormal, the identification condition results from the matrix pre- and post-multiplied by it in (51) and can therefore be re-written identical to (27):

$$\mathrm{H}_0 : \mathrm{IS} = \mathrm{MISS} \quad \Leftrightarrow \quad \left\{ \begin{array}{c} s_{C,mm}^2 \leq a' M_{U_{C,1}} a \\ U_{C,m}' a = 0. \end{array} \right. \tag{52}$$

Alongside the above analogy with $m = 1$, the assumed large values of $s_{C,11}, \ldots, s_{C,(m-1)(m-1)}$ imply that we can consistently estimate $U_{C,1}$, so we can construct a conditional critical value function for the LR test of no-identification along the lines for the one for $m = 1$.

For smaller values of $s_{C,11}, \ldots, s_{C,(m-1)(m-1)}$, a more stringent identification condition than (52) results because the smallest characteristic root of (25) is a nondecreasing function of $s_{C,11}, \ldots, s_{C,(m-1)(m-1)}$; see Kleibergen (2007). For values of $s_{C,(m-1)(m-1)}$ close to $s_{C,mm}$, the limiting distribution of $\hat{s}^2_{C,mm}$ is, however, also more complicated so we, for now, refrain from analyzing such settings.

Theorem 6: *Under Assumptions 1 and 2, the drifting sequence in (24), the SVD of* $\hat{C}$:

$$\hat{C} = \hat{U}_C \hat{S}_C \hat{V}'_C, \tag{53}$$

with $\hat{U}_C = (\hat{U}_{C,1} \vdots \hat{U}_{C,m} \vdots \hat{U}_{C,2})$, $\hat{U}_{C,1} : k_f \times (m-1)$, $\hat{U}_{C,m} : k_f \times 1$, $\hat{U}_{C,2} : k_f \times (k_f - m)$, *and* $\hat{V}_C$ *orthonormal* $k_f \times k_f$ *and* $m \times m$ *dimensional matrices, and* $\hat{S}_C$ *a* $k_f \times m$ *dimensional matrix with the singular values* $\hat{s}_{C,11} \ldots \hat{s}_{C,mm}$ *in decreasing order on the main diagonal,* $\widehat{\text{IS}} = \hat{s}^2_{C,mm}$, *and large values of* $\hat{s}_{C,11} \ldots \hat{s}_{C,(m-1)(m-1)}$, *so we can consistently estimate* $\hat{U}_{C,1}$, *the LR no-identification statistic is approximately equal to:*

$$\begin{aligned} & \text{LR}(\alpha = 0) \\ = \ & \widehat{\text{IS}} - \widehat{\text{MISS}} \qquad (54) \\ = \ & \frac{1}{2}\left[\hat{s}^2_{C,mm} - \hat{a}' M_{\hat{U}_{C,1}} \hat{a} + \sqrt{\left(\hat{s}^2_{C,mm} - \hat{a}' M_{\hat{U}_{C,1}} \hat{a}\right)^2 + 4\hat{s}^2_{C,mm}(\hat{U}'_{C,m}\hat{a})^2}\right] + o_p(1). \end{aligned}$$

Proof. See the Appendix. ■

When U_1 is consistently estimable, we can prove the counterpart of Theorems 3 and 4 as well for the elements of $\text{LR}(\alpha = 0)$ in (54). For reasons of brevity we refrain from doing so and directly state the final result in Corollary 8 below.

Corollary 8. *Under Assumptions 1 and 2, the drifting sequence in (24), large singular values* $s_{C,11} \ldots s_{C,(m-1)(m-1)}$*:* $\hat{U}_{C,1}$ *is a consistent estimator of* $U_{C,1}$ *so Theorems 3-6 and Corollary 7 jointly imply that the asymptotic behaviors under* $\text{H}^*_{\text{non-ind}}$: $U'_{C,m}a = 0$, $s^2_{C,mm} = a' M_{U_{C,1}} a$, *of:*

$$\hat{s}^2_{C,mm} - \hat{a}' M_{\hat{U}_{C,1}} \hat{a}, \ \hat{U}'_{C,m}\hat{a}, \ \textit{and} \ \hat{s}^2_{C,mm} + \hat{a}' M_{\hat{U}_{C,1}} \hat{a}, \tag{55}$$

are approximately independent up to order $N^{-\frac{1}{2}}$, and therefore also that of

$$\text{LR}(\alpha = 0) \textit{ and } \hat{s}^2_{C,mm} + \hat{a}' M_{\hat{U}_{C,1}} \hat{a}. \tag{56}$$

Corollary 8 shows that we can use $\hat{s}^2_{C,mm} + \hat{a}' M_{\hat{U}_{C,1}} \hat{a}$ as a conditioning statistic for the LR test of no-identification. It equals the sum of the diagonal elements of the sample analog of the matrix from which the identification condition results in (51). When $s_{C,11} \dots s_{C,(m-1)(m-1)}$ are large, the eigenvalues of this matrix correspond with the smallest two characteristic roots of the sample analog of (25). Their sum therefore also provides an easy-to-compute conditioning statistic.

Corollary 9. *Under the conditions of Corollary 8, the sum of the smallest two eigenvalues of the sample analog of (25):*

$$\text{rk} = \widehat{\text{MISS}} + \hat{\lambda}_{aC,m}, \tag{57}$$

with $\widehat{\text{MISS}} = \hat{\lambda}_{aC,m+1}$, the smallest characteristic root of the sample analog of (25) and $\hat{\lambda}_{aC,m}$ the second smallest characteristic root, is under $\text{H}^*_{\text{non-ind}}$ approximately independent up to order $N^{-\frac{1}{2}}$ of LR($\alpha = 0$), so it provides a conditioning statistic for the LR test of no-identification.

Comparing the boundary settings and conditioning statistics under $\text{H}^*_{\text{non-ind}}$ (36) for one structural parameter, $m = 1$, and several structural parameters, $m > 1$, we then have:

	$m = 1$	$m > 1$
$\text{H}^*_{\text{non-ind}}$:	$C'C - a'a = 0$	$s^2_{C,mm} - a' M_{U_1} a = 0$
	$C'a = 0$	$U'_{C,m} a = 0$
Conditioning statistic rk:	$\hat{C}'\hat{C} + \hat{a}'\hat{a}$	$\widehat{\text{MISS}} + \hat{\lambda}_{aC,m}$

(58)

We then have two options to compute the conditional critical value function, both of which are based on consistent estimation of $U_{C,1}$ and therefore need large singular values $s_{C,11} \dots s_{C,(m-1)(m-1)}$.

1. Appendix A8 shows the higher order expansion of $\hat{a}$ and $\hat{C}$ as a function of $a^* = U_{C_1}' a$, $s_{C,11} \ldots s_{C,mm}$ and estimators based on standardized random variables $\dot{u}_i$, $\dot{V}_i$ and $\dot{Z}_i$, $i = 1, \ldots, N$. The sampling algorithm for computing the conditional critical value function of the LR test of no-identification is then:

   (a) Use the estimators $\hat{a}^* = \hat{U}_{C_1}' \hat{a}$, $\hat{s}_{C,11} \ldots \hat{s}_{C,(m-1)(m-1)}$ from the data under consideration.

   (b) Generate standardized i.i.d. random variables $\dot{u}_i$, $\dot{V}_i$ and $\dot{Z}_i$, for $i = 1, \ldots, N$.

   (c) Generate $\hat{a}$ and $\hat{C}$ for a range of values of $a' M_{U_1} a = s_{C,mm}^2$ using the expansion in Appendix A8 and the realizations of the standardized random variables from b. and the estimates of $\hat{a}^*$ and $\hat{s}_{C,11} \ldots \hat{s}_{C,(m-1)(m-1)}$ from a.

   (d) Compute LR($\alpha = 0$) and rk and use them to compute conditional distribution of LR($\alpha = 0$) given rk.

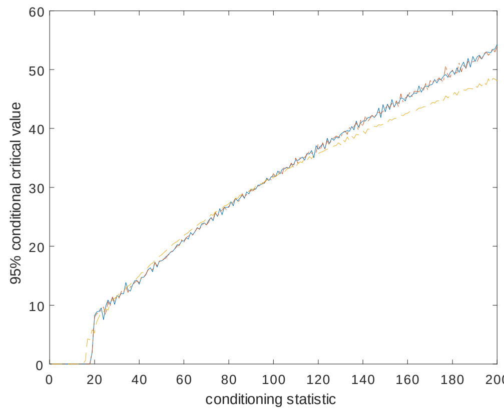
2. Use the algorithm to compute critical values from Subsection 4.4 with the expression of the LR statistic from (54) and conditioning statistic from (56).

We next apply the above algorithms to test the hypothesis of no-identification.

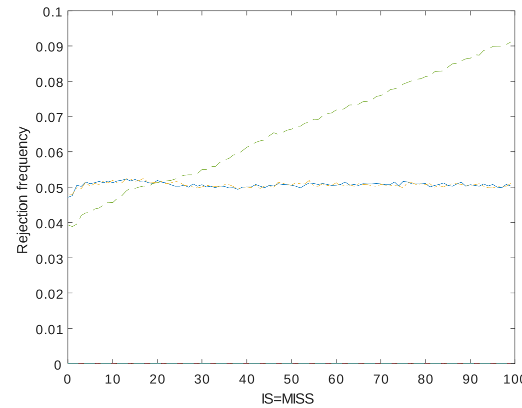
4.6 Simulation and application, $m > 1$

For the Fama-French (1993) three-factor model with and without the zero-β return using data from Lettau, Ludvigson, and Ma (2019), Panel 9.1 in Figure 9 shows the 95% conditional critical value functions that result from incorporating the zero-β return or not. Panel 9.1 shows that these critical value functions are basically identical, but differ from the one which would result when we ignore the estimation error of the covariance matrix estimators. Panel 9.2 in Figure 9 shows the resulting rejection frequencies which are very close to 5% for both specifications. Panel 9.2 also shows that usage of the conditional critical value function that does not incorporate the estimation error of the covariance matrix estimators leads to considerable size distortion.

Figure 9: Conditional critical value function for testing $\mathrm{H}^*_{\text{non-ind}}$ at the 5% significance level using LR($\alpha = 0$) as a function of rk (57) calibrated to the Fama-French (1993) three-factor model with zero-β return (solid), without (dash-dotted), $k_f = 25$, $m = 3$, $N = 201$ and known covariance (dashed) and resulting rejection frequencies.

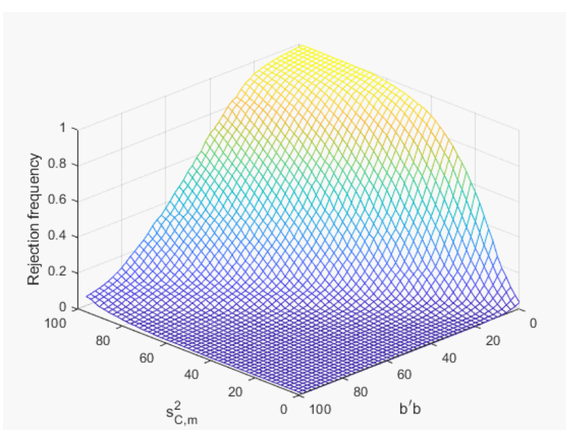


Panel 9.1. Conditional critical values



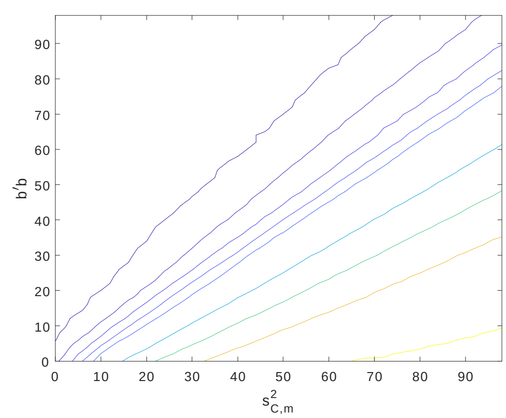
Panel 9.2. Rejection frequencies

Figure 10: Power of 5% significance LR no-identification test calibrated to the Fama-French (1993) three-factor model, $m = 3$, $k_f = 25$, $\lambda_0 \neq 0$.



Panel 10.1

$U'_{C,m} a = 0$



Panel 10.2

Contour lines at 1, 5, 10, 15, 20, 40, 60, 80 and 99%.

Figure 10 illustrates the power of the LR test calibrated to the Fama-French (1993) three-factor model with the zero-β return and using the Lettau, Ludvigson, and Ma (2019) data. Panel 10.1 shows the power surface over IS and MISS, while $U'_{C,m}a = 0$, and Panel 10.2 shows the accompanying contour lines. The contour line at 5% is very close to the 45^0 degree line, where IS = MISS, and the equivalue lines of the population value of the conditioning statistic are orthogonal to it.

Table 2: LR test of no-identification for Fama-French (1993) with market, HML and SMB factors. Significance at 1%,***; 5%,**; 10%,*.

	$\lambda_0 \neq 0$	$\lambda_0 = 0$
$\widehat{\text{IS}}$	106.81^{***}	974.39^{***}
$\widehat{\text{MISS}}$	59.34^{***}	87.47^{***}
$\text{LR}(\alpha = 0) = \widehat{\text{IS}} - \widehat{\text{MISS}}$	47.47^{*}	886.91^{***}
Conditioning statistic: $\widehat{\text{MISS}} + \hat{\lambda}_{aC,m}$	182.33	1109.47
95% conditional critical value	50.30	215.82

Table 2 shows the results of the LR no-identification test for the Fama-French (1993) three-factor model using data from Lettau, Ludvigson, and Ma (2019). Both for the specifications with and without the zero-β return, the IS and MISS statistics are strongly significant, which gives the impression that the risk premia are identified in either specification while they are also misspecified. For the specification which includes the zero-β return, the proper LR no-identification test is, however, not significant at the 5% level but just at the 10% level. It shows that we can not reject no-identification with 5% significance. For the specification which does not include the zero-β return, the risk premia are well identified as reflected by the very large value of the LR statistic, which is well above its 95% conditional critical value.

The Appendix contains additional empirical applications for the models previously presented in Figure 1 and Table 1. Unlike the Fama-French (1993) three-factor model, we can not reject no-identification for any of them.

5 Conclusions

The widely employed Jacobian rank test does not test the appropriate hypothesis of no-identification of the structural parameters in potentially misspecified linear GMM. We pin down the appropriate no-identification hypothesis and propose a conditional LR test for testing it. For applications, alongside its conditioning statistic, the conditional critical value function of the LR test depends on the sample size at hand and a few consistently estimable nuisance parameters. When applying the conditional LR no-identification test for linear asset pricing, we find that for some well known specifications the hypothesis of no-identification can not be rejected at the 5% signficance level, while the Jacobian rank test signals strong identification because it does not test the appropriate no-identification hypothesis.

The conditional critical value function of the LR no-identification test is constructed for a setting of homoskedasticity. In future work, we plan to extend it to more general covariance structures. The null distribution is then much more difficult to pin down as it depends on more parameters. For reasons of brevity, we therefore refrained from analyzing it in the current paper.

References

[1] Adrian, T., E. Etula and T. Muir. Financial Intermediaries and the Cross-Section of Asset Returns. *Journal of Finance*, **69**:2557–2596, 2014.

[2] Andrews, D.W.K. and X. Cheng. Estimation and inference with Weak, Semi-Strong and Strong Identification. *Econometrica*, **80**:2153–2211, 2012.

[3] Andrews, D.W.K., M.J. Moreira and J.H. Stock. Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression. *Econometrica*, **74**:715–752, 2006.

[4] Andrews, I. On the structure of IV estimands. *Journal of Econometrics*, 211(1):294–307, 2019.

[5] Andrews, I. and A. Mikusheva. Conditional inference with a functional nuisance parameter. *Econometrica*, **84**:1571–1612, 2016.

[6] Cragg, J.C. and S.G. Donald. Inferring the rank of a matrix. *Journal of Econometrics*, **76**:223–250, 1997.

[7] Dufour, J.-M. Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica*, **65**:1365–388, 1997.

[8] Fama, E.F. and K.R. French. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, **33**:3–56, 1993.

[9] Gospodinov, N. R. Kan and C. Robotti. Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Factors. *Review of Financial Studies*, **27**:2139–2170, 2014.

[10] Hall, A.R. and A. Inoue. The Large Sample Behaviour of the Generalized Method of Moments Esimator in Misspecified Models. *Journal of Econometrics*, **50**:1029–1054, 2003.

[11] Hansen, B.E. and S. Lee. Inference for Iterated GMM under Misspecification. *Econometrica*, 89(3):1419–1447, 2021.

[12] Hansen, L.P. Large Sample Properties of Generalized Method Moments Estimators. *Econometrica*, **50**:1029–1054, 1982.

[13] Hausman, J.A. Specification and Estimation of Simultaneous Equations Systems. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, volume 1*. Elsevier Science (Amsterdam), 1983.

[14] He, Z., B. Kelly and A. Manela. Intermediary Asset Pricing: New Evidence from Many Asset Classes. *Journal of Financial Economics*, **126**:1–35, 2017.

[15] Hillier, G.. Exact properties of the conditional likelihood ratio test in an IV regression model. *Econometric Theory*, 25(4):915–957, 2009.

[16] Jagannathan, R. and Z. Wang. The Conditional CAPM and the Cross-Section of Expected Returns. *Journal of Finance*, **51**:3–53, 1996.

[17] Kan, R., C. Robotti and J. Shanken. Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology. *Journal of Finance*, **68**:2617–2649, 2013.

[18] Kleibergen, F. Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression. *Econometrica*, **70**:1781–1803, 2002.

[19] Kleibergen, F. Testing Parameters in GMM without assuming that they are identified. *Econometrica*, **73**:1103–1124, 2005.

[20] Kleibergen, F. Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics*, **139**:181–216, 2007.

[21] Kleibergen, F. and R. Paap. Generalized Reduced Rank Tests using the Singular Value Decomposition. *Journal of Econometrics*, **133**:97–126, 2006.

[22] Kleibergen, F. and S. Mavroeidis. Weak instrument robust tests in GMM and the new Keynesian Phillips curve. *Journal of Business and Economic Statistics*, **27**:293–311, 2009.

[23] Kleibergen, F. and Z. Zhan. Robust Inference for Consumption-Based Asset Pricing. *Journal of Finance*, **75**:507–550, 2020.

[24] Kleibergen, F. and Z. Zhan. Double robust inference for continuous updating GMM. *Quantitative Economics*, 16(1):295–327, 2025.

[25] Kroencke, T.A. Asset Pricing without Garbage. *Journal of Finance*, **72**:47–98, 2017.

[26] Lee, S. A consistent variance estimator for 2SLS when instruments identify different LATEs. *Journal of Business and Economic Statistics*, **36**:400–410, 2018.

[27] Lettau, M. and S. Ludvigson. Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia are Time-Varying. *Journal of Political Economy*, **109**:1238–1287, 2001.

[28] Lettau, M., S.C. Ludvigson and S. Ma. Capital Share Risk in U.S. Asset Pricing. *Journal of Finance*, **74**:1753–1792, 2019.

[29] Lewellen, J., S. Nagel and J. Shanken. A Skeptical Appraisal of Asset-Pricing Tests. *Journal of Financial Economics*, **96**:175–194, 2010.

[30] Maasoumi, E.. How to live with misspecification if you must. *Journal of Econometrics*, **44**:67–86, 1990.

[31] Maasoumi, E. and P. Phillips. On the behavior of inconsistent instrumental variable estimators. *Journal of Econometrics*, **19**:183–201, 1982.

[32] Moreira, M.J.,. A Conditional Likelihood Ratio Test for Structural Models. *Econometrica*, **71**:1027–1048, 2003.

[33] Robin, J.-M. and R.J. Smith. Tests of Rank. *Econometric Theory*, **16**:151–175, 2000.

[34] Savov, A. Asset Pricing with Garbage. *Journal of Finance*, **72**:47–98, 2011.

[35] Staiger, D. and J.H. Stock. Instrumental Variables Regression with Weak Instruments. *Econometrica*, **65**:557–586, 1997.

[36] Stock, J.H. and J.H. Wright. GMM with Weak Identification. *Econometrica*, **68**:1055–1096, 2000.

[37] White, H. *Asymptotic Theory for Econometricians.* Academic Press, 1984.

[38] Yogo M. A consumption-based explanation of expected stock returns. *Journal of Finance*, **61**:539–580, 2006.

Supplemental Appendix for "Testing for Identification in Potentially Misspecified Linear GMM"

Page

A1. Specifications for Figure 1 and Table 1. The specifications of the eight different specifications used for Figure 1 and Table 1 are:

1. Fama and French (1993), the prominent three, so-called Fama-French, factors: the market return R_m, SMB (small minus big), and HML (high minus low). We use quarterly data from Lettau, Ludvigson, and Ma (2019) over 1963Q3 to 2013Q4 for the three factors, and the twenty-five size and book-to-market sorted portfolios as test assets.

2. Jagannathan and Wang (1996), three factors: R_m, corporate bond yield spread, and per capita labor income growth. We use their monthly data from July 1963 to December 1990 so $T = 330$, while one hundred size and beta sorted portfolios are used as test assets.

3. Yogo (2006), three factors: R_m, durable and nondurable consumption growth. The sample period is 1951Q1 to 2001Q4 so $T = 204$, with twenty-five size and book-to-market sorted portfolios as test assets.

4. Lettau and Ludvigson (2001), three factors: (lagged) consumption-wealth ratio, consumption growth, and their interaction. We use quarterly data from 1963Q3 to 1998Q3 so $T = 141$, while the test assets are the twenty-five Fama-French portfolios.

5. Savov (2011), one factor: garbage growth. We use the same annual data, 1960 - 2006, while the test assets are the twenty-five Fama-French portfolios augmented by the ten industry portfolios, as suggested by Lewellen, Nagel, and Shanken (2010).

6. Adrian, Etula, and Muir (2014), one factor: leverage. Following Lettau, Ludvigson, and Ma (2019), we extend the time period to 1963Q3 - 2013Q4, and use twenty-five size and book-to-market sorted portfolios as test assets.

7. Kroencke (2017), one factor: unfiltered annual consumption growth. We use the postwar 1960 - 2014 sample from Kroencke (2017), while thirty portfolios, sorted by size, value and investment alongside the market portfolio, are used as test assets.

8. He, Kelly, and Manela (2017), two factors: banking equity-capital ratio and R_m. The data are also taken from Lettau, Ludvigson, and Ma (2019) for the period 1963Q3 - 2013Q4, and twenty-five size and book-to-market sorted portfolios are the test assets.

A2. MISS and IS for linear asset pricing models. Let $R_{i,t}$ be the return on the i-th asset at time t, with $i = 1, ..., N$, and $t = 1, ..., T$. The beta representation of expected returns models it as linear in the beta vector of factor loadings:

$$E(R_{i,t}) = \beta_i' \lambda_F,$$

where λ_F is the $K \times 1$ vector of risk premia, and β_i is the $K \times 1$ vector of factor loadings:

$$\beta_i = var(F_t)^{-1} cov(F_t, R_{i,t}),$$

with F_t the $K \times 1$ vector of the specified risk factors. We can as well represent the beta representation jointly for all assets by stacking the N equations to obtain

$$E(R_t) = \beta \lambda_F,$$

with $R_t = (R_{1,t}, ..., R_{N,t})'$, $\beta = (\beta_1, ..., \beta_N)'$.[11]

A scalar λ_0 is often added to the beta representation, so

$$E(R_t) = \iota_N \lambda_0 + \beta \lambda_F,$$

with ι_N the $N \times 1$ vector of ones, and λ_0 the so-called zero-beta return, or the expected return on an asset with no exposure to priced risks. The $\lambda_0 = 0$ restriction can be achieved by considering R_t as the excess return.

To calculate $\widehat{\text{IS}}$ and $\widehat{\text{MISS}}$, we consider the following two cases.

1. If $\lambda_0 = 0$ is imposed: consider R_t as the observed $N \times 1$ vector of asset returns, and F_t as the $K \times 1$ vector of risk factors.

[11] Here we use the standard notation for linear asset pricing models, so N and K play the roles of k_f and m in the GMM notation used in the main text, while T is the sample size.

2. If $\lambda_0 = 0$ is not imposed: consider $\mathcal{R}_t = (\mathcal{R}_{1,t} \ldots \mathcal{R}_{N+1,t})'$as the $(N+1) \times 1$ vector of asset returns; F_t is the $K \times 1$ vector of risk factors, $t = 1, ..., T$. By subtracting the $(N+1)$-th asset return, we obtain the $N \times 1$ column vector R_t:

$$R_t = (\mathcal{R}_{1,t} \ldots \mathcal{R}_{N,t})' - \iota_N \mathcal{R}_{N+1,t}.$$

The choice of the $(N+1)$-th asset does not affect $\widehat{\text{IS}}$ and $\widehat{\text{MISS}}$ statistics; see Kleibergen and Zhan (2020).

For both cases above, we use the auxiliary linear factor model $R_t = \alpha + \beta F_t + u_t$, which yields

$$\hat{\beta} = \sum_{t=1}^{T} \bar{R}_t \bar{F}_t' \left(\sum_{t=1}^{T} \bar{F}_t \bar{F}_t' \right)^{-1}, \ \hat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t', \ \text{ and } \ \hat{Q}_{\bar{F}\bar{F}} = \frac{1}{T} \sum_{t=1}^{T} \bar{F}_t \bar{F}_t',$$

where $\bar{F}_t = F_t - \bar{F}$, $\bar{F} = \frac{1}{T} \sum_{t=1}^{T} F_t$, $\bar{R}_t = R_t - \bar{R}$, $\bar{R} = \frac{1}{T} \sum_{t=1}^{T} R_t$, and $\hat{u}_t = (R_t - \bar{R}) - \hat{\beta}(F_t - \bar{F})$ is the residual at time t.

Let τ_{min} be the smallest root of

$$\left| \tau \hat{Q}_{\bar{F}\bar{F}}^{-1} - \hat{\beta}' \hat{\Omega}^{-1} \hat{\beta} \right| = 0,$$

which is identical to the smallest eigenvalue of the matrix $\hat{Q}_{\bar{F}\bar{F}} \hat{\beta}' \hat{\Omega}^{-1} \hat{\beta}$. The $\widehat{\text{IS}}$ statistic reads:

$$\widehat{\text{IS}} = T \times \tau_{min}.$$

Similarly, let λ_{min} be the smallest root of

$$\left| \lambda \begin{pmatrix} 1 & 0 \\ 0 & \hat{Q}_{\bar{F}\bar{F}}^{-1} \end{pmatrix} - \begin{pmatrix} \bar{R} & \hat{\beta} \end{pmatrix}' \hat{\Omega}^{-1} \begin{pmatrix} \bar{R} & \hat{\beta} \end{pmatrix} \right| = 0,$$

which is equal to the smallest eigenvalue of $\begin{pmatrix} 1 & 0 \\ 0 & \hat{Q}_{\bar{F}\bar{F}} \end{pmatrix} \begin{pmatrix} \bar{R} & \hat{\beta} \end{pmatrix}' \hat{\Omega}^{-1} \begin{pmatrix} \bar{R} & \hat{\beta} \end{pmatrix}$. The misspecification J-statistic ($\widehat{\text{MISS}}$) reads:

$$J = \widehat{\text{MISS}} = T \times \lambda_{min}.$$

A3. Proof of Theorem 2. We first specify the different components using the asymptotically independent building blocks that make up the weak instrument robust statistics:

$$\begin{aligned}
\hat{\mu}_f(\theta) =& \ \hat{\mu}_f(0) + \hat{J}(0)\theta \\
\hat{D}(\theta) =& \ \hat{J}(0) - \hat{\mu}_f(\theta)\tfrac{\omega_{\mu J}(\theta)}{\omega_{\mu\mu}(\theta)} \\
\hat{D}(\theta = -\Omega_{JJ}^{-1}\omega_{J\mu}) =& \ \hat{J}(0), \\
\hat{\mu}_f(\theta = -\Omega_{JJ}^{-1}\omega_{J\mu}) =& \ \hat{\mu}_J(0),
\end{aligned}$$

with $\omega_{\mu J}(\theta) = \omega_{\mu J} + \theta'\Omega_{JJ}$, $\omega_{\mu\mu}(\theta) = \omega_{\mu\mu} + 2\omega_{\mu J}\theta + \theta'\Omega_{JJ}\theta$, so $\omega_{\mu J}(\theta = -\Omega_{JJ}^{-1}\omega_{J\mu}) = 0$. We have that:

$$\begin{aligned}
\sqrt{N}\left(\hat{\mu}_f(\theta) - \mu_f(\theta)\right) &\underset{d}{\rightarrow} \psi_\mu(\theta) \sim N(0, \omega_{\mu\mu}(\theta)\Sigma), \\
\sqrt{N}\text{vec}\left(\hat{D}(\theta) - D(\theta)\right) &\underset{d}{\rightarrow} \psi_\theta(\theta) \sim N(0, \Omega_{JJ.\mu}(\theta) \otimes \Sigma),
\end{aligned}$$

where $\psi_\mu(\theta)$ and $\psi_\theta(\theta)$ are independently distributed, $\mu_f(\theta) = \mu_f(0) + J(0)\theta$, $D(\theta) = J(0) - \mu_f(\theta)\frac{\omega_{\mu J}(\theta)}{\omega_{\mu\mu}(\theta)}$, $\Omega_{JJ.\mu}(\theta) = \Omega_{JJ} - \omega_{\mu J}(\theta)'\omega_{\mu\mu}(\theta)^{-1}\omega_{\mu J}(\theta)$, from which we next have that for $\theta^1 = -\Omega_{JJ}^{-1}\omega_{J\mu}$, $\Omega_{JJ.\mu}(\theta^1) = \Omega_{JJ}$, $\omega_{\mu\mu}(\theta^1) = \omega_{\mu\mu.J}$, $D(\theta^1) = J(0)$, $\omega_{\mu J}(\theta) = (\theta - \theta^1)'\Omega_{JJ}$, $\omega_{\mu\mu}(\theta) = \omega_{\mu\mu.J} + (\theta - \theta^1)'\Omega_{JJ}(\theta - \theta^1)$, $\Omega_{JJ.\mu}(\theta) = \Omega_{JJ} - \omega_{\mu J}(\theta)'\omega_{\mu\mu}(\theta)^{-1}\omega_{\mu J}(\theta) = \Omega_{JJ} - \Omega_{JJ}(\theta - \theta^1)\left[\omega_{\mu\mu.J} + (\theta - \theta^1)'\Omega_{JJ}(\theta - \theta^1)\right]^{-1}(\theta - \theta^1)'\Omega_{JJ}$, $\Omega_{JJ.\mu}(\theta) = \frac{1}{\Omega_{JJ}^{-1} + (\theta - \theta^1)\omega_{\mu\mu.J}^{-1}(\theta - \theta^1)'} = \frac{\omega_{\mu\mu.J}\Omega_{JJ}}{\omega_{\mu\mu.J} + (\theta - \theta^1)\Omega_{JJ}(\theta - \theta^1)'}$:

$$\begin{aligned}
\sqrt{N}\left(\hat{D}(\theta^1) - D(\theta^1)\right) &\underset{d}{\rightarrow} \psi_\theta(\theta^1) \Leftrightarrow \\
\sqrt{N}\left(\hat{D}(\theta^1) - D(\theta) - \mu_f(\theta)\tfrac{\omega_{\mu J}(\theta)}{\omega_{\mu\mu}(\theta)}\right) &\underset{d}{\rightarrow} \psi_\theta(\theta^1) \Leftrightarrow
\end{aligned}$$

$$\begin{aligned}
\sqrt{N}\Sigma^{-\frac{1}{2}}\hat{D}(\theta^1) &\underset{d}{\rightarrow} \sqrt{N}\Sigma^{-\frac{1}{2}}D(\theta) + \sqrt{N}\Sigma^{-\frac{1}{2}}\mu_f(\theta)\tfrac{(\theta - \theta^1)'\Omega_{JJ}}{\omega_{\mu\mu.J} + (\theta - \theta^1)'\Omega_{JJ}(\theta - \theta^*)} + \Sigma^{-\frac{1}{2}}\psi_\theta(\theta^1) \\
&= \bar{D}(\theta)\Omega_{JJ.\mu}(\theta)^{\frac{1}{2}} + \bar{\mu}(\theta)\tfrac{(\theta - \theta^1)'\Omega_{JJ}}{\sqrt{\omega_{\mu\mu.J} + (\theta - \theta^1)'\Omega_{JJ}(\theta - \theta^1)}} + \Sigma^{-\frac{1}{2}}\psi_\theta(\theta^1) \\
&= \bar{D}(\theta)\tfrac{\sqrt{\omega_{\mu\mu.J}\Omega_{JJ}}}{\sqrt{\omega_{\mu\mu.J} + (\theta - \theta^1)\Omega_{JJ}(\theta - \theta^1)'}} + \bar{\mu}(\theta)\tfrac{(\theta - \theta^1)'\Omega_{JJ}}{\sqrt{\omega_{\mu\mu.J} + (\theta - \theta^1)'\Omega_{JJ}(\theta - \theta^1)}} + \Sigma^{-\frac{1}{2}}\psi_\theta(\theta^1) \\
&= \left[\bar{D}(\theta)\sqrt{\omega_{\mu\mu.J}} + \bar{\mu}(\theta)(\theta - \theta^1)\Omega_{JJ}^{\frac{1}{2}}\right]\sqrt{\tfrac{\Omega_{JJ}}{\omega_{\mu\mu}(\theta)}} + \Sigma^{-\frac{1}{2}}\psi_\theta(\theta^1)
\end{aligned}$$

for $\bar{D}(\theta) = \sqrt{N}\Sigma^{-\frac{1}{2}}D(\theta)\Omega_{JJ.\mu}(\theta)^{-\frac{1}{2}}$, $\bar{\mu}(\theta) = \sqrt{N}\Sigma^{-\frac{1}{2}}\mu_f(\theta)\omega_{\mu\mu}(\theta)^{-\frac{1}{2}}$.

Similarly, using $1-\frac{\omega_{\mu J}(\theta)}{\omega_{\mu\mu}(\theta)}(\theta-\theta^1)=\frac{\omega_{\mu\mu.J}}{\omega_{\mu\mu.J}+(\theta-\theta^1)'\Omega_{JJ}(\theta-\theta^1)}$:

$$\begin{aligned}
\sqrt{N}\left(\hat{\mu}_f(\theta^1)-\mu_f(\theta^1)\right) &\underset{d}{\rightarrow} \psi_\mu(\theta^1) \Leftrightarrow \\
\sqrt{N}\left(\hat{\mu}_f(\theta^1)-\mu_f(\theta)+J(0)(\theta-\theta^1)\right) &\underset{d}{\rightarrow} \psi_\mu(\theta^1) \Leftrightarrow \\
\sqrt{N}\left(\hat{\mu}_f(\theta^1)-\mu_f(\theta)+\left(D(\theta)+\mu_f(\theta)\frac{\omega_{\mu J}(\theta)}{\omega_{\mu\mu}(\theta)}\right)(\theta-\theta^1)\right) &\underset{d}{\rightarrow} \psi_\mu(\theta^1) \Leftrightarrow \\
\sqrt{N}\left(\hat{\mu}_f(\theta^1)-\mu_f(\theta)(1-\frac{\omega_{\mu J}(\theta)}{\omega_{\mu\mu}(\theta)}(\theta-\theta^1))+D(\theta)(\theta-\theta^1)\right) &\underset{d}{\rightarrow} \psi_\mu(\theta^1) \Leftrightarrow \\
\sqrt{N}\left(\hat{\mu}_f(\theta^1)-\left[\mu_f(\theta)\frac{\omega_{\mu\mu.J}}{\omega_{\mu\mu}(\theta)}-D(\theta)(\theta-\theta^1)\right]\right) &\underset{d}{\rightarrow} \psi_\mu(\theta^1) \Leftrightarrow
\end{aligned}$$

$$\begin{aligned}
\sqrt{N}\Sigma^{-\frac{1}{2}}\hat{\mu}_f(\theta^1) &\underset{d}{\rightarrow} \sqrt{N}\Sigma^{-\frac{1}{2}}\left[\mu_f(\theta)\frac{\omega_{\mu\mu.J}}{\omega_{\mu\mu}(\theta)}-D(\theta)(\theta-\theta^1)\right]+\Sigma^{-\frac{1}{2}}\psi_\mu(\theta^1) \\
&= \bar{\mu}(\theta)\frac{\omega_{\mu\mu.J}}{\sqrt{\omega_{\mu\mu}(\theta)}}-\bar{D}(\theta)\Omega_{JJ.\mu}(\theta)^{\frac{1}{2}}(\theta-\theta^1)+\Sigma^{-\frac{1}{2}}\psi_\mu(\theta^1) \\
&= \left[\bar{\mu}(\theta)\sqrt{\omega_{\mu\mu.J}}-\bar{D}(\theta)(\theta-\theta^1)\Omega_{JJ}^{\frac{1}{2}}\right]\sqrt{\frac{\omega_{\mu\mu.J}}{\omega_{\mu\mu}(\theta)}}+\Sigma^{-\frac{1}{2}}\psi_\mu(\theta^1).
\end{aligned}$$

Combining the expressions above, we then have when the DGP has the CUE pseudo-true value θ^* :

$$\begin{aligned}
& N\hat{J}(0)'\Sigma^{-1}\hat{\mu}_J(0) \\
=\ & N\hat{D}(\theta^1)'\Sigma^{-1}\hat{\mu}(\theta^1) \\
\underset{d}{\rightarrow}\ & \left[\left[\bar{D}(\theta^*)\sqrt{\omega_{\mu\mu.J}}+\bar{\mu}(\theta^*)(\theta^*-\theta^1)\Omega_{JJ}^{\frac{1}{2}}\right]\sqrt{\frac{\Omega_{JJ}}{\omega_{\mu\mu}(\theta^*)}}+\Sigma^{-\frac{1}{2}}\psi_\theta(\theta^1)\right]' \\
& \left[\left[\bar{\mu}(\theta^*)\sqrt{\omega_{\mu\mu.J}}-\bar{D}(\theta^*)(\theta^*-\theta^1)\Omega_{JJ}^{\frac{1}{2}}\right]\sqrt{\frac{\omega_{\mu\mu.J}}{\omega_{\mu\mu}(\theta^*)}}+\Sigma^{-\frac{1}{2}}\psi_\mu(\theta^1)\right]
\end{aligned}$$

Because θ^* is a CUE pseudo-true value, $\bar{D}(\theta^*)'\bar{\mu}(\theta^*)\equiv 0$. Thus, we get:

$$\begin{aligned}
& N\hat{J}(0)'\Sigma^{-1}\hat{\mu}_J(0) \\
\underset{d}{\rightarrow}\ & \left[\bar{\mu}(\theta^*)'\bar{\mu}(\theta^*)-\bar{D}(\theta^*)'\bar{D}(\theta^*)\right]\frac{\omega_{\mu\mu.J}\Omega_{JJ}}{\omega_{\mu\mu}(\theta)}(\theta^*-\theta^1)+ \\
& \sqrt{\frac{\Omega_{JJ}}{\omega_{\mu\mu}(\theta^*)}}\left[\bar{D}(\theta^*)\sqrt{\omega_{\mu\mu.J}}+\bar{\mu}(\theta^*)(\theta^*-\theta^1)\Omega_{JJ}^{\frac{1}{2}}\right]'\Sigma^{-\frac{1}{2}}\psi_\mu(\theta^1)+ \\
& \sqrt{\frac{\omega_{\mu\mu.J}}{\omega_{\mu\mu}(\theta^*)}}\left[\bar{\mu}(\theta^*)\sqrt{\omega_{\mu\mu.J}}-\bar{D}(\theta)(\theta^*-\theta^1)\Omega_{JJ}^{\frac{1}{2}}\right]'\Sigma^{-\frac{1}{2}}\psi_\theta(\theta^1)+\psi_\theta(\theta^1)'\Sigma^{-1}\psi_\mu(\theta^1).
\end{aligned}$$

Hence,

$$
\begin{aligned}
& N\Omega_{JJ}^{-\frac{1}{2}}\hat{J}(0)'\Sigma^{-1}\hat{\mu}_J\omega_{\mu\mu.J}^{-\frac{1}{2}} \\
\underset{d}{\rightarrow} \; & \left[\bar{\mu}(\theta^*)'\bar{\mu}(\theta^*) - \bar{D}(\theta^*)'\bar{D}(\theta^*)\right] \frac{\omega_{\mu\mu.J}^{\frac{1}{2}}\Omega_{JJ}^{\frac{1}{2}}}{\omega_{\mu\mu}(\theta^*)}(\theta^* - \theta^1) + \\
& \frac{1}{\sqrt{\omega_{\mu\mu.J}}}\left[\bar{D}(\theta^*)\sqrt{\omega_{\mu\mu.J}} + \bar{\mu}(\theta^*)(\theta^* - \theta^1)\Omega_{JJ}^{\frac{1}{2}}\right]' \Sigma^{-\frac{1}{2}} \frac{1}{\sqrt{\omega_{\mu\mu}(\theta^*)}}\psi_\mu(\theta^1) + \\
& \frac{\sqrt{\omega_{\mu\mu.J}}}{\omega_{\mu\mu}(\theta^*)}\left[\bar{\mu}(\theta^*)\sqrt{\omega_{\mu\mu.J}} - \bar{D}(\theta^*)(\theta^* - \theta^1)\Omega_{JJ}^{\frac{1}{2}}\right]' \Sigma^{-\frac{1}{2}}\sqrt{\frac{\omega_{\mu\mu}(\theta^*)}{\omega_{\mu\mu.J}\Omega_{JJ}}}\psi_\theta(\theta^1) + \\
& \left[\sqrt{\frac{\omega_{\mu\mu}(\theta^*)}{\omega_{\mu\mu.J}\Omega_{JJ}}}\psi_\theta(\theta^1)\right]' \Sigma^{-1}\left[\frac{1}{\sqrt{\omega_{\mu\mu}(\theta^*)}}\psi_\mu(\theta^1)\right]. \\
= \; & \left[\bar{\mu}(\theta^*)'\bar{\mu}(\theta^*) - \bar{D}(\theta^*)'\bar{D}(\theta^*)\right] \frac{\omega_{\mu\mu.J}^{\frac{1}{2}}\Omega_{JJ}^{\frac{1}{2}}}{\omega_{\mu\mu}(\theta^*)}(\theta^* - \theta^1) + \\
& \frac{1}{\sqrt{\omega_{\mu\mu.J}}}\left[\bar{D}(\theta^*)\sqrt{\omega_{\mu\mu.J}} + \bar{\mu}(\theta^*)(\theta^* - \theta^1)\Omega_{JJ}^{\frac{1}{2}}\right]' \psi_\mu^*(\theta^1) + \\
& \frac{\sqrt{\omega_{\mu\mu.J}}}{\omega_{\mu\mu}(\theta^*)}\left[\bar{\mu}(\theta^*)\sqrt{\omega_{\mu\mu.J}} - \bar{D}(\theta^*)(\theta^* - \theta^1)\Omega_{JJ}^{\frac{1}{2}}\right]' \psi_\theta^*(\theta^1) + \\
& \psi_\theta^*(\theta^1)'\psi_\mu^*(\theta^1),
\end{aligned}
$$

for $\psi_\theta^*(\theta^1)$ and $\psi_\mu^*(\theta^1)$ independent k_f dimensional standard normal random vectors.

This shows that when expressing tests for the two-stage pseudo-true value to equal $-\Omega_{JJ}^{-1}\omega_{J\mu}$ using the CUE pseudo-true value, these tests can only have power when $\bar{D}(\theta^*)'\bar{D}(\theta^*) > \bar{\mu}(\theta^*)'\bar{\mu}(\theta^*)$ which corresponds with IS $>$ MISS.

A4. Algorithm to compute the initial estimate of the conditional critical value function of the LR test of $\mathbf{H}^*_{\textbf{non-ind}}$. The algorithm to compute the critical value function of the LR test of $\text{H}^*_{\text{non-ind}}$ (36) uses the entier function, [..], and the first and second columns of I_{k_f} indicated by e_{1,k_f} and e_{2,k_f} respectively:

- Set all elements of the array "sum" to zero and for a range of values of c from $0...c_{\max}$, $a = \sqrt{c}e_{1,k_f}$, $C = \sqrt{c}e_{2,k_f}$ and set up the two-dimensional array Z :

  1. Generate ψ_J^* and $\psi_{\mu.J}^*$ from independent $N(0, I_{k_f})$ distributions;
  2. Compute $\hat{a} = a + \psi_{\mu.J}^*$, $\hat{C} = C + \psi_J^*$;
  3. Compute $\text{LR}(\alpha = 0) = \frac{1}{2}\left[\hat{C}'\hat{C} - \hat{a}'\hat{a} + \sqrt{\left(\hat{C}'\hat{C} - \hat{a}'\hat{a}\right)^2 + 4(\hat{C}'\hat{a})^2}\right]$;
  4. Compute conditioning statistic: rk$= \hat{a}'\hat{a} + \hat{C}'\hat{C}$ and $i = [\text{rk}]$;
  5. $\text{sum}_i = \text{sum}_i + 1$;

6. Set: $Z(i, sum_i)$ =LR($\alpha = 0$);

- Sort $Z(i,:)$ in ascending order;
- The critical value $cv(r,\alpha)$ equals $(1-\alpha)\times 100$-th percentile of sorted $Z(r,:)$.

A5. Proof of Theorem 4. We specify the covariance matrix estimators $\hat{\Omega} = \begin{pmatrix} \hat{\omega}_{\mu\mu} & \hat{\omega}_{\mu J} \\ \hat{\omega}_{J\mu} & \hat{\Omega}_{JJ} \end{pmatrix}$ and $\hat{\Sigma}$ as

$$\begin{array}{llllll} \hat{\Omega} = & \Omega^{\frac{1}{2}\prime}\dot{\Omega}\Omega^{\frac{1}{2}}, & \hat{\Omega}^{-1} = & \Omega^{-\frac{1}{2}}\dot{\Omega}^{-1}\Omega^{-\frac{1}{2}\prime} = \hat{\Omega}^{-\frac{1}{2}\prime}\hat{\Omega}^{-\frac{1}{2}}, & \hat{\Omega}^{-\frac{1}{2}} = & \dot{\Omega}^{-\frac{1}{2}}\Omega^{-\frac{1}{2}\prime}, \\ \hat{\Sigma} = & \Sigma^{\frac{1}{2}}\dot{\Sigma}\Sigma^{\frac{1}{2}\prime}, & \hat{\Sigma}^{-1} = & \Sigma^{-\frac{1}{2}\prime}\dot{\Sigma}^{-1}\Sigma^{-\frac{1}{2}} = \hat{\Sigma}^{-\frac{1}{2}\prime}\hat{\Sigma}^{-\frac{1}{2}}, & \hat{\Sigma}^{-\frac{1}{2}} = & \dot{\Sigma}^{-\frac{1}{2}}\Sigma^{-\frac{1}{2}}, \end{array}$$

so $\dot{\Omega} = \frac{1}{N}\sum_{i=1}^{N}\binom{\dot{u}_i}{\dot{V}_i}\binom{\dot{u}_i}{\dot{V}_i}' = \begin{pmatrix} \dot{\omega}_{\mu\mu} & \vdots & \dot{\omega}_{\mu J} \\ \dot{\omega}_{J\mu} & \vdots & \dot{\Omega}_{JJ} \end{pmatrix}$, $\dot{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}\dot{Z}_i\dot{Z}_i'$, with $\binom{\dot{u}_i}{\dot{V}_i} = \Omega^{-\frac{1}{2}\prime}\binom{u_i}{V_i}$, $\dot{u}_i = (u_i - \omega_{\mu J}\Omega_{JJ}^{-1}V_i)\omega_{\mu\mu.J}^{-\frac{1}{2}}$, $\dot{V}_i = \Omega_{JJ}^{-\frac{1}{2}\prime}V_i$, $\dot{Z}_i = \Sigma^{-\frac{1}{2}}Z_i$, with $\Omega^{\frac{1}{2}} = \begin{pmatrix} \omega_{\mu\mu.J}^{\frac{1}{2}} & \vdots & 0 \\ \Omega_{JJ}^{-\frac{1}{2}}\omega_{J\mu} & \vdots & \Omega_{JJ}^{\frac{1}{2}} \end{pmatrix}$, $\dot{\omega}_{\mu\mu.J} = \dot{\omega}_{\mu\mu} - \dot{\omega}_{\mu J}\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}$. For $\begin{pmatrix}\mu_f(0) \vdots J(0)\end{pmatrix} = \frac{1}{\sqrt{N}}\Sigma^{\frac{1}{2}}(a \vdots C)\Omega^{\frac{1}{2}}$, $\mu_f(0) = \frac{1}{\sqrt{N}}\Sigma^{\frac{1}{2}}a\omega_{\mu\mu.J}^{\frac{1}{2}} + \frac{1}{\sqrt{N}}\Sigma^{\frac{1}{2}}C\Omega_{JJ}^{-\frac{1}{2}}\omega_{J\mu}$, $J(0) = \frac{1}{\sqrt{N}}\Sigma^{\frac{1}{2}}C\Omega_{JJ}^{\frac{1}{2}}$, $\mu_J(0) = \mu_f(0) - J(0)\Omega_{JJ}^{-1}\omega_{J\mu} = \frac{1}{\sqrt{N}}\Sigma^{\frac{1}{2}}a\omega_{\mu\mu.J}^{\frac{1}{2}}$.

Under i.i.d. data and finite moments, the standardized random components $\dot{\omega}_{\mu\mu}$, $\dot{\omega}_{J\mu}$, $\dot{\Omega}_{JJ}$ and $\dot{\Sigma}$ converge according to:

$$\sqrt{N}\begin{pmatrix} \dot{\omega}_{\mu\mu} - 1 \\ \dot{\omega}_{J\mu} \\ \text{vech}(\dot{\Omega}_{JJ} - I_m) \\ \text{vech}(\dot{\Sigma} - I_{k_f}) \\ \frac{1}{N}\sum_{i=1}^{N}\dot{Z}_i\dot{u}_i \\ \frac{1}{N}\sum_{i=1}^{N}\dot{Z}_i\dot{V}_i \end{pmatrix} \underset{d}{\rightarrow} \begin{pmatrix} \psi_{\mu\mu} \\ \psi_{J\mu} \\ \psi_{JJ} \\ \psi_{\Sigma} \\ \psi^*_{\mu.J} \\ \psi^*_J \end{pmatrix},$$

with $\psi_{\mu\mu}$, $\psi_{J\mu}$, ψ_{JJ} and ψ_Σ : 1, m, $\frac{1}{2}m(m+1)$ and $\frac{1}{2}k_f(k_f+1)$ dimensional mean zero normally distributed random vectors, and $\psi^*_{\mu.J}$, ψ^*_J independent normal k_f dimensional random vectors.

The different elements of the covariance matrix estimator $\hat{\Omega}$ can then be expressed as:

$$\begin{aligned}
\hat{\omega}_{\mu\mu} =& \ \omega_{\mu\mu.J}\dot{\omega}_{\mu\mu} + \omega_{\mu J}\Omega_{JJ}^{-\frac{1}{2}\prime}\dot{\omega}_{J\mu}\omega_{\mu\mu.J}^{\frac{1}{2}} + \omega_{\mu\mu.J}^{\frac{1}{2}}\dot{\omega}_{\mu J}\Omega_{JJ}^{-\frac{1}{2}}\omega_{J\mu} + \omega_{\mu J}\Omega_{JJ}^{-\frac{1}{2}\prime}\dot{\Omega}_{JJ}\Omega_{JJ}^{-\frac{1}{2}}\omega_{J\mu} \\
\hat{\omega}_{J\mu} =& \ \Omega_{JJ}^{\frac{1}{2}\prime}\dot{\omega}_{J\mu}\omega_{\mu\mu.J}^{\frac{1}{2}} + \Omega_{JJ}^{\frac{1}{2}\prime}\dot{\Omega}_{JJ}\Omega_{JJ}^{-\frac{1}{2}}\omega_{J\mu} \\
\hat{\omega}_{\mu J} =& \ \hat{\omega}_{J\mu}' \\
\hat{\Omega}_{JJ} =& \ \Omega_{JJ}^{\frac{1}{2}\prime}\dot{\Omega}_{JJ}\Omega_{JJ}^{\frac{1}{2}} \\
\hat{\Omega}_{JJ}^{-1} =& \ \Omega_{JJ}^{-\frac{1}{2}}\dot{\Omega}_{JJ}^{-1}\Omega_{JJ}^{-\frac{1}{2}\prime} \\
-\hat{\Omega}_{JJ}^{-1}\hat{\omega}_{J\mu} =& \ -\Omega_{JJ}^{-\frac{1}{2}}\dot{\Omega}_{JJ}^{-1}\Omega_{JJ}^{-\frac{1}{2}\prime}\left(\Omega_{JJ}^{\frac{1}{2}\prime}\dot{\omega}_{J\mu}\omega_{\mu\mu.J}^{\frac{1}{2}} + \Omega_{JJ}^{\frac{1}{2}\prime}\dot{\Omega}_{JJ}\Omega_{JJ}^{-\frac{1}{2}}\omega_{J\mu}\right) \\
=& \ -\Omega_{JJ}^{-\frac{1}{2}}\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}\omega_{\mu\mu.J}^{\frac{1}{2}} - \Omega_{JJ}^{-1}\omega_{J\mu} \\
\hat{\omega}_{\mu\mu.J} =& \ \omega_{\mu\mu.J}\dot{\omega}_{\mu\mu} + \omega_{\mu J}\Omega_{JJ}^{-\frac{1}{2}\prime}\dot{\omega}_{J\mu}\omega_{\mu\mu.J}^{\frac{1}{2}} + \omega_{\mu\mu.J}^{\frac{1}{2}}\dot{\omega}_{\mu J}\Omega_{JJ}^{-\frac{1}{2}}\omega_{J\mu} + \omega_{\mu J}\Omega_{JJ}^{-\frac{1}{2}\prime}\dot{\Omega}_{JJ}\Omega_{JJ}^{-\frac{1}{2}}\omega_{J\mu} - \\
& \left(\omega_{\mu\mu.J}^{\frac{1}{2}}\dot{\omega}_{\mu J}\Omega_{JJ}^{\frac{1}{2}} + \omega_{\mu J}\Omega_{JJ}^{-\frac{1}{2}\prime}\dot{\Omega}_{JJ}\Omega_{JJ}^{\frac{1}{2}}\right)\left(\Omega_{JJ}^{-\frac{1}{2}}\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}\omega_{\mu\mu.J}^{\frac{1}{2}} + \Omega_{JJ}^{-1}\omega_{J\mu}\right) \\
=& \ \omega_{\mu\mu.J}\dot{\omega}_{\mu\mu} + \omega_{\mu J}\Omega_{JJ}^{-\frac{1}{2}\prime}\dot{\omega}_{J\mu}\omega_{\mu\mu.J}^{\frac{1}{2}} + \omega_{\mu\mu.J}^{\frac{1}{2}}\dot{\omega}_{\mu J}\Omega_{JJ}^{-\frac{1}{2}}\omega_{J\mu} + \omega_{\mu J}\Omega_{JJ}^{-\frac{1}{2}\prime}\dot{\Omega}_{JJ}\Omega_{JJ}^{-\frac{1}{2}}\omega_{J\mu} - \\
& \omega_{\mu\mu.J}^{\frac{1}{2}}\dot{\omega}_{\mu J}\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}\omega_{\mu\mu.J}^{\frac{1}{2}} - \omega_{\mu\mu.J}^{\frac{1}{2}}\dot{\omega}_{\mu J}\Omega_{JJ}^{-\frac{1}{2}\prime}\omega_{J\mu} - \omega_{\mu J}\Omega_{JJ}^{-\frac{1}{2}\prime}\dot{\omega}_{J\mu}\omega_{\mu\mu.J}^{\frac{1}{2}} - \\
& \omega_{\mu J}\Omega_{JJ}^{-\frac{1}{2}\prime}\dot{\Omega}_{JJ}\Omega_{JJ}^{-\frac{1}{2}}\omega_{J\mu} \\
=& \ \omega_{\mu\mu.J}\left(\dot{\omega}_{\mu\mu} - \dot{\omega}_{\mu J}\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}\right) = \omega_{\mu\mu.J}\dot{\omega}_{\mu\mu.J}
\end{aligned}$$

For $\dot{u}_i$, $\dot{V}_i$ and Z_i, resp. one, m and k_f dimensional i.i.d. random variables with mean zero and identity covariance matrices, we then have:

$$\begin{aligned}
\left(\hat{\mu}_f(0) \vdots \hat{J}(0)\right) =& \ \left(\mu_f(0) \vdots J(0)\right) + \Sigma^{\frac{1}{2}}\left(\tfrac{1}{N}\textstyle\sum_{i=1}^N \dot{Z}_i\binom{\dot{u}_i}{\dot{V}_i}'\right)\Omega^{\frac{1}{2}} \\
\sqrt{N}\left(\hat{\mu}_f(0) \vdots \hat{J}(0)\right) =& \ \Sigma^{\frac{1}{2}}\left[(a \vdots C) + \left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N \dot{Z}_i\binom{\dot{u}_i}{\dot{V}_i}'\right)\right]\Omega^{\frac{1}{2}}
\end{aligned}$$

$$\begin{aligned}
\hat{a} =& \ \sqrt{N}\hat{\Sigma}^{-\frac{1}{2}}\left(\hat{\mu}_f(0) - \hat{J}(0)\hat{\Omega}_{JJ}^{-1}\hat{\omega}_{J\mu}\right)\hat{\omega}_{\mu\mu.J}^{-\frac{1}{2}} \\
=& \ \dot{\Sigma}^{-\frac{1}{2}}\left((a \vdots C) + \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N \dot{Z}_i\binom{\dot{u}_i}{\dot{V}_i}'\right)\begin{pmatrix} \omega_{\mu\mu.J}^{\frac{1}{2}} & 0 \\ \Omega_{JJ}^{-\frac{1}{2}}\omega_{J\mu} & \Omega_{JJ}^{\frac{1}{2}} \end{pmatrix}\begin{pmatrix} 1 \\ -\Omega_{JJ}^{-\frac{1}{2}}\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}\omega_{\mu\mu.J}^{\frac{1}{2}} - \Omega_{JJ}^{-1}\omega_{J\mu} \end{pmatrix}\omega_{\mu\mu.J}^{-\frac{1}{2}} \\
& \left(\dot{\omega}_{\mu\mu} - \dot{\omega}_{\mu J}\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}\right)^{-\frac{1}{2}} \\
=& \ \dot{\Sigma}^{-\frac{1}{2}}\left((a \vdots C) + \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N \dot{Z}_i\binom{\dot{u}_i}{\dot{V}_i}'\right)\begin{pmatrix} 1 \\ -\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu} \end{pmatrix}\left(\dot{\omega}_{\mu\mu} - \dot{\omega}_{\mu J}\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}\right)^{-\frac{1}{2}} \\
=& \ \dot{\Sigma}^{-\frac{1}{2}}\left((a - C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}) + \left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N \dot{Z}_i(\dot{u}_i - \dot{V}_i'\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})\right)\right)\left(\dot{\omega}_{\mu\mu} - \dot{\omega}_{\mu J}\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}\right)^{-\frac{1}{2}}
\end{aligned}$$

$$
\begin{aligned}
\hat{C} = & \sqrt{N}\hat{\Sigma}^{-\frac{1}{2}}\hat{J}(0)\hat{\Omega}_{JJ}^{-\frac{1}{2}} \\
= & \dot{\Sigma}^{-\frac{1}{2}}\left(C + \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\dot{Z}_i\dot{V}_i'\right)\Omega_{JJ}^{\frac{1}{2}}\Omega_{JJ}^{-\frac{1}{2}}\dot{\Omega}_{JJ}^{-\frac{1}{2}} \\
= & \dot{\Sigma}^{-\frac{1}{2}}\left(C\dot{\Omega}_{JJ}^{-\frac{1}{2}} + \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\dot{Z}_i\dot{V}_i'\dot{\Omega}_{JJ}^{-\frac{1}{2}}\right). \\
\sqrt{N}\hat{\Sigma}^{-\frac{1}{2}}\left(\hat{\mu}_f(0) \vdots \hat{J}(0)\right)\hat{\Omega}^{-\frac{1}{2}} = & \sqrt{N}\hat{\Sigma}^{-\frac{1}{2}}\left(\hat{\mu}_f(0) \vdots \hat{J}(0)\right)\begin{pmatrix} \hat{\omega}_{\mu\mu.J}^{-\frac{1}{2}} & 0 \\ -\hat{\Omega}_{JJ}^{-1}\hat{\omega}_{J\mu}\hat{\omega}_{\mu\mu.J}^{-\frac{1}{2}} & \hat{\Omega}_{JJ}^{-\frac{1}{2}} \end{pmatrix} \\
= & \dot{\Sigma}^{-\frac{1}{2}}\left[(a \vdots C) + \left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\dot{Z}_i\binom{\dot{u}_i}{\dot{V}_i}'\right)\right]\dot{\Omega}^{-\frac{1}{2}},
\end{aligned}
$$

with $\dot{\Omega}^{-\frac{1}{2}} = \begin{pmatrix} \dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}} & 0 \\ -\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}\dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}} & \dot{\Omega}_{JJ}^{-\frac{1}{2}} \end{pmatrix}$, so

$$
\begin{aligned}
\hat{a} = & \sqrt{N}\hat{\Sigma}^{-\frac{1}{2}}\left(\hat{\mu}_f(0) - \hat{J}(0)\hat{\Omega}_{JJ}^{-1}\hat{\omega}_{J\mu}\right)\hat{\omega}_{\mu\mu.J}^{-\frac{1}{2}} \\
= & \dot{\Sigma}^{-\frac{1}{2}\prime}\left((a - C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}) + \left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\dot{Z}_i(\dot{u}_i - \dot{V}_i'\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})\right)\right)\dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}} \\
\hat{C} = & \sqrt{N}\hat{\Sigma}^{-\frac{1}{2}}\hat{J}(0)\hat{\Omega}_{JJ}^{-\frac{1}{2}} \\
= & \dot{\Sigma}^{-\frac{1}{2}\prime}\left(C\dot{\Omega}_{JJ}^{-\frac{1}{2}} + \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\dot{Z}_i\dot{V}_i'\dot{\Omega}_{JJ}^{-\frac{1}{2}}\right).
\end{aligned}
$$

The convergence of $\dot{\omega}_{\mu\mu}$, $\dot{\omega}_{J\mu}$, $\dot{\Omega}_{JJ}$ and $\dot{\Sigma}$ implies the intermediate results:

$$
\begin{aligned}
\dot{\Omega}_{JJ}^{-1} &= I_m - \tfrac{1}{\sqrt{N}}\Psi_{JJ} + O_p(N^{-1}) \\
\dot{\Omega}_{JJ}^{-\frac{1}{2}} &= I_m - \tfrac{1}{2\sqrt{N}}\Psi_{JJ} + O_p(N^{-1}) \\
\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu} &= \tfrac{1}{\sqrt{N}}\psi_{J\mu} + O_p(N^{-1}) \\
\dot{\omega}_{\mu J}\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu} &= \tfrac{1}{N}\psi'_{J\mu}\psi_{J\mu} + O_p(N^{-1\frac{1}{2}}) \\
\dot{\omega}_{\mu\mu.J} = \dot{\omega}_{\mu\mu} - \dot{\omega}_{\mu J}\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu} &= 1 + \tfrac{1}{\sqrt{N}}\psi_{\mu\mu} - \tfrac{1}{N}\psi'_{J\mu}\psi_{J\mu} + O_p(N^{-1\frac{1}{2}}) \\
\dot{\omega}_{\mu\mu.J}^{-1} &= 1 - \tfrac{1}{\sqrt{N}}\psi_{\mu\mu} + O_p(N^{-1}) \\
\dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}} &= 1 - \tfrac{1}{2\sqrt{N}}\psi_{\mu\mu} + O_p(N^{-1}) \\
\dot{\Sigma}^{-1} &= I_{k_f} - \tfrac{1}{\sqrt{N}}\Psi_{\Sigma} + O_p(N^{-1}) \\
a - C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu} &= a - \tfrac{1}{\sqrt{N}}C\psi_{J\mu} + O_p(N^{-1}) \\
\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\dot{Z}_i(\dot{u}_i - \dot{V}_i'\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}) &= \psi_{\mu.J}^{*} - \tfrac{1}{\sqrt{N}}\psi_J^{*}\psi_{J\mu} + O_p(N^{-1})
\end{aligned}
$$

for ψ_{Σ} =vech(Ψ), which we use to establish the higher order behavior of $\hat{C}'\hat{C}$, $\hat{a}'\hat{a}$ and

$\hat{C}'\hat{a}$:

$$\begin{aligned}
\hat{a}'\hat{a} &= \dot{\omega}_{\mu\mu.J}^{-1}\left((a - C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}) + \left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N \dot{Z}_i(\dot{u}_i - \dot{V}_i'\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})\right)\right)\dot{\Sigma}^{-1} \\
&\quad \left((a - C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}) + \left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N \dot{Z}_i(\dot{u}_i - \dot{V}_i'\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})\right)\right) \\
&= \dot{\omega}_{\mu\mu.J}^{-1}(a - C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})'\dot{\Sigma}^{-1}(a - C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}) + \\
&\quad 2\dot{\omega}_{\mu\mu.J}^{-1}(a - C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})'\dot{\Sigma}^{-1}\left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N \dot{Z}_i(\dot{u}_i - \dot{V}_i'\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})\right) + \\
&\quad \dot{\omega}_{\mu\mu.J}^{-1}\left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N \dot{Z}_i(\dot{u}_i - \dot{V}_i'\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})\right)\dot{\Sigma}^{-1}\left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N \dot{Z}_i(\dot{u}_i - \dot{V}_i'\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})\right)
\end{aligned}$$

$$\begin{aligned}
&\dot{\omega}_{\mu\mu.J}^{-1}(a - C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})'\dot{\Sigma}^{-1}(a - C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}) \\
= &\left(1 - \tfrac{1}{\sqrt{N}}\psi_{\mu\mu}\right)\left(a - \tfrac{1}{\sqrt{N}}C\psi_{J\mu}\right)'\left(I_{k_f} - \tfrac{1}{\sqrt{N}}\Psi_\Sigma\right)\left(a - \tfrac{1}{\sqrt{N}}C\psi_{J\mu}\right) + O_p(N^{-1}) \\
= &\ a'a - \tfrac{1}{\sqrt{N}}\left(a'\Psi_\Sigma a + a'a\psi_{\mu\mu} + 2a'C\psi_{J\mu}\right) + O_p(N^{-1})
\end{aligned}$$

$$\begin{aligned}
&\dot{\omega}_{\mu\mu.J}^{-1}(a - C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})'\dot{\Sigma}^{-1}\left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N \dot{Z}_i(\dot{u}_i - \dot{V}_i'\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})\right) \\
= &\left(1 - \tfrac{1}{\sqrt{N}}\psi_{\mu\mu}\right)\left(a - \tfrac{1}{\sqrt{N}}C\psi_{J\mu}\right)'\left(I_{k_f} - \tfrac{1}{\sqrt{N}}\Psi_\Sigma\right)\left(\psi_{\mu.J}^* - \tfrac{1}{\sqrt{N}}\psi_J^*\psi_{J\mu}\right) + O_p(N^{-1}) \\
= &\ a'\psi_{\mu.J}^* - \tfrac{1}{\sqrt{N}}\left(\psi_{J\mu}'C'\psi_{\mu.J}^* + a'\psi_J^*\psi_{J\mu} + \psi_{\mu\mu}a'\psi_{\mu.J}^* + a'\Psi_\Sigma\psi_{\mu.J}^*\right) + O_p(N^{-1})
\end{aligned}$$

$$\begin{aligned}
&\dot{\omega}_{\mu\mu.J}^{-1}\left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N \dot{Z}_i(\dot{u}_i - \dot{V}_i'\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})\right)\dot{\Sigma}^{-1}\left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N \dot{Z}_i(\dot{u}_i - \dot{V}_i'\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})\right) \\
= &\left(1 - \tfrac{1}{\sqrt{N}}\psi_{\mu\mu}\right)\left(\psi_{\mu.J}^* - \tfrac{1}{\sqrt{N}}\psi_J^*\psi_{J\mu}\right)'\left(I_{k_f} - \tfrac{1}{\sqrt{N}}\Psi_\Sigma\right)\left(\psi_{\mu.J}^* - \tfrac{1}{\sqrt{N}}\psi_J^*\psi_{J\mu}\right) + O_p(N^{-1}) \\
= &\ \psi_{\mu.J}^{*\prime}\psi_{\mu.J}^* - \tfrac{1}{\sqrt{N}}\left(\psi_{\mu.J}^{*\prime}\Psi_\Sigma\psi_{\mu.J}^* + \psi_{\mu.J}^{*\prime}\psi_{\mu.J}^*\psi_{\mu\mu} + 2\psi_{\mu.J}^{*\prime}\psi_J^*\psi_{J\mu}\right) + O_p(N^{-1})
\end{aligned}$$

Combining the above, we obtain:

$$\begin{aligned}
\hat{a}'\hat{a} &= a'a + 2a'\psi_{\mu.J}^* + \psi_{\mu.J}^{*\prime}\psi_{\mu.J}^* - \tfrac{1}{\sqrt{N}}\left(a'\Psi_\Sigma a + a'a\psi_{\mu\mu} + 2a'C\psi_{J\mu}\right) - \\
&\quad \tfrac{2}{\sqrt{N}}\left(\psi_{J\mu}'C'\psi_{\mu.J}^* + a'\psi_J^*\psi_{J\mu} + \psi_{\mu\mu}a'\psi_{\mu.J}^* + a'\Psi_\Sigma\psi_{\mu.J}^*\right) - \\
&\quad \tfrac{1}{\sqrt{N}}\left(\psi_{\mu.J}^{*\prime}\Psi_\Sigma\psi_{\mu.J}^* + \psi_{\mu.J}^{*\prime}\psi_{\mu.J}^*\psi_{\mu\mu} + 2\psi_{\mu.J}^{*\prime}\psi_J^*\psi_{J\mu}\right) + O_p(N^{-1}).
\end{aligned}$$

Similarly, we derive the higher order behavior of $\hat{C}'\hat{C}$, and $\hat{C}'\hat{a}$, as follows.

$$
\begin{aligned}
\hat{C}'\hat{C} = \;& \dot{\Omega}_{JJ}^{-1}\left(C + \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\dot{Z}_i\dot{V}_i'\right)'\dot{\Sigma}^{-1}\left(C + \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\dot{Z}_i\dot{V}_i'\right) \\
= \;& \left(I_m - \tfrac{1}{\sqrt{N}}\Psi_{JJ}\right)(C+\psi_J^*)'\left(I_{k_f} - \tfrac{1}{\sqrt{N}}\Psi_\Sigma\right)(C+\psi_J^*) + O_p(N^{-1}) \\
= \;& C'C + 2C'\psi_J^* + \psi_J^{*\prime}\psi_J^* - \tfrac{1}{\sqrt{N}}\Psi_{JJ}\left(C'C + 2C'\psi_J^* + \psi_J^{*\prime}\psi_J^*\right) - \\
& \tfrac{1}{\sqrt{N}}\left(C'\Psi_\Sigma C + 2C\Psi_\Sigma\psi_J^* + \psi_J^{*\prime}\Psi_\Sigma\psi_J^*\right) + O_p(N^{-1})
\end{aligned}
$$

$$
\begin{aligned}
\hat{a}'\hat{C} \;=\;& \dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}}\left((a - C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}) + \left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\dot{Z}_i(\dot{u}_i - \dot{V}_i'\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})\right)\right)\dot{\Sigma}^{-1} \\
& \left(C + \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\dot{Z}_i\dot{V}_i'\right)\dot{\Omega}_{JJ}^{-\frac{1}{2}} \\
=\;& \dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}}(a - C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})'\dot{\Sigma}^{-1}\left(C + \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\dot{Z}_i\dot{V}_i'\right)\dot{\Omega}_{JJ}^{-\frac{1}{2}} + \\
& \dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}}\left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\dot{Z}_i(\dot{u}_i - \dot{V}_i'\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})\right)\dot{\Sigma}^{-1}\left(C + \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\dot{Z}_i\dot{V}_i'\right)\dot{\Omega}_{JJ}^{-\frac{1}{2}} \\
=\;& \left(1 - \tfrac{1}{2\sqrt{N}}\psi_{\mu\mu}\right)\left(a - \tfrac{1}{\sqrt{N}}C\psi_{J\mu}\right)'\left(I_{k_f} - \tfrac{1}{\sqrt{N}}\Psi_\Sigma\right)(C+\psi_J^*)\left(1 - \tfrac{1}{2\sqrt{N}}\Psi_{JJ}\right) + \\
& \left(1 - \tfrac{1}{2\sqrt{N}}\psi_{\mu\mu}\right)\left(\psi_{\mu.J}^* - \tfrac{1}{\sqrt{N}}\psi_J^*\psi_{J\mu}\right)'\left(I_{k_f} - \tfrac{1}{\sqrt{N}}\Psi_\Sigma\right)(C+\psi_J^*)\left(1 - \tfrac{1}{2\sqrt{N}}\Psi_{JJ}\right) + O_p(N^{-1}) \\
=\;& a'C + a'\psi_J^* + \psi_{\mu.J}^{*\prime}C + \psi_{\mu.J}^{*\prime}\psi_J^* - \tfrac{1}{2\sqrt{N}}\left(\psi_{\mu\mu} + \Psi_{JJ}\right)(a'C + a'\psi_J^*) - \\
& \tfrac{1}{\sqrt{N}}\left(a'\Psi_\Sigma C + a'\Psi_\Sigma\psi_J^* + \psi_{J\mu}C'(C+\psi_J^*)\right) - \tfrac{1}{2\sqrt{N}}\left(\psi_{\mu\mu} + \Psi_{JJ}\right)\left(\psi_{\mu.J}^{*\prime}C + \psi_{\mu.J}^{*\prime}\psi_J^*\right) - \\
& \tfrac{1}{\sqrt{N}}\left(\psi_{\mu.J}^{*\prime}\Psi_\Sigma C + \psi_{\mu.J}^{*\prime}\Psi_\Sigma\psi_J^* + \psi_{J\mu}\psi_J^{*\prime}(C+\psi_J^*)\right) + O_p(N^{-1})
\end{aligned}
$$

Under $\mathrm{H}^*_{\text{non-ind}} : C'C = a'a$ and $C'a = 0$:

$$
\begin{aligned}
& \hat{C}'\hat{C} - \hat{a}'\hat{a} \\
=\;& 2C'\psi_J^* + \psi_J^{*\prime}\psi_J^* - 2a'\psi_{\mu.J}^* - \psi_{\mu.J}^{*\prime}\psi_{\mu.J}^* - \tfrac{1}{\sqrt{N}}\Psi_{JJ}\left(C'C + 2C'\psi_J^* + \psi_J^{*\prime}\psi_J^*\right) - \\
& \tfrac{1}{\sqrt{N}}\left(C'\Psi_\Sigma C + 2C\Psi_\Sigma\psi_J^* + \psi_J^{*\prime}\Psi_\Sigma\psi_J^*\right) + \tfrac{1}{\sqrt{N}}\left(a'\Psi_\Sigma a + a'a\psi_{\mu\mu}\right) + \\
& \tfrac{2}{\sqrt{N}}\left(\psi_{J\mu}'C'\psi_{\mu.J}^* + a'\psi_J^*\psi_{J\mu} + \psi_{\mu\mu}a'\psi_{\mu.J}^* + a'\Psi_\Sigma\psi_{\mu.J}^*\right) + \\
& \tfrac{1}{\sqrt{N}}\left(\psi_{\mu.J}^{*\prime}\Psi_\Sigma\psi_{\mu.J}^* + \psi_{\mu.J}^{*\prime}\psi_{\mu.J}^*\psi_{\mu\mu} + 2\psi_{\mu.J}^{*\prime}\psi_J^*\psi_{J\mu}\right) + O_p(N^{-1})
\end{aligned}
$$

$$
\begin{aligned}
& \hat{a}'\hat{C} \\
=\;& a'\psi_J^* + \psi_{\mu.J}^{*\prime}C + \psi_{\mu.J}^{*\prime}\psi_J^* - \tfrac{1}{2\sqrt{N}}\left(\psi_{\mu\mu} + \Psi_{JJ}\right)a'\psi_J^* - \\
& \tfrac{1}{\sqrt{N}}\left(a'\Psi_\Sigma C + a'\Psi_\Sigma\psi_J^* + \psi_{J\mu}C'(C+\psi_J^*)\right) - \tfrac{1}{2\sqrt{N}}\left(\psi_{\mu\mu} + \Psi_{JJ}\right)\left(\psi_{\mu.J}^{*\prime}C + \psi_{\mu.J}^{*\prime}\psi_J^*\right) - \\
& \tfrac{1}{\sqrt{N}}\left(\psi_{\mu.J}^{*\prime}\Psi_\Sigma C + \psi_{\mu.J}^{*\prime}\Psi_\Sigma\psi_J^* + \psi_{J\mu}\psi_J^{*\prime}(C+\psi_J^*)\right) + O_p(N^{-1})
\end{aligned}
$$

$$
\begin{aligned}
& \hat{C}'\hat{C} + \hat{a}'\hat{a} \\
= \; & C'C + a'a + 2C'\psi_J^* + \psi_J^{*\prime}\psi_J^* + 2a'\psi_{\mu.J}^* + \psi_{\mu.J}^{*\prime}\psi_{\mu.J}^* - \tfrac{1}{\sqrt{N}}\Psi_{JJ}\left(C'C + 2C'\psi_J^* + \psi_J^{*\prime}\psi_J^*\right) - \\
& \tfrac{1}{\sqrt{N}}\left(C'\Psi_\Sigma C + 2C\Psi_\Sigma\psi_J^* + \psi_J^{*\prime}\Psi_\Sigma\psi_J^*\right) - \tfrac{1}{\sqrt{N}}\left(a'\Psi_\Sigma a + a'a\psi_{\mu\mu}\right) - \\
& \tfrac{2}{\sqrt{N}}\left(\psi_{J\mu}'C'\psi_{\mu.J}^* + a'\psi_J^*\psi_{J\mu} + \psi_{\mu\mu}a'\psi_{\mu.J}^* + a'\Psi_\Sigma\psi_{\mu.J}^*\right) - \\
& \tfrac{1}{\sqrt{N}}\left(\psi_{\mu.J}^{*\prime}\Psi_\Sigma\psi_{\mu.J}^* + \psi_{\mu.J}^{*\prime}\psi_{\mu.J}^*\psi_{\mu\mu} + 2\psi_{\mu.J}^{*\prime}\psi_J^*\psi_{J\mu}\right) + O_p(N^{-1})
\end{aligned}
$$

The above expressions are equivalent to those presented in Theorem 4, which are re-organized.

A6. Proof of Theorem 5. Using the specfication from (25)-(26), IS and MISS are defined by:

$$
\begin{aligned}
\text{IS} = \; & \text{smallest root of the characteristic polynomial: } \left|\tau I_m - \tfrac{1}{N}C'C\right| = 0 \\
\text{MISS} = \; & \text{smallest root of the characteristic polynomial: } \left|\lambda I_{m+1} - \tfrac{1}{N}(a \vdots C)'(a \vdots C)\right| = 0.
\end{aligned}
$$

To specify the hypothesis of no-identification, $\text{H}_{\text{non-ind}}$: IS = MISS, more explicitly as a function of a and C, we first use the SVD of C (49) to specify:

$$
\begin{aligned}
& (a \vdots C)'(a \vdots C) \\
= \; & \left[U_C'(a \vdots C)\begin{pmatrix} 1 & 0 \\ 0 & V_C \end{pmatrix}'\begin{pmatrix} 1 & 0 \\ 0 & V_C \end{pmatrix}\right]' U_C'(a \vdots C)\begin{pmatrix} 1 & 0 \\ 0 & V_C \end{pmatrix}'\begin{pmatrix} 1 & 0 \\ 0 & V_C \end{pmatrix} \\
= \; & \begin{pmatrix} 1 & 0 \\ 0 & V_C \end{pmatrix}'\begin{pmatrix} a^* & S_{C,1} & 0 \\ U_{C,m}'a & 0 & s_{C,mm} \\ b & 0 & 0 \end{pmatrix}'\begin{pmatrix} a^* & S_{C,1} & 0 \\ U_{C,m}'a & 0 & s_{C,mm} \\ b & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & V_C \end{pmatrix}
\end{aligned}
$$

for $a^* = U_{C,1}'a$, $b = U_{C,2}'a$, which used that U_C and $\begin{pmatrix} 1 & 0 \\ 0 & V_C \end{pmatrix}$ are orthonormal, and $S_{C,1}$ is the $(m-1)$ by $(m-1)$ diagonal matrix with $s_{C,11}$, ..., $s_{C,(m-1)(m-1)}$ on the diagonal.

Since $\begin{pmatrix} 1 & 0 \\ 0 & V_C \end{pmatrix}$ is also an orthonormal matrix, MISS equals the smallest eigenvalue of

$$\begin{pmatrix} a^* & S_{C,1} & 0 \\ U'_{C,m}a & 0 & s_{C,mm} \\ b & 0 & 0 \end{pmatrix}' \begin{pmatrix} a^* & S_{C,1} & 0 \\ U'_{C,m}a & 0 & s_{C,mm} \\ b & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} a^{*\prime}a^* + (U'_{C,m}a)^2 + b'b & a^{*\prime}S_{C,1} & (U'_{C,m}a)s_{C.mm} \\ S'_{C,1}a^* & S'_{C,1}S_{C,1} & 0 \\ (U'_{C,m}a)s_{C.mm} & 0 & (s_{C,mm})^2 \end{pmatrix},$$

which we can similarly specify as:

$$R \begin{pmatrix} a_1^{*2} + s_{C,11}^2 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & \vdots & \vdots \\ 0 & 0 & a_{m-1}^{*2} + s_{C,(m-1)(m-1)}^2 & 0 & 0 \\ 0 & \cdots & 0 & b'b & 0 \\ 0 & \cdots & 0 & 0 & (U'_{C,m}a)^2 + (s_{C,mm})^2 \end{pmatrix} R',$$

for $R = (R_1 \vdots R_m \vdots R_{m+1})$, $R_m = e_{1,m+1}$, $e_{i,m+1}$ the i-th $(m+1)$-dimensional unity vector (or i-th column of I_{m+1}), $a^* = (a_1^*, \ldots, a_{m-1}^*)'$,

$$R_1 = \begin{pmatrix} \frac{a_1^*}{\sqrt{a_1^{*2}+s_{C,11}^2}} & \cdots & \frac{a_{m-1}^*}{\sqrt{a_{m-1}^{*2}+s_{C,(m-1)(m-1)}^2}} \\ \frac{s_{C,11}}{\sqrt{a_1^{*2}+s_{C,11}^2}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{s_{C,(m-1)(m-1)}}{\sqrt{a_{m-1}^{*2}+s_{C,(m-1)(m-1)}^2}} \\ 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{pmatrix} : (m+1) \times (m-1),$$

$$R_{m+1} = \frac{s_{C,mm}}{\sqrt{(U'_{C,m}a)^2+s_{C,mm}^2}} e_{m+1,m+1} + \frac{(U'_{C,m}a)}{\sqrt{(U'_{C,m}a)^2+s_{C,mm}^2}} e_{1,m+1}.$$

When $U'_{C,m}a = 0$, $b'b$ and $(s_{C,mm})^2$ are two of the eigenvalues, so a necessary and sufficient condition for IS = MISS is: $U'_{C,m}a = 0$ and $b'b \geq (s_{C,mm})^2$.

A7. Proof of Theorem 6. We use that

$$(\hat{a} \vdots \hat{C}) = \sqrt{N}\hat{\Sigma}^{-\frac{1}{2}}\left(\hat{\mu}_f(0) \vdots \hat{J}(0)\right)\hat{\Omega}^{-\frac{1}{2}},$$

so $\hat{C} = \sqrt{N}\hat{J}(0)\hat{\Omega}_{JJ}^{-\frac{1}{2}}$, $\hat{a} = \sqrt{N}\hat{\Sigma}^{-\frac{1}{2}}\hat{\mu}_f(0) - \hat{C}\hat{\Omega}_{JJ}^{-\frac{1}{2}}\hat{\omega}_{Ju}$. A SVD of $\hat{C}$ results in

$$\hat{C} = \hat{U}_C\hat{S}_C\hat{V}_C',$$

with $\hat{U}_C = (\hat{U}_{C,1} \vdots \hat{U}_{C,m} \vdots \hat{U}_{C,2})$, $\hat{U}_{C,1} : k_f \times (m-1)$, $\hat{U}_{C,m} : k_f \times 1$, $\hat{U}_{C,2} : k_f \times (k_f - m)$, and $\hat{V}_C$ orthonormal $k_f \times k_f$ and $m \times m$ dimensional matrices, and $\hat{S}_C = \begin{pmatrix} \hat{S}_{C,1} & 0 \\ 0 & \hat{s}_{C,mm} \\ 0 & 0 \end{pmatrix}$, $\hat{S}_{C,1} : (m-1) \times (m-1)$, $\hat{s}_{C,mm} : 1 \times 1$, a $k_f \times m$ dimensional matrix with the singular values in decreasing order on the main diagonal so $\widehat{\text{IS}} = \hat{s}_{C,mm}^2$.

$\widehat{\text{MISS}}$ equals the smallest root of the characteristic polynomial:

$$\left|\lambda I_m - (\hat{a} \vdots \hat{C})'(\hat{a} \vdots \hat{C})\right| = 0.$$

We specify $(\hat{a} \vdots \hat{C})'(\hat{a} \vdots \hat{C})$ using the SVD of $\hat{C}$ as

$$\begin{aligned}
& (\hat{a} \vdots \hat{C})'(\hat{a} \vdots \hat{C}) \\
= \; & (\hat{a} \vdots \hat{C})'\hat{U}_C\hat{U}_C'(\hat{a} \vdots \hat{C}) \\
= \; & \begin{pmatrix} \hat{U}_{C,1}'\hat{a} & \hat{U}_{C,1}'\hat{C} \\ \hat{U}_{C,m}'\hat{a} & \hat{U}_{C,m}'\hat{C} \\ \hat{U}_{C,2}'\hat{a} & \hat{U}_{C,2}\hat{C} \end{pmatrix}' \begin{pmatrix} \hat{U}_{C,1}'\hat{a} & \hat{U}_{C,1}'\hat{C} \\ \hat{U}_{C,m}'\hat{a} & \hat{U}_{C,m}'\hat{C} \\ \hat{U}_{C,2}'\hat{a} & \hat{U}_{C,2}\hat{C} \end{pmatrix} \\
= \; & \begin{pmatrix} 1 & 0 \\ 0 & \hat{V}_C' \end{pmatrix}' \begin{pmatrix} \hat{U}_{C,1}'\hat{a} & \hat{S}_{C,1} & 0 \\ \hat{U}_{C,m}'\hat{a} & 0 & \hat{s}_{C,mm} \\ \hat{U}_{C,2}'\hat{a} & 0 & 0 \end{pmatrix}' \begin{pmatrix} \hat{U}_{C,1}'\hat{a} & \hat{S}_{C,1} & 0 \\ \hat{U}_{C,m}'\hat{a} & 0 & \hat{s}_{C,mm} \\ \hat{U}_{C,2}'\hat{a} & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \hat{V}_C' \end{pmatrix} \\
= \; & \begin{pmatrix} 1 & 0 \\ 0 & \hat{V}_C' \end{pmatrix}' \begin{pmatrix} \hat{a}'\hat{U}_{C,1}\hat{U}_{C,1}'\hat{a} + (\hat{a}'\hat{U}_{C,m})^2 + \hat{a}'\hat{U}_{C,2}\hat{U}_{C,2}'\hat{a} & \hat{a}'\hat{U}_{C,1}\hat{S}_{C,1} & \hat{a}'\hat{U}_{C,m}\hat{s}_{C,mm} \\ \hat{S}_{C,1}'\hat{U}_{C,1}'\hat{a} & \hat{S}_{C,1}'\hat{S}_{C,1} & 0 \\ \hat{s}_{C,mm}\hat{U}_{C,m}'\hat{a} & 0 & \hat{s}_{C,mm}^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \hat{V}_C' \end{pmatrix}.
\end{aligned}$$

Because $\begin{pmatrix} 1 & 0 \\ 0 & \hat{V}_C' \end{pmatrix}$ is an orthonormal matrix, the characteristic roots are identical to the eigenvalues of:

$$\begin{pmatrix} \hat{a}^{*\prime}\hat{a}^* + (\hat{a}'\hat{U}_{C,m})^2 + \hat{b}^{*\prime}\hat{b}^* & \hat{a}^{*\prime}\hat{S}_{C,1} & \hat{a}'\hat{U}_{C,m}\hat{s}_{C,mm} \\ \hat{S}_{C,1}'\hat{a}^* & \hat{S}_{C,1}'\hat{S}_{C,1} & 0 \\ \hat{s}_{C,mm}\hat{U}_{C,m}'\hat{a} & 0 & \hat{s}_{C,mm}^2 \end{pmatrix}$$

$$= \hat{R}\begin{pmatrix} \hat{a}_1^{*2} + \hat{s}_{C,11}^2 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & \vdots & \vdots \\ 0 & 0 & \hat{a}_{m-1}^{*2} + \hat{s}_{C,(m-1)(m-1)}^2 & 0 & 0 \\ 0 & \cdots & 0 & \hat{b}^{*\prime}\hat{b}^* & 0 \\ 0 & \cdots & 0 & 0 & (\hat{a}'\hat{U}_{C,m})^2 + (\hat{s}_{C,mm}^2)^2 \end{pmatrix}\hat{R}',$$

for $\hat{U}_{C,1}'\hat{a} = (\hat{a}_1^* \dots \hat{a}_{m-1}^*)' = \hat{a}^*$, $\hat{U}_{C,2}'\hat{a} = \hat{b}^*$, $\hat{R} = (\hat{R}_1 \vdots \hat{R}_m \vdots \hat{R}_{m+1})$, $\hat{R}_m = e_{1,m+1}$, with $e_{i,m+1}$ the i-th $(m+1)$-dimensional unity vector (or i-th column of I_{m+1}),

$$\hat{R}_1 = \begin{pmatrix} \frac{\hat{a}_1^*}{\sqrt{\hat{a}_1^{*2}+\hat{s}_{C,11}^2}} & \cdots & \frac{\hat{a}_{m-1}^*}{\sqrt{\hat{a}_{m-1}^{*2}+\hat{s}_{C,(m-1)(m-1)}^2}} \\ \frac{\hat{s}_{C,1,11}}{\sqrt{\hat{a}_1^{*2}+\hat{s}_{C,11}^2}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{\hat{s}_{C,(m-1)(m-1)}}{\sqrt{\hat{a}_{m-1}^{*2}+\hat{s}_{C,(m-1)(m-1)}^2}} \\ 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{pmatrix} : (m+1) \times (m-1),$$

$$R_{m+1} = \frac{\hat{s}_{C,mm}}{\sqrt{(\hat{a}'\hat{U}_{C,m})^2+\hat{s}_{C,mm}^2}}e_{m+1,m+1} + \frac{(\hat{a}'\hat{U}_{C,m})}{\sqrt{(\hat{a}'\hat{U}_{C,m})^2+\hat{s}_{C,mm}^2}}e_{1,m+1}.$$

For relative large values of $\hat{s}_{C,ii}^2$ compared to $\hat{a}_i^{*2}$, $i = 1, \dots, m-1$, $\hat{R}_1$ becomes orthogonal to $\hat{R}_m$ and $\hat{R}_{m+1}$, so the two smallest eigenvalues then result from

$$\begin{pmatrix} \hat{a}'\hat{U}_{C,m}\hat{U}_{C,m}'\hat{a} + \hat{b}^{*\prime}\hat{b}^* & \hat{a}'\hat{U}_{C,m}\hat{s}_{C,mm} \\ \hat{s}_{C,mm}\hat{U}_{C,m}'\hat{a} & \hat{s}_{C,mm}^2 \end{pmatrix} =$$

$$\begin{pmatrix} \hat{a}'(\hat{U}_{C,m} \vdots \hat{U}_{C,2})(\hat{U}_{C,m} \vdots \hat{U}_{C,2})'\hat{a} & \hat{a}'\hat{U}_{C,m}\hat{s}_{C,mm} \\ \hat{s}_{C,mm}\hat{U}_{C,m}'\hat{a} & \hat{s}_{C,mm}^2 \end{pmatrix}.$$

We can next analytically solve for the smallest root which leads to the approximate expression of the LR statistic:

$$\begin{aligned}&\text{LR}(\alpha=0)\\ \approx\;& \tfrac{1}{2}\left[\hat{s}^2_{C,mm}-\hat{a}'M_{\hat{U}_{C,1}}\hat{a}+\sqrt{\left(\hat{s}^2_{C,mm}-\hat{a}'M_{\hat{U}_{C,1}}\hat{a}\right)^2+4\hat{s}^2_{C,mm}(\hat{U}'_{C,m}\hat{a})^2}\right].\end{aligned}$$

A8. Expansion for computing conditional critical value function for $m>1$. Using the proof of Theorem 4 and the SVD from Theorem 5:

$$\begin{aligned}
&\sqrt{N}\hat{\Sigma}^{-\frac{1}{2}}\left(\hat{\mu}_f(0)\vdots\hat{J}(0)\right)\hat{\Omega}^{-\frac{1}{2}}\\
=\;&\dot{\Sigma}^{-\frac{1}{2}}\left((a\vdots C)+\left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N\dot{Z}_i\binom{\dot{u}_i}{\dot{V}_i}'\right)\right)\dot{\Omega}^{-\frac{1}{2}}\\
=\;&\dot{\Sigma}^{-\frac{1}{2}}\left[(a\vdots U_CS_CV_C')+\left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N\dot{Z}_i\binom{\dot{u}_i}{\dot{V}_i}'\right)\right]\dot{\Omega}^{-\frac{1}{2}}\\
=\;&\dot{\Sigma}^{-\frac{1}{2}}U_C\left[(U_C'a\vdots S_C)+\left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N\ddot{Z}_i\binom{\dot{u}_i}{V_C'\dot{V}_i}'\right)\right]\begin{pmatrix}1&0\\0&V_C'\end{pmatrix}\dot{\Omega}^{-\frac{1}{2}}\\
=\;&\ddot{\Sigma}^{-\frac{1}{2}}\left[(\dot{a}\vdots S_C)+\left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N\ddot{Z}_i\binom{\dot{u}_i}{V_C'\dot{V}_i}'\right)\right]\begin{pmatrix}\dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}}&0\\-V_C'\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}\dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}}&V_C'\dot{\Omega}_{JJ}^{-\frac{1}{2}}\end{pmatrix}\\
=\;&\ddot{\Sigma}^{\frac{1}{2}}\left[(\dot{a}\vdots S_C)+\left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N\ddot{Z}_i\binom{\dot{u}_i}{V_C'\dot{V}_i}'\right)\right]\begin{pmatrix}\dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}}&0\\-(V_C'\dot{\Omega}_{JJ}V_C)^{-1}V_C\dot{\omega}_{J\mu}\dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}}&(V_C\dot{\Omega}_{JJ}V_C')^{-\frac{1}{2}}\end{pmatrix}\\
=\;&\ddot{\Sigma}^{\frac{1}{2}}\left[(\dot{a}\vdots S_C)+\left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N\ddot{Z}_i\binom{\dot{u}_i}{\ddot{V}_i}'\right)\right]\begin{pmatrix}\dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}}&0\\-\ddot{\Omega}_{JJ}^{-1}\ddot{\omega}_{J\mu}\dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}}&\ddot{\Omega}_{JJ}^{-\frac{1}{2}}\end{pmatrix}\\
=\;&\ddot{\Sigma}^{\frac{1}{2}}\Big[(\dot{a}-S_C\ddot{\Omega}_{JJ}^{-1}\ddot{\omega}_{J\mu})\dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}}+\left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N\ddot{Z}_i(\dot{u}_i-\ddot{V}_i\ddot{\Omega}_{JJ}^{-1}\ddot{\omega}_{J\mu})\dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}}\right)\\
&S_C\ddot{\Omega}_{JJ}^{-\frac{1}{2}}+\left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N\ddot{Z}_i\ddot{V}_i'\ddot{\Omega}_{JJ}^{-\frac{1}{2}}\right)\Big],
\end{aligned}$$

for $\dot{a}=U_C'a$, $\ddot{Z}_i=U_C'\dot{Z}_i$, $\ddot{\Sigma}=U_C'\dot{\Sigma}U_C=\frac{1}{N}\sum_{i=1}^N\ddot{Z}_i\ddot{Z}_i'$, $\ddot{\Sigma}^{-\frac{1}{2}}=\dot{\Sigma}^{-\frac{1}{2}}U_C$, $\ddot{V}_i=V_C'\dot{V}_i$, $\ddot{\Omega}_{JJ}=V_C'\dot{\Omega}_{JJ}V_C$, so $\ddot{\Omega}_{JJ}^{-1}=(V_C'\dot{\Omega}_{JJ}V_C)^{-1}=V_C^{-1}\dot{\Omega}_{JJ}V_C'^{-1}=V_C'\dot{\Omega}_{JJ}V_C$ because V_C is orthonormal, $V_C^{-1}=V_C'$, and $\ddot{\omega}_{J\mu}=V_C'\dot{\omega}_{J\mu}$.

A9. Additional empirical applications.

Table A1: LR test of no-identification for Jagannathan and Wang (1996)

Significance at 1%,***; 5%,**; 10%,*.

	$\lambda_0 \neq 0$	$\lambda_0 = 0$
$\widehat{\text{IS}}$	103.54	103.56
$\widehat{\text{MISS}}$	75.07	86.46
$\text{LR}(\alpha = 0) = \widehat{\text{IS}} - \widehat{\text{MISS}}$	28.47	17.10
Conditioning statistic: $\widehat{\text{MISS}} + \hat{\lambda}_{aC,m}$	251.45	272.29
95% conditional critical value	40.91	45.86

Table A2: LR test of no-identification for Lettau and Ludvigson (2001)

Significance at 1%,***; 5%,**; 10%,*.

	$\lambda_0 \neq 0$	$\lambda_0 = 0$
$\widehat{\text{IS}}$	31.75*	40.90**
$\widehat{\text{MISS}}$	31.11*	37.15**
$\text{LR}(\alpha = 0) = \widehat{\text{IS}} - \widehat{\text{MISS}}$	0.65	3.75
Conditioning statistic: $\widehat{\text{MISS}} + \hat{\lambda}_{aC,m}$	70.55	79.77
95% conditional critical value	23.56	25.80

Table A3: LR test of no-identification for Yogo (2006)

Significance at 1%,***; 5%,**; 10%,*.

	$\lambda_0 \neq 0$	$\lambda_0 = 0$
$\widehat{\text{IS}}$	17.34	19.60
$\widehat{\text{MISS}}$	17.14	19.42
$\text{LR}(\alpha = 0) = \widehat{\text{IS}} - \widehat{\text{MISS}}$	0.20	0.19
Conditioning statistic: $\widehat{\text{MISS}} + \hat{\lambda}_{aC,m}$	41.91	44.24
95% conditional critical value	14.91	15.28

Table A4: LR test of no-identification for Savov (2011)

Significance at 1%,***; 5%,**; 10%,*.

	$\lambda_0 \neq 0$	$\lambda_0 = 0$
$\widehat{\text{IS}}$	140.68***	296.78***
$\widehat{\text{MISS}}$	134.27***	268.60***
$\text{LR}(\alpha = 0) = \widehat{\text{IS}} - \widehat{\text{MISS}}$	6.41	28.18
Conditioning statistic: $\widehat{\text{MISS}} + \hat{\lambda}_{aC,m}$	812.86	984.98
95% conditional critical value	743.88	909.73

Table A5: LR test of no-identification for Adrian, Etula, and Muir (2014)

Significance at 1%,***; 5%,**; 10%,*.

	$\lambda_0 \neq 0$	$\lambda_0 = 0$
$\widehat{\text{IS}}$	31.97	42.03**
$\widehat{\text{MISS}}$	28.42	30.41
$\text{LR}(\alpha = 0) = \widehat{\text{IS}} - \widehat{\text{MISS}}$	3.56	11.62
Conditioning statistic: $\widehat{\text{MISS}} + \hat{\lambda}_{aC,m}$	118.57	157.56
95% conditional critical value	72.88	122.51

Table A6: LR test of no-identification for Kroencke (2017)

Significance at 1%,***; 5%,**; 10%,*.

	$\lambda_0 \neq 0$	$\lambda_0 = 0$
$\widehat{\text{IS}}$	78.47***	102.77***
$\widehat{\text{MISS}}$	59.84***	60.03***
$\text{LR}(\alpha = 0) = \widehat{\text{IS}} - \widehat{\text{MISS}}$	18.63	42.74
Conditioning statistic: $\widehat{\text{MISS}} + \hat{\lambda}_{aC,m}$	160.74	226.73
95% conditional critical value	111.56	178.03

Table A7: LR test of no-identification for He, Kelly, and Manela (2017)
Significance at 1%,***; 5%,**; 10%,*.

	$\lambda_0 \neq 0$	$\lambda_0 = 0$
$\widehat{\text{IS}}$	35.88**	59.74***
$\widehat{\text{MISS}}$	35.32**	44.44***
$\text{LR}(\alpha = 0) = \widehat{\text{IS}} - \widehat{\text{MISS}}$	0.57	15.29
Conditioning statistic: $\widehat{\text{MISS}} + \hat{\lambda}_{aC,m}$	114.40	167.37
95% conditional critical value	36.10	47.91