Testing for Identification in Potentially Misspecified Linear GMM*

Frank Kleibergen† Zhaoguo Zhan‡

August 10, 2026

Abstract

Identification in correctly specified linear generalized method of moments (GMM) requires the Jacobian to have full rank. The identification condition is more stringent in potentially misspecified, over-identified linear GMM. Many applications in, *e.g.*, macroeconomics and finance, concern over-identified misspecified models, where identification is tested using a Jacobian rank test. The appropriate identification condition is, however, that the difference between the population analogs of the traditional Jacobian rank statistic and the misspecification J-statistic exceeds zero. Their sample difference forms a (quasi) likelihood ratio statistic for testing no-identification. Under homoskedasticity, we establish a conditional critical value function for a size-correct no-identification likelihood ratio test. Extensions allow for covariance matrix estimators, multiple structural parameters, and general covariance structures. No-identification is then often not rejected in asset pricing applications.

Keywords: Misspecification, weak identification, conditional likelihood ratio test.

*We thank participants of the "Celebrating Don Andrews" conference at Yale, 34th EC^2 meeting, 2024 IAAE meeting, "Econometrics in Rio 2024" workshop, 2025 ESWC, 2026 AFES, Aarhus Workshop in Econometrics 1, and seminar participants at Brown, Cornell, LSE, Penn State, Penn, Princeton, Tilburg, Toulouse, UCL and Warwick for comments and discussion. The research of Frank Kleibergen has been funded partly by the NWO (Dutch Research Council) grant 401.21.EB.002: "Double robust inference for structural economic models (DRISEM)".

†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). Robust 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 identification strength 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. Misspecification is currently well recognized for, *e.g.*, over-identified instrumental variables (IV) estimation of treatment effects where local average treatment effects differ over the instruments because of heterogeneity; linear asset pricing models where observed risk factors are insufficient to span the cross-section of asset returns; and New-Keynesian Phillips curve models. See, *e.g.*, Imbens and Angrist (1994), and Gospodinov, Kan, and Robotti (2014). 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 interpreting 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 may 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), Kitamura (2005), Gospodinov, Kan, and Robotti (2014), Lee (2018), Hansen and Lee (2021), and Kleibergen and Zhan (2025).

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 continuous updating estimator (CUE) is reflected by a positive difference between the population analogs of the Jacobian rank statistic and the misspecification J-statistic. The same measure is key for identification of the pseudo-true value of the two-stage estimator, which equals a weighted average of the pseudo-true value of the CUE and the, so-called, endogeneity parameter. The weight of the pseudo-true value of the CUE equals zero when it is not identified, so the pseudo-true value of the two-stage estimator then coincides with the endogeneity parameter and does not identify a structural parameter of interest. The resulting tendency of the two-stage estimator to align with the endogeneity parameter is for the same reason as the bias towards it in case of many instruments; see, *e.g.*, Bekker (1994). It all shows that the pseudo-true values of both the CUE and the two-stage estimator do not identify a structural parameter of interest if the population analogs of the traditional rank statistic and the misspecification J-statistic are equal. Because the traditional rank statistic results from constrained optimization of the objective function involved in the J-statistic, their difference, which is our proposed identification measure, is non-negative by construction. 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 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 uncorrelated 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. We also obtain a fixed bounding critical value for a normalized LR statistic. Hereafter, we provide empirically important generalizations which incorporate: covariance matrix estimators, multiple structural parameters, and general covariance structures. Because misspecification allows population moments to be 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), Hansen and Lee (2021), and Kleibergen and Zhan (2025). Alongside the dependence on the conditioning statistic, the conditional critical value function 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 relates the identification measure to the CUE and two-stage estimator, and 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 5 discusses the necessary extensions and also applies the LR test to the Fama and 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, *i.e.*, 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 reject when the zero-beta return is incorporated. Given the importance of the Fama and French (1993) three-factor model, our empirical application illustrates the relevance of using the appropriate identification test. The sixth section draws some conclusions. Technical details and additional findings are relegated to the Supplemental Appendix.

2 Identification in misspecified linear GMM

2.1 Over-identified linear moment equations

We conduct inference on a k-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 $k > m$. The linear moment equations for GMM are specified accordingly:

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

with $J = \frac{\partial}{\partial\theta'}\mu_f(\theta)$. Many widely used econometric models, such as 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 μ, $\hat{\mu}$, and the Jacobian J, $\hat{J}$, whose joint convergence when the sample size N increases results from Assumption 1.

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

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

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

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

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

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

[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. In such cases, an inverse of V should be interpreted as a generalized inverse.

[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-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 $\mu = E(Z_iY_i)$, $J = -E(Z_iX_i')$. Assumption 1 then just imposes that a central limit theorem applies to $\hat{\mu} = \frac{1}{N}\sum_{i=1}^N Z_iY_i$ and the vectorization of $\hat{J} = -\frac{1}{N}\sum_{i=1}^N Z_iX_i'$. If linear moment functions are further adjusted by using a weight matrix, then μ, J change accordingly.

2.2 Identification and misspecification measures

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

$$\mu_f(\theta_0) = \mu + J\theta_0 = 0. \tag{4}$$

The moment vector μ is then spanned by the Jacobian J, so $(\mu \vdots J)$ is a $k \times (m+1)$ dimensional matrix which is at most of rank m:

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

The true value of θ, θ_0, is identified when J 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, like, for example, the Cragg and Donald (1997) rank test. We therefore define the traditional identification measure denoted by IS (identification strength) accordingly.

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, *see also Kleibergen (2007) and Kleibergen and Mavroeidis (2009):*

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

In the setting of one structural parameter so $m = 1$, IS *in (6) is just* $J'V_{JJ}^{-1}J$.

When the moment equations are misspecified, there is no longer a unique 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 may lead to distinct pseudo-true values, while we focus on the pseudo-true value of the CUE. We do so for the invariance properties of the CUE, because of which inference from weak identification robust procedures is centered around it. Later on we show that the identification condition of the CUE is also relevant for the popular two-stage estimator, *i.e.*, the pseudo-true value of the two-stage estimator does not represent a structural parameter of interest and becomes a mere function of the covariance parameters in V when the identification condition for the CUE fails to hold. This important finding for the two-stage estimator thus differs from the IS measure for identification 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:

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

so

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

Next, we define the misspecification measure as the minimum of the CUE population objective function, which is the population analog of the misspecification J-statistic (divided by the sample size).

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

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

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

2.3 IS $-$ MISS as the identification measure

Both IS and MISS defined above are well known, and they are typically discussed separately. In contrast, we show that their difference, IS $-$ MISS, is an appropriate identification measure in over-identified misspecified linear GMM, as follows.

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

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 linear moment equation, so the same value results when we use the linear moment equation:

$$\eta_f(\alpha, \varphi) = J_1 + \mu\alpha + J_2\varphi, \tag{11}$$

with $J = (J_1 \vdots J_2)$, $J_1 : k \times 1$, $J_2 : k \times (m-1)$, 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 $\theta_1 \neq 0$.[4] Thus,

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

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

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

and the proof above similarly holds. For reasons of brevity, we assume the existence of the minimum in (13), but the argument could similarly be extended to cover the infimum. ∎

The MISS $\leq$ IS inequality 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, as shown by Corollary 1.

[4] Since we ultimately care about the sample moment equation, for which the CUE of θ_1 is unequal to zero with probability one, we can without loss of generality assume that $\theta_1 \neq 0$.

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, a GMM identification measure for the pseudo-true value of the CUE which applies to both misspecified and correctly specified linear moment equations is:*

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

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

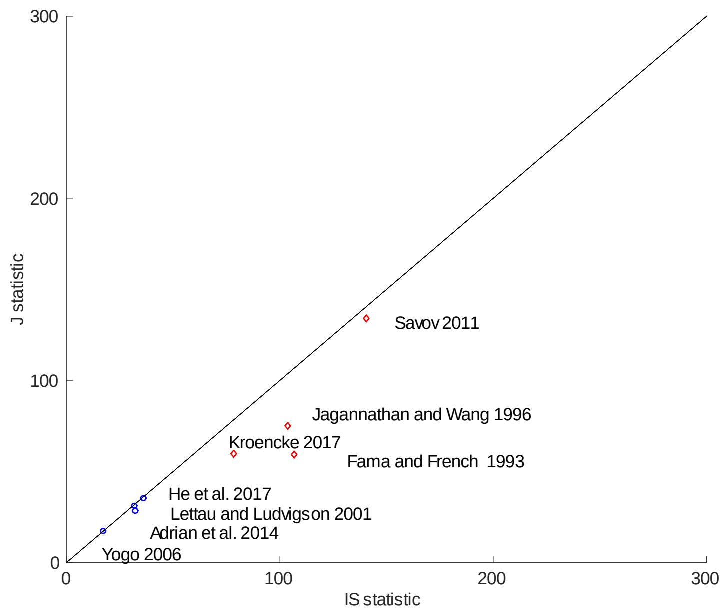
Corollary 3: *The sample value of N times* MISS, $\widehat{\text{MISS}}$, *is always strictly smaller than the sample value of N times* IS, $\widehat{\text{IS}}$*:*

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

Taken together, Theorem 1 and its Corollaries 1-3 show that just using $\widehat{\text{IS}}$ to gauge identification overstates the identification strength when the linear moment equations are potentially misspecified. Many commonly used empirical models may involve misspecified moment equations, so their identification strength tends to be overstated when using $\widehat{\mathrm{IS}}$ as the measure of identification.

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

Figure 1: Scatter plot of $\widehat{\mathrm{IS}}$ and $\widehat{\mathrm{MISS}}$ (=J) statistics for different specifications.



Notes: Figure 1 shows $\widehat{\mathrm{IS}}$ and $\widehat{\mathrm{MISS}}$ (=J) 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.

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 $\widehat{\mathrm{IS}}$ and $\widehat{\mathrm{MISS}}$ (=J-statistic) 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).[5] Appendix A1-A2 provide the specifications of the above linear asset pricing models, as well as the detailed expressions of $\widehat{\text{IS}}$ and $\widehat{\text{MISS}}$ for gauging the identification of risk premia (*i.e.*, the parameters of interest), respectively.

Because $\widehat{\text{MISS}} < \widehat{\text{IS}}$, all scatter points in Figure 1 are below the 45^0-degree line, but their proximity to it is striking. For all points, the distance to the 45^0-degree line measured by $\widehat{\text{IS}} - \widehat{\text{MISS}}$, is much smaller than the value of $\widehat{\text{IS}}$. Overall, Figure 1 shows that $\widehat{\text{IS}}$ can severely overstate the identification strength in empirical studies, so it is important to use $\widehat{\text{IS}} - \widehat{\text{MISS}}$ as the identification measure.

Table 1: $\widehat{\text{IS}}$ and $\widehat{\text{MISS}}$ statistics

Notes: Panel A contains the $\widehat{\text{IS}}$ and $\widehat{\text{MISS}}$ statistics from Figure 1, for which the zero-beta return, denoted 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***

Table 1 illustrates the effect on identification and misspecification from imposing further restrictions. The left hand side of Table 1 contains the estimates of IS and MISS 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

[5]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).

zero-β return, λ_0, is zero, as described in Appendix A2. 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 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 outweighed 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.

To summarize, Figure 1 and Table 1 above just showed the empirical relevance of the $\widehat{\text{IS}}-\widehat{\text{MISS}}$ identification measure by using eight examples of asset pricing models, while we could have done so using other important empirical models as well. For example, Kleibergen and Zhan (2025) consider the return to schooling study of Card (1995) for which different local average treatment effects of the (three) involved instruments make the linear IV regression model misspecified; see Imbens and Angrist (1994). For this model, $\widehat{\text{IS}}-\widehat{\text{MISS}}$ is then also the appropriate identification measure to use.

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 the over-identified linear GMM that is potentially misspecified. We do so in a sequence of steps, which correspond to Subsections 4.1-4.4, respectively:[6]

1. For a Kronecker product form covariance structure, we characterize the data generating process (DGP) parameter setting which leads to equality of IS and MISS.

[6] For ease of exposition, further extensions are provided in the later Section 5.

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, $\widehat{\mathrm{IS}} - \widehat{\mathrm{MISS}}$ 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 uncorrelated with the LR statistic under our hypothesis of interest. It enables us to compute a conditional critical value function for the LR test of no-identification with correct size.

4. We present a simulation study that illustrates the size and power properties of the LR test.

4.1 DGP parameter setting for IS = MISS

To obtain the parameter setting of the DGP for equality of IS and MISS, we start by assuming a Kronecker product structure (KPS) of the joint covariance matrix V used in Assumption 1. In the later Subsection 5.3, we show how to extend the no-identification LR test from the KPS covariance matrix to more general covariance matrix structures.

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

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

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

The KPS of the covariance matrix results, for example, under i.i.d. homoskedastic 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:

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

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

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

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

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

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* $\mu_J = \mu - J\Omega_{JJ}^{-1}\omega_{J\mu}$*, 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* μ_J*,* J*, respectively:*

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

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, $m = 1$, 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: *Under Assumptions 1 and 2, 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 show the implications of the no-identification DGP parameter setting for the pseudo-true values of the CUE and two-stage estimator for a single structural parameter, $m = 1$. For this purpose, we use θ^*_{2S} and θ^*_{CUE} in this subsection for the pseudo-true values of the two-stage estimator and CUE, respectively.

By employing the K-class notation (see, *e.g.*, Hausman (1983)), we express θ^*_{2S} and θ^*_{CUE} using C and a from (24):

$$\begin{array}{rcll} \theta^*(K) & = & -\omega^{\frac{1}{2}}_{\mu\mu.J}\Omega^{-\frac{1}{2}}_{JJ}(\frac{C'C}{N} - K)^{-1}\frac{C'a}{N} - \Omega^{-1}_{JJ}\omega_{J\mu} & : K\text{-class notation} \\ \theta^*_{2S} & = & \theta^*(K = 0) & : \text{Two-stage} \\ \theta^*_{\text{CUE}} & = & \theta^*(K = \text{MISS}). & : \text{CUE} \end{array} \tag{28}$$

which implies a linear relationship between θ^*_{2S} and θ^*_{CUE}, as stated in Corollary 5.

Corollary 5: *Under Assumptions 1 and 2, for the single structural parameter setting,* $m = 1$, θ^*_{2S} *is a weighted average of* θ^*_{CUE} *and the endogeneity parameter* $\rho \equiv -\Omega^{-1}_{JJ}\omega_{J\mu}$*:*

$$\theta^*_{2S} \quad = \quad \left(1 - \tfrac{\text{MISS}}{\text{IS}}\right)\theta^*_{\text{CUE}} + \tfrac{\text{MISS}}{\text{IS}}\rho. \tag{29}$$

Because MISS $\leq$ IS as shown by Theorem 1, both weights in (29) are non-negative and add up to one. When IS = MISS, the pseudo-true value of the CUE is not identified, and (29) shows that the pseudo-true value of the two-stage estimator just equals the endogeneity parameter ρ, which is a function of the (residual) covariance parameters and does not reflect a structural parameter of interest. Hence, no-identification of the CUE implies that the two-stage estimator does not identify a structural relationship where the moment vector is spanned by its Jacobian.

Overall, Corollary 5 shows that neither the pseudo-true value of the CUE nor the pseudo-true value of the two-stage estimator represents a structural parameter of interest under the no-identification setting from Corollary 4. Therefore, the no-identification setting we consider is crucial for both the CUE and the two-stage estimator. Consequently, our proposed LR no-identification test is practically relevant, regardless of whether practitioners adopt the two-stage estimator or the CUE.

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

In this subsection, we develop our LR no-identificaton test. 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 multiple structural parameters in the later Section 5. 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 in (6) and MISS in (9) equal the CUE objective function at set or minimized values respectively, $\widehat{\mathrm{IS}} - \widehat{\mathrm{MISS}}$ equals the (quasi-) likelihood ratio (LR) statistic that tests $\mathrm{H}_0 : \alpha = 0$ in (14):

$$\mathrm{LR}(\alpha = 0) = \widehat{\mathrm{IS}} - \widehat{\mathrm{MISS}} = N \times \hat{Q}_{CUE}(\alpha = 0) - N \times \min_{\alpha \in \mathbb{R}} \hat{Q}_{CUE}(\alpha), \tag{30}$$

with $\hat{Q}_{CUE}(\alpha)$ the sample analog of $Q_{CUE}(\alpha)$.[7]

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

$$\mathrm{LR}(\alpha = 0) = \tfrac{1}{2}\left[\hat{C}'\hat{C} - \hat{a}'\hat{a} + \sqrt{\left(\hat{C}'\hat{C} - \hat{a}'\hat{a}\right)^2 + 4(\hat{C}'\hat{a})^2}\right], \tag{31}$$

for $\hat{a} = \sqrt{N}\Sigma^{-\frac{1}{2}}\hat{\mu}_J\omega_{\mu\mu.J}^{-\frac{1}{2}}$, $\hat{C} = \sqrt{N}\Sigma^{-\frac{1}{2}}\hat{J}\Omega_{JJ}^{-\frac{1}{2}}$ and $\hat{\mu}_J = \hat{\mu} - \hat{J}\Omega_{JJ}^{-1}\omega_{J\mu}$. While the expression of the LR statistic (31) is for testing $\mathrm{H}_0 : \alpha = 0$, our no-identification hypothesis of interest concerns a specific setting of (a, C) that implies $\alpha = 0$ as

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

provided in Corollary 4:

$$\mathrm{H}_{\text{non-ind}} : \mathrm{IS} = \mathrm{MISS} \Longleftrightarrow C'C - a'a \leq 0 \text{ and } C'a = 0. \tag{32}$$

It is clear that the expression of the LR statistic (31) contains the sample analogs of all elements of $\mathrm{H}_{\text{non-ind}}$.

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

$$\mathrm{H}^*_{\text{non-ind}} : C'C - a'a = 0,\ C'a = 0. \tag{33}$$

A conditional critical value function that controls the size of the LR test for $\mathrm{H}^*_{\text{non-ind}}$ also does so for $\mathrm{H}_{\text{non-ind}}$, because the derivative of $\mathrm{LR}(\alpha = 0)$ 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 $\mathrm{LR}(\alpha = 0)$ and will not lead to size distortion.

Corollary 6 states the limit behavior for $\hat{a}$ and $\hat{C}$ in the LR statistic (31), which results from Assumptions 1 and 2.

Corollary 6: *When Assumptions 1 and 2, $m = 1$, and the drifting sequence in (24) apply, the limit behavior of $\hat{a} = \sqrt{N}\Sigma^{-\frac{1}{2}}\hat{\mu}_J\omega_{\mu\mu.J}^{-\frac{1}{2}}$, $\hat{C} = \sqrt{N}\Sigma^{-\frac{1}{2}}\hat{J}\Omega_{JJ}^{-\frac{1}{2}}$ with $\hat{\mu}_J = \hat{\mu} - \hat{J}\Omega_{JJ}^{-1}\omega_{J\mu}$, is:*

$$\begin{pmatrix} \hat{a} \\ \hat{C} \end{pmatrix} \underset{d}{\rightarrow} \begin{pmatrix} a \\ C \end{pmatrix} + \begin{pmatrix} \psi^*_{\mu.J} \\ \psi^*_J \end{pmatrix}, \tag{34}$$

where $\psi^*_{\mu.J}$ and ψ^*_J are independent k-dimensional standard normal random vectors.

The LR statistic (31) is an invariant statistic and therefore, as all invariant statistics are, a function of the maximal invariant which equals the normalized quadratic form of $(\hat{\mu} \vdots \hat{J})$, see, *e.g.*, Andrews, Moreira, and Stock (2006):

$$\begin{aligned} \mathrm{MAXINV} = &\ N \times \Omega^{-\frac{1}{2}\prime}(\hat{\mu} \vdots \hat{J})'\Sigma^{-1}(\hat{\mu} \vdots \hat{J})\Omega^{-\frac{1}{2}} \\ = &\ (\hat{a} \vdots \hat{C})'(\hat{a} \vdots \hat{C}) \\ = &\ \begin{pmatrix} \hat{a}'\hat{a} & \vdots & \hat{a}'\hat{C} \\ \hat{C}'\hat{a} & \vdots & \hat{C}'\hat{C} \end{pmatrix}. \end{aligned} \tag{35}$$

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 to get:

$$\text{MAXINV} = (\hat{C}'\hat{C} - \hat{a}'\hat{a},\ \hat{C}'\hat{a},\ \hat{C}'\hat{C} + \hat{a}'\hat{a}), \tag{36}$$

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 (31) is only a function of the first two elements of the maximal invariant (36): $\hat{C}'\hat{C} - \hat{a}'\hat{a}$, $\hat{C}'\hat{a}$, and not of the third element, $\hat{C}'\hat{C} + \hat{a}'\hat{a}$.
2. Under $\mathrm{H}^*_{\text{non-ind}}$, the third element of (36) is asymptotically uncorrelated with the first two, as shown by Theorem 2 below.

Theorem 2: *Under the conditions of Corollary 6 and* $\mathrm{H}^*_{\text{non-ind}}$*, the joint limit behavior of the three components* $\hat{C}'\hat{C} - \hat{a}'\hat{a}$, $\hat{C}'\hat{a}$, $\hat{C}'\hat{C} + \hat{a}'\hat{a}$ *can be written as:*

$$\begin{aligned}
&\begin{pmatrix} \hat{C}'\hat{C} - \hat{a}'\hat{a} \\ \hat{C}'\hat{a} \\ \hat{C}'\hat{C} + \hat{a}'\hat{a} \end{pmatrix} \\
\underset{d}{\to}\ & \begin{pmatrix} 0 \\ 0 \\ C'C + a'a \end{pmatrix} + \begin{pmatrix} 2\binom{C}{-a}' \\ \binom{a}{C}' \\ 2\binom{C}{a}' \end{pmatrix} \begin{pmatrix} \psi^*_J \\ \psi^*_{\mu.J} \end{pmatrix} + \begin{pmatrix} \binom{\psi^*_J}{-\psi^*_{\mu.J}}' \\ \binom{0}{\psi_{J^*}}' \\ \binom{\psi^*_J}{\psi^*_{\mu.J}}' \end{pmatrix} \begin{pmatrix} \psi^*_J \\ \psi^*_{\mu.J} \end{pmatrix} \qquad (37) \\
=\ & \mathbf{N}\left(\begin{pmatrix} 0 \\ 0 \\ C'C + a'a \end{pmatrix},\ (C'C + a'a) \begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix} \right) + \begin{pmatrix} \binom{\psi^*_J}{-\psi^*_{\mu.J}}' \\ \binom{0}{\psi_{J^*}}' \\ \binom{\psi^*_J}{\psi^*_{\mu.J}}' \end{pmatrix} \begin{pmatrix} \psi^*_J \\ \psi^*_{\mu.J} \end{pmatrix}.
\end{aligned}$$

Under $\mathrm{H}^*_{\text{non-ind}}$*, the dependence between the three components thus only results from the last elements of their expressions in (37) which are uncorrelated.*

Proof. Results from Corollary 6 and applying $\mathrm{H}^*_{\text{non-ind}}$ (33), 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. ■

Theorem 2 shows that under Assumptions 1, 2, the drifting sequence in (24) and $\mathrm{H}^*_{\text{non-ind}}$, the only parameter where the limiting distribution of the maximal invariant (36) depends on is $C'C + a'a$, and therefore so does $\mathrm{LR}(\alpha = 0)$. We therefore use:

$$\mathrm{rk} = \hat{C}'\hat{C} + \hat{a}'\hat{a}, \tag{38}$$

as a conditioning statistic for the only parameter in the distribution of $\mathrm{LR}(\alpha = 0)$ (31). The decomposition in (37) shows that the dependence between the different components becomes negligible for larger values of $C'C + a'a$.

The conditioning statistic (38) 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:

$$\mathrm{rk}_{\text{Moreira (2003)}} = \hat{a}'\hat{a}. \tag{39}$$

Under the hypothesis of interest of Moreira (2003), (39) provides a sufficient statistic for the only parameter which the LR statistic then depends on and is also independently distributed of. Our conditioning statistic (38) provides an estimator for the only parameter the limiting distribution of the LR statistic (31) depends on under $\mathrm{H}^*_{\text{non-ind}}$, *i.e.*, $C'C + a'a$, but it is not a sufficient statistic. It is also asymptotically uncorrelated with the LR statistic (31), because of which we use a different simulation algorithm to obtain conditional critical values than Moreira (2003).[8]

Remark 1: The algorithm for computing the conditional critical value function is stated in Supplemental Appendix SA1. 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 $\mathrm{H}^*_{\text{non-ind}}$, it computes the conditional distribution of $\mathrm{LR}(\alpha = 0)$ given rk$= \hat{C}'\hat{C} + \hat{a}'\hat{a}$. The conditional critical value function for conducting a 5% significance test of $\mathrm{H}^*_{\text{non-ind}}$ using $\mathrm{LR}(\alpha = 0)$ then corresponds to the 95% percentiles of the computed conditional distribution of $\mathrm{LR}(\alpha = 0)$ given

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

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

Remark 2: Theorem 2 shows that the leading variances of the different components of LR($\alpha = 0$) in (31) are all proportional to $C'C + a'a$, so we can standardize LR($\alpha = 0$) by dividing $\sqrt{\text{rk}}$, which results in a standardized bounding distribution for the LR no-identification test:

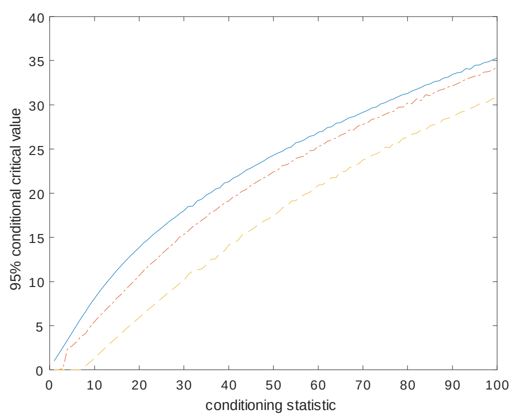
$$\tfrac{\text{LR}(\alpha=0)}{\sqrt{\text{rk}}} \preceq \psi_1 + \sqrt{\psi_1^2 + \psi_2^2}, \tag{40}$$

with ψ_1 and ψ_2 independent standard normal random variables and "$\preceq$" indicates stochastic dominance.[9] The bounding argument results because (37) shows that rk by construction over-estimates the variance of the components of LR($\alpha = 0$). For larger values of $C'C + a'a$, the upper bound becomes an equality.

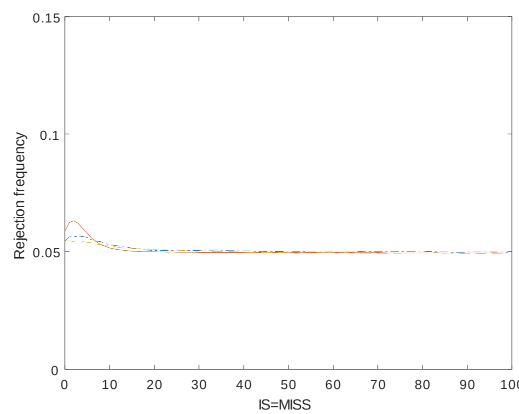
4.4 Simulation

4.4.1 Size of the LR test for $m = 1$

Figure 2: 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 in (38), and the resulting rejection frequencies for $k = 3$ (solid), 10 (dash-dotted) and 25 (dashed).



Panel 2.1. Conditional critical values



Panel 2.2. Rejection frequencies

Figure 2 shows the 95% conditional critical value functions for different numbers

[9]For two random variables x and y: $x \preceq y$ implies that $\Pr(x \geq cv) \leq \Pr(y \geq cv)$, where cv stands for a given critical value.

of moment equations in Panel 2.1, and the resulting rejection frequencies when we conduct a 5% significance LR test of $\mathrm{H}^*_{\text{non-ind}}$ in Panel 2.2. It is striking how close the rejection frequencies in Panel 2.2 are to the nominal 5%. There is only some minor over-rejection 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 uncorrelatedness of the conditioning statistic (38) allows us to compute a conditional critical value function for the LR test of no-identification with excellent size properties.

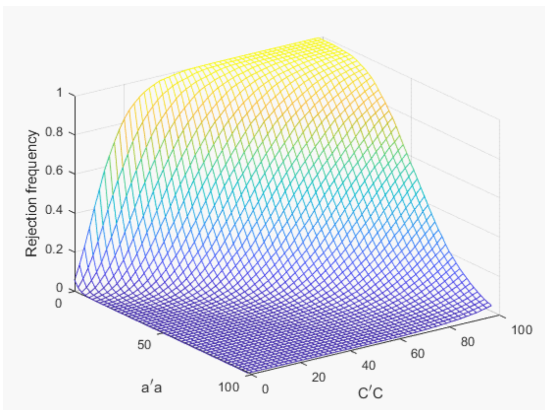
Additional simulation results are provided in the Supplemental Appendix SA7, where we show that using the bounding distribution for the LR no-identification test in (40) leads to conservative sizes, *i.e.*, the rejection frequencies at the null are bounded by the nominal 5%, which is as expected.

4.4.2 Power of the LR test for $m = 1$

Figure 3 shows power surfaces of the LR test of no-identification. The involved conditional critical value function is calibrated to the boundary setting of no-identification $\mathrm{H}^*_{\text{non-ind}}$ (33), but our main hypothesis of interest is $\mathrm{H}_{\text{non-ind}}$ (32). The power surfaces in Figure 3 show the rejection frequencies with respect to potential violation of one of the two components in $\mathrm{H}_{\text{non-ind}}$ (32) while the other one is kept at the hypothesized value. Panel 3.1, and its contour lines in Panel 3.3, therefore show the power surface of the LR test for violations of $C'C \leq a'a$ while $C'a = 0$, and Panel 3.2 shows it for violations of $C'a = 0$ while $C'C = a'a$.

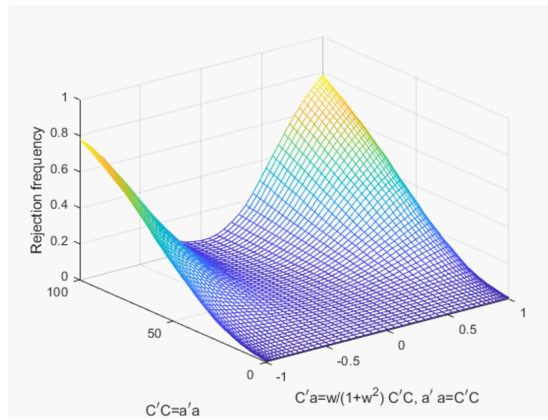
Panels 3.1 and 3.3 show that the conditional LR test controls the size of testing $\mathrm{H}_{\text{non-ind}}$ (32) because, while the conditional critical values are calibrated to $\mathrm{H}^*_{\text{non-ind}}$ (33), the rejection frequency is at most 5% when $C'C \leq a'a$ as indicated by the contour lines in Panel 3.3. When $\mathrm{H}_{\text{non-ind}}$ does not hold, so $C'C$ exceeds $a'a$, Panels 3.1 and 3.3 show that the LR test has discriminatory power for rejecting $\mathrm{H}_{\text{non-ind}}$. Interestingly, the contour line at 5% in Panel 3.3 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 3.2 similarly shows that the LR test has discriminatory power for detecting non-zero values of $C'a$ when $C'C = a'a$.

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



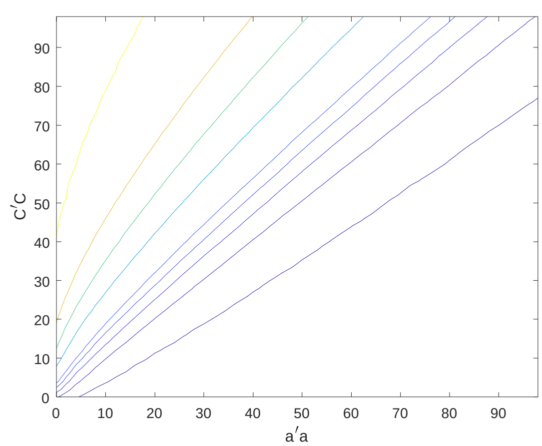
Panel 3.1.

$C'a = 0$



Panel 3.2.

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



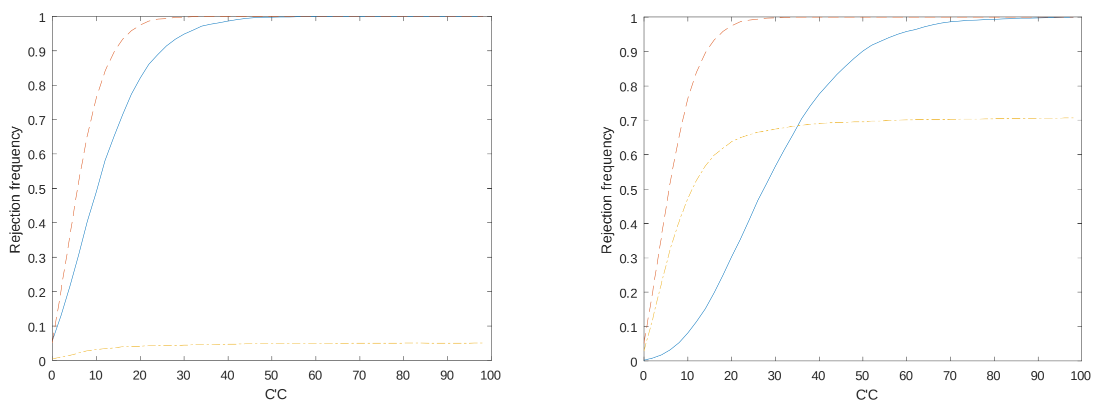
Panel 3.3. Contour lines of Panel 3.1 at 1, 5, 10, 15, 20, 40, 60, 80 and 99%.

Notes: In Panel 3.2, w varies between -1 and 1 on the x-axis, so that $C'a$ deviates from 0 accordingly. The y-axis shows the variation of $C'C = a'a$ between 0 and 100.

4.4.3 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 4 therefore shows the power of: the LR test of no-identification, the IS test whose test statistic equals the F-statistic for testing $J = 0$ when $m = 1$, and the MISS test whose test statistic equals the J-statistic. The level of significance is 5%.

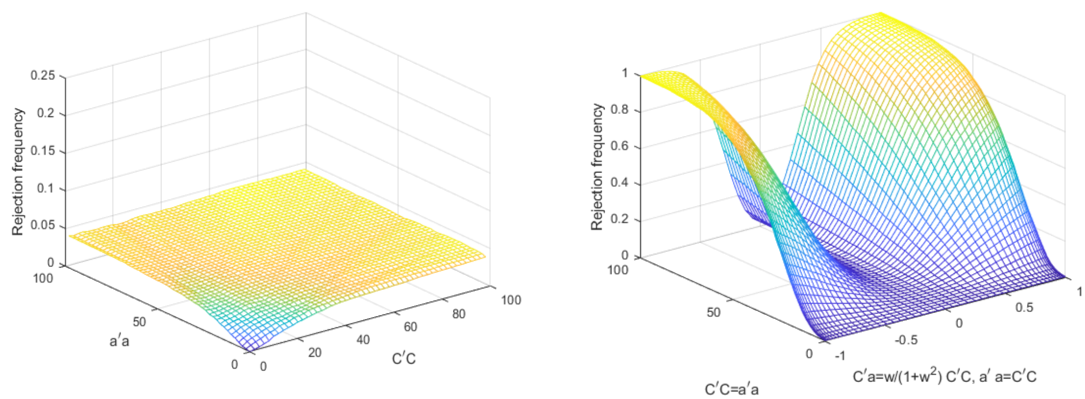
Figure 4: Power of 5% significance conditional LR no-identification test (solid), IS (dashed) and MISS (dash-dotted) tests, $k = 3$.



Panel 4.1. $a'a = 0,\ C'a = 0$ Panel 4.2. $a'a = 10,\ C'a = 0$

Panel 4.1 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 4.2 has an increased level of misspecification, $a'a = 10$, for which the power curve of the 5% IS test has not changed compared to Panel 4.1. Up to $C'C = 10$ there is, however, no identification. The rejection frequency of the MISS test at $C'C = 10$, is around 45%. Panel 4.2 is interesting because the MISS test still mostly does not reject up to $C'C = 10$, 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 4 also shows the LR test is size-correct, *i.e.*, its rejection frequencies are near the nominal 5% at $C'C = a'a = 0$ in Panel 4.1, and at $C'C = a'a = 10$ in Panel 4.2, respectively.

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



Panel 5.1.
$C'a = 0$

Panel 5.2.
$C'C = a'a,\ C'a = \frac{w}{1+w^2}C'C$

Notes: In Panel 5.2, w varies between -1 and 1 on the x-axis, so that $C'a$ deviates from 0 accordingly. The y-axis shows the variation of $C'C = a'a$ between 0 and 100.

Figure 5 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 3. The DRLM test is a size-correct test of $\mathrm{H}_0 : \alpha = 0$ 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 5.1, for which $C'a = 0$ is imposed. In contrast, Panel 5.2 shows that the DRLM test has good power, exceeding that of the LR test in Panel 3.2. Similar to the IS test in Panel 4.1, Figure 5 shows that for specific settings, the power of tests of just one component of the composite hypothesis $\mathrm{H}_{\text{non-ind}}$ (32) can exceed that of the LR test, but these tests have misleading power or no power when the other component of the composite hypothesis $\mathrm{H}_{\text{non-ind}}$ (32) gets violated.

5 Extension and application

In this section, we generalize the LR test presented before to allow for covariance matrix estimators and multiple structural parameters. The LR no-identification test is then applied to the asset pricing studies previously presented in Section 3. We also extend to general covariance structures.

5.1 Covariance matrix estimators and multiple parameters

5.1.1 KPS covariance matrix estimator, $m = 1$

The covariance matrices Ω and Σ in Assumption 2 are, for ease of exposition, previously considered to be known. In empirical settings, these covariance matrices are unknown so we use consistent estimators $\hat{\Omega}$ and $\hat{\Sigma}$ instead, which can be written as:

$$\begin{aligned} \hat{\Omega} = & \begin{pmatrix} \hat{\omega}_{\mu\mu} & \vdots & \hat{\omega}_{\mu J} \\ \hat{\omega}_{J\mu} & \vdots & \hat{\Omega}_{JJ} \end{pmatrix} = \Omega + O_p(N^{-\frac{1}{2}}) \\ \hat{\Sigma} = & \ \Sigma + O_p(N^{-\frac{1}{2}}), \end{aligned} \tag{41}$$

with $O_p(N^{-\frac{1}{2}})$ reflects the estimation errors that occur in the estimation of Ω and Σ, respectively. We correspondingly replace the covariance matrices in the expressions of a and C in (24) by their estimators in (41).

Unlike correctly specified moment equations, the non-zero moment equations at the pseudo-true value imply that we have to incorporate the estimation error of the covariance matrix for an accurate approximation of the conditional distribution of the LR statistic; see, *e.g.*, Maasoumi and Phillips (1982), Hall and Inoue (2003), Gospodinov, Kan, and Robotti (2014), Lee (2018), and Hansen and Lee (2021). A higher-order approximation of the components in Theorem 2, $\hat{C}'\hat{C} - \hat{a}'\hat{a}$, $\hat{C}'\hat{a}$, $\hat{C}'\hat{C} + \hat{a}'\hat{a}$, shows that their asymptotic uncorrelatedness remains present, as in Theorem 3.

Theorem 3: *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, incorporating the covariance estimators in (41) for $\hat{C}'\hat{C} - \hat{a}'\hat{a}$, $\hat{C}'\hat{a}$ and $\hat{C}'\hat{C} + \hat{a}'\hat{a}$ makes the asymptotic uncorrelatedness from Theorem 2 extend to their higher-order elements up to order $N^{-\frac{1}{2}}$, because lower order* *components are uncorrelated while dependent components are of a higher negative order in the sample size and therefore dominated by lower order ones.*

Proof. See Supplemental Appendix SA2. ■

Alongside the asymptotic uncorrelatedness of the different elements of the LR statistic, the higher-order expansion constructed in the proof of Theorem 3 also shows that larger identification and misspecification strengths affect the conditional distribution of the LR statistic under our hypothesis of interest compared to the known covariance matrix setting in Theorem 2. This finding is stated in Corollary 7.

Corollary 7: *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 $\sqrt{N}$, the higher-order elements of $\hat{C}'\hat{C} - \hat{a}'\hat{a}$, $\hat{C}'\hat{a}$ and $\hat{C}'\hat{C} + \hat{a}'\hat{a}$ affect the large sample behavior of the LR statistic and have to be accounted for in the conditional critical value function of the LR test.*

Proof. See Supplemental Appendix SA2. ■

Theorem 3 and Corollary 7 show that we have to account for the estimation error of the covariance matrix for the conditional distribution of the LR statistic. They also show that the key conditioning argument for obtaining the conditional critical values remains unaltered, so we can adapt the simulation algorithm from Supplemental Appendix SA1 accordingly. This adaptation, which accounts for higher-order elements, is provided for the multiple parameters setting in Supplemental Appendix SA5.

5.1.2 Multiple structural parameters, known KPS covariance

For more than one structural parameter, $m > 1$, the notation is heavier, while the underlying idea remains the same as the $m = 1$ case in Section 4. The difference between the sample analogs of IS and MISS thus equals the LR statistic that tests for a zero value of α in the population moment equation:[10]

$$\eta_f(\alpha, \varphi) = J_1 + \mu\alpha + J_2\varphi, \tag{42}$$

[10] See also (11). The moment equation in (42) is normalized using the first column of J. The LR statistic is invariant with respect to this normalization, so an identical value results when normalized using any other column of J.

with $J=(J_1 \vdots J_2)$, $J_1 : k \times 1$, $J_2 : k \times (m-1)$, so

$$\begin{aligned}
\mathrm{LR}(\alpha=0) &= \widehat{\mathrm{IS}} - \widehat{\mathrm{MISS}} \\
&= N \times \min_{\varphi \in \mathbb{R}^{m-1}} \hat{Q}_{CUE}(\alpha=0,\varphi) - N \times \min_{\alpha \in \mathbb{R}, \varphi \in \mathbb{R}^{m-1}} \hat{Q}_{CUE}(\alpha,\varphi), \quad (43)
\end{aligned}$$

where $\hat{Q}_{CUE}(\alpha,\varphi)$ uses $\hat{\eta}_f(\alpha,\varphi) = \hat{J}_1 + \hat{\mu}\alpha + \hat{J}_2\varphi$:

$$\hat{Q}_{CUE}(\alpha,\varphi) = \hat{\eta}_f(\alpha,\varphi)' \left[\begin{pmatrix} (\alpha \otimes I_k) \\ I_k \\ (\varphi \otimes I_k) \end{pmatrix}' \hat{V} \begin{pmatrix} (\alpha \otimes I_k) \\ I_k \\ (\varphi \otimes I_k) \end{pmatrix} \right]^{-1} \hat{\eta}_f(\alpha,\varphi). \qquad (44)$$

While the expression of the LR statistic in (43) 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 a known KPS covariance matrix from Assumption 2 that implies equality of IS and MISS, as shown in Theorem 4 which generalizes Corollary 4 to multiple structural parameters.

Theorem 4: *Under Assumptions 1, 2 and using the specification from (24),* IS *equals the squared smallest singular value that results from a singular value decomposition (SVD) of C :*

$$C = U_C S_C V_C', \qquad (45)$$

with $U_C = (U_{C,1} \vdots U_{C,m} \vdots U_{C,2})$ an orthonormal $k \times k$ dimensional matrix, $U_{C,1}$: $k \times (m-1)$, $U_{C,m} : k \times 1$, $U_{C,2} : k \times (k-m)$ dimensional matrices, $V_C = (V_{C,1} \ldots V_{C,m})$ an $m \times m$ dimensional orthonormal matrix, and S_C a $k \times m$ 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:*

$$\mathrm{H}_{\text{non-ind}} : U_{C,m}' a = 0 \text{ and } b'b \geq s_{C,mm}^2 \Leftrightarrow \mathrm{H}_{\text{non-ind}} : \mathrm{IS} = \mathrm{MISS}, \qquad (46)$$

where $b = U_{C,2}' a$.

Proof. See Supplemental Appendix SA3. ■

Theorem 4 extends Corollary 4 to more than one structural parameter, and states the no-identification condition. 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 along the lines for the single structural parameter setting.

Theorem 5: *Under Assumptions 1 and 2, the drifting sequence in (24), the SVD of* $\hat{C} = \sqrt{N}\hat{\Sigma}^{-\frac{1}{2}}\hat{J}\hat{\Omega}_{JJ}^{-\frac{1}{2}}$:

$$\hat{C} = \hat{U}_C \hat{S}_C \hat{V}_C', \tag{47}$$

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

$$\begin{aligned} \mathrm{LR}(\alpha = 0) &= \widehat{\mathrm{IS}} - \widehat{\mathrm{MISS}} & (48) \\ &= \frac{1}{2}\left[\hat{s}_{C,mm}^2 - \hat{a}' M_{\hat{U}_{C,1}} \hat{a} + \sqrt{\left(\hat{s}_{C,mm}^2 - \hat{a}' M_{\hat{U}_{C,1}} \hat{a}\right)^2 + 4\hat{s}_{C,mm}^2 (\hat{U}_{C,m}' \hat{a})^2}\right] \\ &\quad + O(\hat{s}_{C,(m-1)(m-1)}^{-2}), \end{aligned}$$

for $\hat{a} = \sqrt{N}\hat{\Sigma}^{-\frac{1}{2}}\left(\hat{\mu} - \hat{J}\hat{\Omega}_{JJ}^{-1}\hat{\omega}_{J\mu}\right)\hat{\omega}_{\mu\mu.J}^{-\frac{1}{2}}$.

Proof. See Supplemental Appendix SA4. ■

Theorem 5 shows that the expression of the no-identification LR statistic (31) from the single structural parameter setting (approximately) extends to multiple structural parameters. Likewise, we can also extend Theorem 2 to multiple structural parameters. For reasons of brevity we directly state the extension in Corollary 8 below.

Corollary 8: *Under Assumptions 1 and 2, the drifting sequence in (24), large singular values* $s_{C,11} \ldots s_{C,(m-1)(m-1)} \gg s_{C,mm}$: $\hat{U}_{C,1}$ *is a consistent estimator of* $U_{C,1}$ *because of which Theorem 2 implies that under* $\mathrm{H}_{\text{non-ind}}^* : U_{C,m}' a = 0,\ s_{C,mm}^2 = b'b =$

[11] The $O(\hat{s}_{C,(m-1)(m-1)}^{-2})$ indicates that the approximation error is proportional to $\hat{s}_{C,(m-1)(m-1)}^{-2}$ and therefore small when it is assumed that $\hat{s}_{C,(m-1)(m-1)}$ is large.

$a'M_{U_{C,1}}a$, *the asymptotic behaviors of:*

$$\hat{s}^2_{C,mm} - \hat{a}'M_{\hat{U}_{C,1}}\hat{a},\ \hat{U}'_{C,m}\hat{a},\ \textit{and}\ \hat{s}^2_{C,mm} + \hat{a}'M_{\hat{U}_{C,1}}\hat{a}, \tag{49}$$

are approximately uncorrelated up to order $N^{-\frac{1}{2}}$, *and therefore the asymptotic uncorrelatedness also holds for:*

$$\mathrm{LR}(\alpha = 0)\ \textit{and}\ \hat{s}^2_{C,mm} + \hat{a}'M_{\hat{U}_{C,1}}\hat{a}. \tag{50}$$

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 no-identification LR test. When $s_{C,11} \dots s_{C,(m-1)(m-1)}$ are large and under $\mathrm{H}^*_{\text{non-ind}}$, the proof of Theorem 5 shows that it also corresponds to the sum of the smallest two characteristic roots of the sample analog of (25), which therefore provides an easy-to-compute conditioning statistic, as stated in Corollary 9.

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

$$\mathrm{rk} = \widehat{\mathrm{MISS}} + \hat{\lambda}_{aC,m}, \tag{51}$$

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

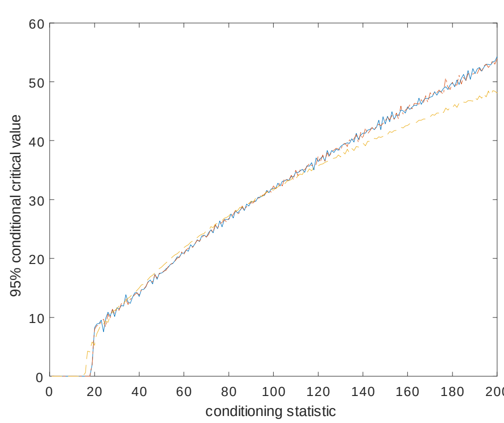
5.1.3 Multiple structural parameters, unknown covariance matrix

We have previously shown that for the single structural parameter, the approximate uncorrelatedness of the conditioning statistic for the LR test remains to hold when we incorporate the KPS covariance matrix estimators in (41). Supplemental Appendix SA5 therefore extends SA1 and provides an algorithm for computing the conditional distribution of the LR statistic given the conditioning statistic that accounts for the higher-order elements in the conditional distribution resulting from the covariance matrix estimator and multiple structural parameters. We can also similarly extend the bounding distribution in (40) to account for higher-order approximations. The application of the LR test to the Fama and French (1993) three-factor model is thus discussed next.

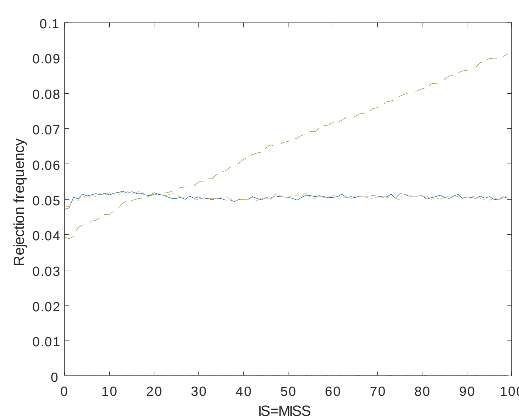
5.2 Fama and French (1993) model, $m = 3$, $k = 25$

For the Fama and French (1993) three-factor model using data from Lettau, Ludvigson, and Ma (2019), Panel 6.1 in Figure 6 shows the 95% conditional critical value functions by using the algorithm from Supplemental Appendix SA5. Panel 6.1 shows that these critical value functions for the specifications with or without the zero-β return are basically identical, but differ from the one which would result when we ignore the estimation error of covariance matrix estimators. Panel 6.2 in Figure 6 shows the resulting rejection frequencies with IS = MISS, which are very close to the nominal 5% for both specifications. Panel 6.2 also shows that usage of the conditional critical value function without incorporating the estimation error of covariance matrix estimators leads to considerable size distortion (dashed).

Figure 6: 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 in (51) calibrated to the Fama and French (1993) three-factor model with zero-β return (solid), without (dash-dotted), known covariance (dashed) and resulting rejection frequencies, $m = 3$, $k = 25$.

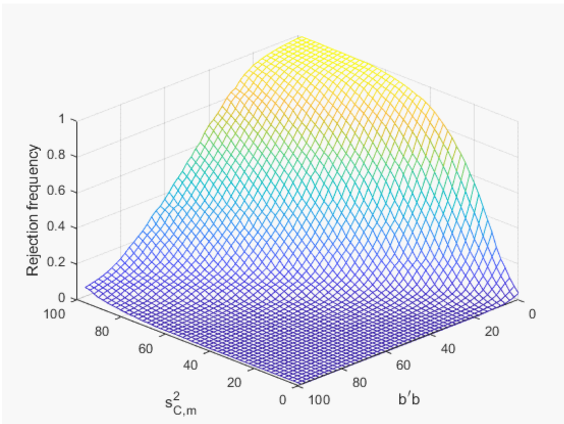


Panel 6.1. Conditional critical values



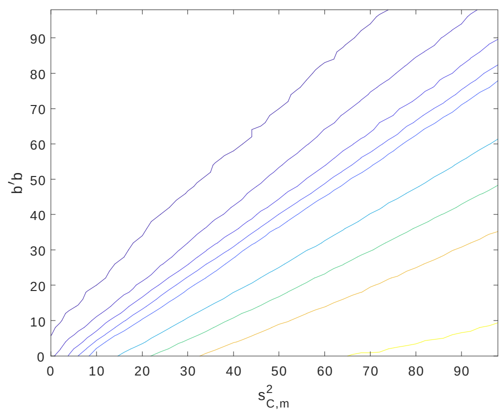
Panel 6.2. Rejection frequencies

Figure 7: Power of 5% significance LR no-identification test calibrated to the Fama and French (1993) three-factor model, $m = 3$, $k = 25$, $\lambda_0 \neq 0$.



Panel 7.1.

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



Panel 7.2.

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

Figure 7 illustrates the power of the LR test calibrated to the Fama and French (1993) three-factor model with the zero-β return and using the Lettau, Ludvigson, and Ma (2019) data. Panel 7.1 shows the power surface over IS and MISS, while $U'_{C,m}a = 0$, and Panel 7.2 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.

Overall, Figures 6-7, which are the counterparts of the previous Figures 2-3, show that the size and power of the proposed LR test work well when calibrated to empirical settings with covariance matrix estimation and multiple parameters.

Table 2 shows the results of the LR no-identification test for the Fama and 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. It shows that we can not reject no-identification with 5% significance. For the specification which does not include the zero-β return, the pseudo-true 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.

Table 2: LR test of no-identification for Fama and French (1993) with R_m, 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

Supplemental Appendix SA6 contains additional empirical applications for the models previously presented in Figure 1 and Table 1. Unlike the Fama and French (1993) three-factor model, we can not reject no-identification for any of them, since their associated $\widehat{\text{IS}}$ and $\widehat{\text{MISS}}$ statistics are all close. Supplemental Appendix SA7 contains the simulation and empirical findings based on the bounding distribution of the LR test. Consistent with our expectations, using the bounding distribution leads to conservative sizes in simulation, while the statistical conclusions remain the same as those in Table 2 when applying the fixed bounding critical value to the Fama and French (1993) model.

5.3 General covariance structure

The $\text{IS} - \text{MISS}$ larger than zero identification condition applies to general covariance structures. It is, however, for such covariance structures not possible to pin down a null distribution for the data generating process that leads to equality of IS and MISS. We therefore extend the null distribution from the KPS setting to general covariance structures and compute critical values from it.

Theorem 2 states the null distribution under $\mathrm{H}^*_{\text{non-ind}}$ and the KPS for V, so $\binom{\psi^*_J}{\psi^*_{\mu.J}} \sim \mathbf{N}(0, I_{2k})$. Under general covariance structures, $\binom{\psi^*_J}{\psi^*_{\mu.J}} \sim \mathbf{N}(0, W)$, with $W = (\Omega' \otimes \Sigma)^{-\frac{1}{2}} V (\Omega' \otimes \Sigma)^{-\frac{1}{2}\prime}$, $\Omega^{-\frac{1}{2}} = \begin{pmatrix} \omega_{\mu\mu.J}^{-\frac{1}{2}} & 0 \\ -\Omega_{JJ}^{-1}\omega_{J\mu}\omega_{\mu\mu.J}^{-\frac{1}{2}} & \Omega_{JJ}^{-\frac{1}{2}} \end{pmatrix}$. The leading element of the limit behavior in Theorem 2 then becomes:

$$
\begin{aligned}
\begin{pmatrix} \hat{C}'\hat{C} - \hat{a}'\hat{a} \\ \hat{C}'\hat{a} \\ \hat{C}'\hat{C} + \hat{a}'\hat{a} \end{pmatrix} & \underset{d}{\rightarrow} \begin{pmatrix} 0 \\ 0 \\ C'C + a'a \end{pmatrix} + \begin{pmatrix} 2\binom{C}{-a}' \\ \binom{a}{C}' \\ 2\binom{C}{a}' \end{pmatrix} \binom{\psi^*_J}{\psi^*_{\mu.J}} \\
& \sim \mathbf{N}\left(\begin{pmatrix} 0 \\ 0 \\ C'C + a'a \end{pmatrix}, \begin{pmatrix} 2\binom{C}{-a}' \\ \binom{a}{C}' \\ 2\binom{C}{a}' \end{pmatrix} W \begin{pmatrix} 2\binom{C}{-a}' \\ \binom{a}{C}' \\ 2\binom{C}{a}' \end{pmatrix}' \right).
\end{aligned} \tag{52}
$$

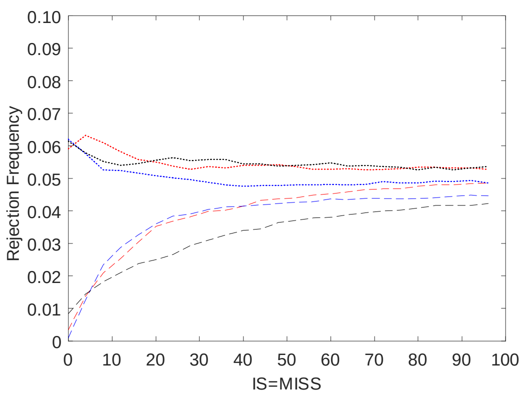
Because W does not equal the identity matrix, the resulting covariance matrix in (52) is not diagonal, so the different components are no longer approximately asymptotically uncorrelated. We therefore do not compute critical values by conditioning on rk in (38) ($= \hat{C}'\hat{C} + \hat{a}'\hat{a}$)), but by directly simulating from the joint distribution of the first two components in (52):

$$
\begin{pmatrix} \hat{C}'\hat{C} - \hat{a}'\hat{a} \\ \hat{C}'\hat{a} \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2\binom{C}{-a}' \\ \binom{a}{C}' \end{pmatrix} W \begin{pmatrix} 2\binom{C}{-a}' \\ \binom{a}{C}' \end{pmatrix}' \right), \tag{53}
$$

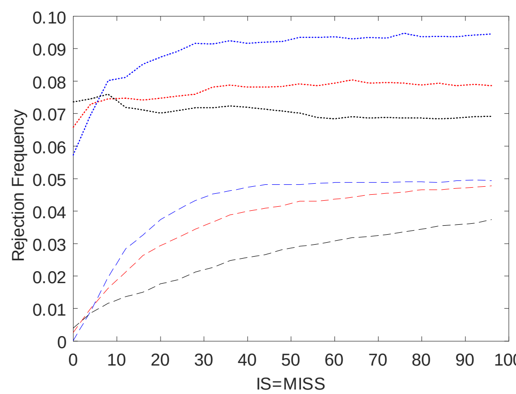
using $\hat{C}$ and $\hat{a}$ as estimators of C and a and a heteroskedasticity and autocorrelation consistent covariance matrix estimator for V.

Figure 8 illustrates the rejection frequencies of testing $\mathrm{H}^*_{\text{non-ind}}$ with the LR test using either the KPS conditional critical values or the critical values that allow for general covariance structures. Panel 8.1 shows the results for a DGP with a KPS covariance matrix, where both critical values (approximately) control the size of the test, and the KPS conditional critical values lead to rejection frequencies (much) closer to the nominal 5% level especially for smaller values of IS = MISS. The general covariance matrix critical values then lead to a somewhat conservative test in Panel 8.1. In contrast, Panel 8.2 uses a non-KPS covariance matrix DGP, for which the KPS conditional critical values lead to size distortion, while the general covariance matrix critical values lead to a size-correct no-identification LR test which is conservative for smaller values of IS = MISS.

Figure 8: Rejection frequencies of testing $\mathrm{H}^*_{\text{non-ind}}$ at the 5% significance level by the no-identification LR test using either the KPS conditional critical values (dotted), or the critical values that allow for general covariance structures (dashed) for $k = 3$ (blue), 10 (red) and 25 (black).



Panel 8.1. KPS covariance matrix



Panel 8.2. non-KPS covariance matrix

The null distribution of the no-identification LR test under general covariance structures resembles that of some widely used rank tests. For general covariance structures, the null distribution for a reduced rank value of the Jacobian is clear and can therefore be tested using the IS statistic which corresponds to the Cragg and Donald (1997) rank statistic. Given the computational difficulty of minimizing the objective function involved in the IS statistic for general covariance structures, other less computationally demanding rank tests have been developed. They provide extensions of the original, KPS covariance matrices only, Anderson (1951) rank test, which is just a function of the easy-to-compute normalized singular values of the sample Jacobian. The popular Kleibergen and Paap (2006) and Robin and Smith (2000) rank tests therefore specify the reduced rank hypothesis using the normalized singular values of the sample Jacobian and thereafter incorporate the general covariance structure in the null distribution. The null distribution of the no-identification LR test under general covariance structures is thus obtained in a similar way.

For our extensions to multiple structural parameters and incorporating the estimation error of covariance matrix estimators, we use the setting of a KPS covariance matrix, but the same arguments for those extensions can be made for the general covariance matrix setting which we, for reasons of brevity, leave for future work.

6 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. An unconditional bounding critical value function is also provided. 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% significance 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 a KPS covariance structure. We show how to extend it to more general covariance structures on which we plan to work further.

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] Anderson, T. W. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, 327–351, 1951.

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

[4] 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.

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

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

[7] Bekker, P.A. Alternative Approximations to the Distributions of Instrumental Variable Estimators. *Econometrica*, **62**:657–681, 1994.

[8] Card, David. Using geographic variation in college proximity to estimate the return to schooling. In Louis Christofides, Kenneth Grant and Robert Swidinsky, editor, *Aspects of Labour Market Behaviour: essays in honor of John Vanderkamp*, pages 201–222. University of Toronto Press, Toronto, Canada, 1995. (NBER Working Paper 4483 (1993)).

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

[10] Dufour, J.-M. Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica*, **65**:1365–1388, 1997.

[11] 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.

[12] 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.

[13] Hall, A.R. and A. Inoue. The Large Sample Behaviour of the Generalized Method of Moments Estimator in Misspecified Models. *Journal of Econometrics*, **114**:361–394, 2003.

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

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

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

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

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

[19] Imbens, G.W. and J.D. Angrist. Identification and estimation of local Average treatment effects. *Econometrica*, **62**:467–475, 1994.

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

[21] 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.

[22] Kitamura, Y. Specification Tests with Instrumental Variables and Rank Deficiency. In S. Corbae, S. Durlauf and B. Hansen, editor, *Economic Theory and Practice: Econometric Essays in Honor of Peter Phillips*, Cambridge University Press, 2005.

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

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

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

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

[27] 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.

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

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

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

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

[32] 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.

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

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

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

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

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

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

[39] Savov, A. Asset Pricing with Garbage. *Journal of Finance*, **66**:177–201, 2011.

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

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

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

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

Appendix

A1. Specifications for Figure 1 and Table 1.

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. 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.

4. 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.

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. The general notation $\mu + J\theta$ used in the main text corresponds to $\mu_R - \beta\lambda_F$ for linear asset pricing models, which is further explained below.

$$\mu + J\theta \Leftrightarrow \mu_R - \beta\lambda_F$$

Let $R_{i,t}$ be the return on the i-th asset at time t, with $i = 1, ..., k$, and $t = 1, ..., T$. Linear asset pricing models are concerned with analyzing the expected asset return:

$$\mu_R = E(R_t),$$

for $R_t = (R_{1,t}...R_{k,t})'$ and for the notation of the paper the sample size N corresponds to T so $N = T$.

The beta representation of expected returns models it as linear in the beta vector of factor loadings:

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

with λ_F the $m \times 1$ vector of risk premia, and β_i the $m \times 1$ vector of factor loadings:

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

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

$$\mu_R = \beta\lambda_F \Leftrightarrow \mu_R - \beta\lambda_F = 0$$

with $\beta = (\beta_1, ..., \beta_k)'$ so for the notation of the paper, the Jacobian J corresponds to

$-\beta : J = -\beta$, and the structural parameter θ corresponds to $\lambda_F : \theta = \lambda_F$.

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

$$\mu_R = \iota_k \lambda_0 + \beta \lambda_F,$$

with ι_k the $k \times 1$ 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 $k \times 1$ vector of asset returns, and F_t as the $m \times 1$ vector of risk factors.

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

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

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

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

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

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.

The $\widehat{\text{IS}}$ statistic results from: $\widehat{\text{IS}} = T \times \tau_{min}$, where τ_{min} is the smallest root of

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

which equals the smallest eigenvalue of the matrix $\hat{Q}_{\bar{F}\bar{F}} \hat{\beta}' \hat{\Omega}^{-1} \hat{\beta}$.

Similarly, the misspecification J-statistic ($\widehat{\text{MISS}}$) results from: J= $\widehat{\text{MISS}} = T \times \lambda_{min}$, where λ_{min} is 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 $\begin{pmatrix} 1 & 0 \\ 0 & \hat{Q}_{\bar{F}\bar{F}} \end{pmatrix} \begin{pmatrix} \bar{R} & \hat{\beta} \end{pmatrix}' \hat{\Omega}^{-1} \begin{pmatrix} \bar{R} & \hat{\beta} \end{pmatrix}$.

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

SA1. 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 $\mathrm{H}^*_{\text{non-ind}}$ (33) uses the entier function, [..], and the first and second columns of I_k indicated by $e_{1,k}$ and $e_{2,k}$ respectively:

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

  1. Generate ψ^*_J and $\psi^*_{\mu.J}$ from independent $N(0, I_k)$ distributions;
  2. Compute $\hat{a} = a + \psi^*_{\mu.J}$, $\hat{C} = C + \psi^*_J$;
  3. Compute $\mathrm{LR}(\alpha = 0) = \frac{1}{2}\left[\hat{C}'\hat{C} - \hat{a}'\hat{a} + \sqrt{\left(\hat{C}'\hat{C} - \hat{a}'\hat{a}\right)^2 + 4(\hat{C}'\hat{a})^2}\right]$;
  4. Compute conditioning statistic: rk$= \hat{C}'\hat{C} + \hat{a}'\hat{a}$ and $i = [\mathrm{rk}]$;
  5. $\mathrm{sum}_i = \mathrm{sum}_i + 1$;
  6. Set $Z(i, sum_i) = \mathrm{LR}(\alpha = 0)$;

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

SA2. Proof of Theorem 3. We specify the covariance matrix estimators $\hat{\Omega}$ and $\hat{\Sigma}$ as

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

with $\binom{u_i}{V_i}$ and Z_i realizations of $m+1$ and k dimensional random vectors that lead to consistent estimation of Ω and Σ, respectively.

In addition, we use the notation:

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

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

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:

$$\sqrt{N}\begin{pmatrix} \dot{\omega}_{\mu\mu} - 1 \\ \dot{\omega}_{J\mu} \\ \mathrm{vech}(\dot{\Omega}_{JJ} - I_m) \\ \mathrm{vech}(\dot{\Sigma} - I_k) \\ \frac{1}{N}\sum_{i=1}^{N}\dot{Z}_i\dot{u}_i \\ \frac{1}{N}\sum_{i=1}^{N}\dot{Z}_i\dot{V}_i \end{pmatrix} \underset{d}{\rightarrow} \begin{pmatrix} \psi_{\mu\mu} \\ \psi_{J\mu} \\ \psi_{JJ} \\ \psi_{\Sigma} \\ \psi_{\mu.J}^* \\ \psi_J^* \end{pmatrix} \sim N(0, V_{\Sigma}),$$

with $\psi_{\mu\mu}$, $\psi_{J\mu}$, ψ_{JJ} and ψ_{Σ} : 1, m, $\frac{1}{2}m(m+1)$ and $\frac{1}{2}k(k+1)$ dimensional mean zero normally distributed random vectors, and $\psi_{\mu.J}^*$, ψ_J^* independent normal k dimensional random vectors.

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

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

and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

which can also be expressed by

$$\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)'\operatorname{vec}(\Psi_\Sigma)+\\ &\tfrac{1}{\sqrt{N}}\left(\psi_{\mu.J}^*\otimes\psi_{\mu.J}^*-\psi_J^*\otimes\psi_J^*\right)'\operatorname{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)'\operatorname{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)'\operatorname{vec}(\Psi_\Sigma)-\\ &\tfrac{1}{\sqrt{N}}\left(\psi_{\mu.J}^*\otimes\psi_{\mu.J}^*+\psi_J^*\otimes\psi_J^*\right)'\operatorname{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}$$

These expansions therefore imply the statements in Theorem 3 and Corollary 7.

Remark: For the case that $\frac{1}{\sqrt{N}}a'a=\frac{1}{\sqrt{N}}C'C$ converging to fixed constants, the leading terms in the above three expansions of $(\hat{C}'\hat{C}-\hat{a}'\hat{a},\hat{C}'\hat{a},\hat{C}'\hat{C}+\hat{a}'\hat{a})$ are jointly normal distributed:

$$\begin{pmatrix}\frac{1}{\sqrt{N}}a'a & 0 & -\frac{1}{\sqrt{N}}a'a\\ 0 & -\frac{1}{\sqrt{N}}C'C & 0\\ -\frac{1}{\sqrt{N}}C'C & 0 & -\frac{1}{\sqrt{N}}C'C\\ \frac{1}{\sqrt{N}}(a\otimes a-C\otimes C) & \frac{1}{\sqrt{N}}(C\otimes a) & -\frac{1}{\sqrt{N}}(a\otimes a+C\otimes C)\\ -2a & C & 2C\\ 2C & a & 2a\end{pmatrix}\begin{pmatrix}\psi_{\mu\mu}\\ \psi_{J\mu}\\ \psi_{JJ}\\ \psi_\Sigma\\ \psi_{\mu.J}^*\\ \psi_J^*\end{pmatrix},$$

for which the covariance can be written as:

$$\begin{pmatrix} \frac{1}{\sqrt{N}}a'a & 0 & -\frac{1}{\sqrt{N}}a'a \\ 0 & -\frac{1}{\sqrt{N}}C'C & 0 \\ -\frac{1}{\sqrt{N}}C'C & 0 & -\frac{1}{\sqrt{N}}C'C \\ \frac{1}{\sqrt{N}}\left(a\otimes a - C\otimes C\right) & \frac{1}{\sqrt{N}}\left(C\otimes a\right) & -\frac{1}{\sqrt{N}}\left(a\otimes a + C\otimes C\right) \\ -2a & C & 2C \\ 2C & a & 2a \end{pmatrix}' V_\Sigma \begin{pmatrix} \frac{1}{\sqrt{N}}a'a & 0 & -\frac{1}{\sqrt{N}}a'a \\ 0 & -\frac{1}{\sqrt{N}}C'C & 0 \\ -\frac{1}{\sqrt{N}}C'C & 0 & -\frac{1}{\sqrt{N}}C'C \\ \frac{1}{\sqrt{N}}\left(a\otimes a - C\otimes C\right) & \frac{1}{\sqrt{N}}\left(C\otimes a\right) & -\frac{1}{\sqrt{N}}\left(a\otimes a + C\otimes C\right) \\ -2a & C & 2C \\ 2C & a & 2a \end{pmatrix}$$

For a diagonal covariance matrix V_Σ with the variances of $(\psi_{\mu\mu},\, \psi_{JJ},\, \psi_\Sigma)$ equal to three and those of $(\psi_{J\mu},\, \psi^*_{\mu.J},\, \psi^*_J)$ equal to one, the covariance matrix of $(\hat{C}'\hat{C} - \hat{a}'\hat{a}$, $\hat{C}'\hat{a}$, $\hat{C}'\hat{C} + \hat{a}'\hat{a})$ thus reads:

$$\begin{pmatrix} 4(C'C + a'a) + \frac{6((a'a)^2 + (C'C)^2)}{N} & 0 & 0 \\ 0 & C'C + a'a + \frac{(C'C)^2 + 3(C'C)(a'a)}{N} & 0 \\ 0 & 0 & 4(C'C + a'a) + \frac{6((a'a)^2 + (C'C)^2)}{N} \end{pmatrix}$$

SA3. Proof of Theorem 4. Using the specfication from (25)-(26), IS and MISS are defined by:

$$\text{IS} = \;\; \text{smallest root of the characteristic polynomial: } \left|\tau I_m - \tfrac{1}{N}C'C\right| = 0$$

and

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

To specify the hypothesis of no-identification, $\mathrm{H}_{\text{non-ind}}$: IS = MISS, more explicitly as a function of a and C, we first use the SVD of C (45) to specify:

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

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

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

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 IS = MISS, $(s_{C,mm})^2$ is the smallest eigenvalue so $U_{C,m}'a = 0$ and $b'b \geq (s_{C,mm})^2$.

SA4. Proof of Theorem 5. We use that

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

so

$$\hat{C} = \sqrt{N}\hat{\Sigma}^{-\frac{1}{2}} \hat{J}\hat{\Omega}_{JJ}^{-\frac{1}{2}}, \quad \hat{a} = \sqrt{N}\hat{\Sigma}^{-\frac{1}{2}} \left(\hat{\mu} - \hat{J}\hat{\Omega}_{JJ}^{-1}\hat{\omega}_{J\mu}\right) \hat{\omega}_{\mu\mu.J}^{-\frac{1}{2}}.$$

A SVD of $\hat{C}$ results in $\hat{C} = \hat{U}_C\hat{S}_C\hat{V}_C'$, with $\hat{U}_C = (\hat{U}_{C,1} \vdots \hat{U}_{C,m} \vdots \hat{U}_{C,2})$, $\hat{U}_{C,1} : k \times (m-1)$, $\hat{U}_{C,m} : k \times 1$, $\hat{U}_{C,2} : k \times (k-m)$, and $\hat{V}_C$ orthonormal $k \times k$ 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 $k \times m$ 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:

$$\left| \lambda I_{m+1} - (\hat{a} \vdots \hat{C})'(\hat{a} \vdots \hat{C}) \right| = 0.$$

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

$$(\hat{a} \vdots \hat{C})'(\hat{a} \vdots \hat{C}) = (\hat{a} \vdots \hat{C})'\hat{U}_C\hat{U}_C'(\hat{a} \vdots \hat{C}) = \begin{pmatrix} \hat{U}_{C,1}'\hat{a} & \hat{U}_{C,1}'\hat{C} \\ \hat{U}_{C,m}'\hat{a} & \hat{U}_{C,m}'\hat{C} \\ \hat{U}_{C,2}'\hat{a} & \hat{U}_{C,2}\hat{C} \end{pmatrix}' \begin{pmatrix} \hat{U}_{C,1}'\hat{a} & \hat{U}_{C,1}'\hat{C} \\ \hat{U}_{C,m}'\hat{a} & \hat{U}_{C,m}'\hat{C} \\ \hat{U}_{C,2}'\hat{a} & \hat{U}_{C,2}\hat{C} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & \hat{V}_C' \end{pmatrix}' \begin{pmatrix} \hat{U}_{C,1}'\hat{a} & \hat{S}_{C,1} & 0 \\ \hat{U}_{C,m}'\hat{a} & 0 & \hat{s}_{C,mm} \\ \hat{U}_{C,2}'\hat{a} & 0 & 0 \end{pmatrix}' \begin{pmatrix} \hat{U}_{C,1}'\hat{a} & \hat{S}_{C,1} & 0 \\ \hat{U}_{C,m}'\hat{a} & 0 & \hat{s}_{C,mm} \\ \hat{U}_{C,2}'\hat{a} & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \hat{V}_C' \end{pmatrix}$$

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

The characteristic roots are identical to the roots of the polynomial:

$$\left|\lambda I_{m+1} - \begin{pmatrix} \hat{a}^{*\prime}\hat{a}^* + (\hat{a}'\hat{U}_{C,m})^2 + \hat{b}^{*\prime}\hat{b}^* & \hat{a}^{*\prime}\hat{S}_{C,1} & \hat{a}'\hat{U}_{C,m}\hat{s}_{C,mm} \\ \hat{S}_{C,1}'\hat{a}^* & \hat{S}_{C,1}'\hat{S}_{C,1} & 0 \\ \hat{s}_{C,mm}\hat{U}_{C,m}'\hat{a} & 0 & \hat{s}_{C,mm}^2 \end{pmatrix}\right| = 0 \iff$$

$$\left|\lambda I_{m-1} - \hat{S}_{C,1}'\hat{S}_{C,1}\right| \left|\lambda I_2 - \begin{pmatrix} \hat{a}^{*\prime}\hat{a}^* + (\hat{a}'\hat{U}_{C,m})^2 + \hat{b}^{*\prime}\hat{b}^* & \hat{a}'\hat{U}_{C,m}\hat{s}_{C,mm} \\ \hat{s}_{C,mm}\hat{U}_{C,m}'\hat{a} & \hat{s}_{C,mm}^2 \end{pmatrix} - \right.$$
$$\left. + \begin{pmatrix} (\hat{S}_{C,1}'\hat{a}^*)' \\ 0 \end{pmatrix} \left(\lambda I_{m-1} - \hat{S}_{C,1}'\hat{S}_{C,1}\right)^{-1} \begin{pmatrix} (\hat{S}_{C,1}'\hat{a}^*)' \\ 0 \end{pmatrix}' \right| = 0 \iff$$

$$\left|\lambda I_{m-1} - \hat{S}_{C,1}'\hat{S}_{C,1}\right| \left|\lambda I_2 - \begin{pmatrix} (\hat{a}'\hat{U}_{C,m})^2 + \hat{b}^{*\prime}\hat{b}^* & \hat{a}'\hat{U}_{C,m}\hat{s}_{C,mm} \\ \hat{s}_{C,mm}\hat{U}_{C,m}'\hat{a} & \hat{s}_{C,mm}^2 \end{pmatrix} - \right.$$
$$\left. \begin{pmatrix} \hat{a}^{*\prime}\hat{a}^* - \hat{a}^{*\prime}\hat{S}_{C,1}\left(\lambda I_{m-1} - \hat{S}_{C,1}'\hat{S}_{C,1}\right)^{-1}\hat{S}_{C,1}'\hat{a}^* & 0 \\ 0 & 0 \end{pmatrix} \right| = 0 \iff$$

$$\left|\lambda I_{m-1} - \hat{S}_{C,1}'\hat{S}_{C,1}\right| \left|\lambda I_2 - \begin{pmatrix} (\hat{a}'\hat{U}_{C,m})^2 + \hat{b}^{*\prime}\hat{b}^* & \hat{a}'\hat{U}_{C,m}\hat{s}_{C,mm} \\ \hat{s}_{C,mm}\hat{U}_{C,m}'\hat{a} & \hat{s}_{C,mm}^2 \end{pmatrix} - \right.$$
$$\left. \begin{pmatrix} \hat{a}^{*\prime} \begin{pmatrix} \frac{\lambda}{(\lambda - \hat{s}_{C,11}^2)} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{\lambda}{(\lambda - \hat{s}_{C,(m-1)(m-1)}^2)} \end{pmatrix} \hat{a}^* & 0 \\ 0 & 0 \end{pmatrix} \right| = 0 \iff$$

$$\left|\lambda I_{m-1}-\hat{S}_{C,1}'\hat{S}_{C,1}\right|\left|\lambda I_2-\begin{pmatrix}(\hat{a}'\hat{U}_{C,m})^2+\hat{b}^{*\prime}\hat{b}^*-\lambda\sum_{i=1}^{m-1}\frac{\hat{a}_i^{*2}}{(\lambda-\hat{s}_{C,ii}^2)} & \hat{a}'\hat{U}_{C,m}\hat{s}_{C,mm}\\ \hat{s}_{C,mm}\hat{U}_{C,m}'\hat{a} & \hat{s}_{C,mm}^2\end{pmatrix}\right| = 0,$$

where we used that $\begin{pmatrix}1 & 0\\ 0 & \hat{V}_C'\end{pmatrix}$ is an orthonormal matrix, $\hat{a}^*=\hat{U}_{C,1}'\hat{a}$, $\hat{b}^*=\hat{U}_{C,2}'\hat{a}$,

$$\left(\lambda I_{m-1}-\hat{S}_{C,1}'\hat{S}_{C,1}\right)^{-1} = -\left(\hat{S}_{C,1}'\hat{S}_{C,1}\right)^{-1}+\begin{pmatrix}\frac{\lambda}{\hat{s}_{C,11}^2(\lambda-\hat{s}_{C,11}^2)} & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \frac{\lambda}{\hat{s}_{C,(m-1)(m-1)}^2(\lambda-\hat{s}_{C,(m-1)(m-1)}^2)}\end{pmatrix}.$$

We next write the second determinant in polynomial form:

$$\begin{aligned}
&\left|\lambda I_2-\begin{pmatrix}(\hat{a}'\hat{U}_{C,m})^2+\hat{b}^{*\prime}\hat{b}^*-\lambda\sum_{i=1}^{m-1}\frac{\hat{a}_i^{*2}}{(\lambda-\hat{s}_{C,ii}^2)} & \hat{a}'\hat{U}_{C,m}\hat{s}_{C,mm}\\ \hat{s}_{C,mm}\hat{U}_{C,m}'\hat{a} & \hat{s}_{C,mm}^2\end{pmatrix}\right| =\\
&\left(\lambda-(\hat{a}'\hat{U}_{C,m})^2-\hat{b}^{*\prime}\hat{b}^*+\lambda\sum_{i=1}^{m-1}\frac{\hat{a}_i^{*2}}{(\lambda-\hat{s}_{C,ii}^2)}\right)\left(\lambda-\hat{s}_{C,mm}^2\right)-\left(\hat{s}_{C,mm}\hat{U}_{C,m}'\hat{a}\right)^2 =\\
&\lambda^2-\lambda\left((\hat{a}'\hat{U}_{C,m})^2+\hat{b}^{*\prime}\hat{b}^*+\hat{s}_{c,mm}^2\right)+\left(\hat{b}^{*\prime}\hat{b}^*\right)\hat{s}_{c,mm}^2+\\
&\lambda\left(\lambda-\hat{s}_{C,mm}^2\right)\left(\sum_{i=1}^{m-1}\frac{\hat{a}_i^{*2}}{(\lambda-\hat{s}_{C,ii}^2)}\right) .
\end{aligned}$$

The smallest root results from the directly above part of the polynomial since $\hat{s}_{1,mm}^2>\ldots>\hat{s}_{C,(m-1)(m-1)}^2\gg\hat{s}_{C,mm}^2$ and is less than or equal to $\hat{s}_{C,mm}^2$. The approximation error of ignoring the last part of the above expression is of the order of $O(\hat{s}_{C,(m-1)(m-1)}^{-2})$ so when we solve the smallest root from the first part which results in the expression for $\widehat{\text{MISS}}$:

$$\begin{aligned}
&\widehat{\text{MISS}}\\
=\ &\tfrac{1}{2}\left[(\hat{a}'\hat{U}_{C,m})^2+\hat{b}^{*\prime}\hat{b}^*+\hat{s}_{c,mm}^2-\sqrt{\left((\hat{a}'\hat{U}_{C,m})^2+\hat{b}^{*\prime}\hat{b}^*+\hat{s}_{c,mm}^2\right)^2-4\hat{s}_{C,mm}^2\left(\hat{b}^{*\prime}\hat{b}^*\right)}\right]\\
&+O(\hat{s}_{C,(m-1)(m-1)}^{-2})
\end{aligned}$$

$$
\begin{aligned}
&= \tfrac{1}{2}\left[(\hat{a}'\hat{U}_{C,m})^2 + \hat{b}^{*\prime}\hat{b}^* + \hat{s}^2_{c,mm} - \sqrt{\left(\hat{s}^2_{c,mm} - (\hat{a}'\hat{U}_{C,m})^2 - \hat{b}^{*\prime}\hat{b}^*\right)^2 + 4\hat{s}^2_{C,mm}((\hat{a}'\hat{U}_{C,m})^2)^2}\right] \\
&\quad + O(\hat{s}^{-2}_{C,(m-1)(m-1)}) \\
&= \tfrac{1}{2}\left[\hat{a}'M_{\hat{U}_{C,1}}\hat{a} + \hat{s}^2_{c,mm} - \sqrt{\left(\hat{s}^2_{c,mm} - \hat{a}'M_{\hat{U}_{C,1}}\hat{a}\right)^2 + 4\hat{s}^2_{C,mm}((\hat{a}'\hat{U}_{C,m})^2)^2}\right] \\
&\quad + O(\hat{s}^{-2}_{C,(m-1)(m-1)}),
\end{aligned}
$$

where we used $\hat{a}'M_{\hat{U}_{C,1}}\hat{a} = (\hat{a}'\hat{U}_{C,m})^2 + \hat{b}^{*\prime}\hat{b}^*$. Therefore,

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

SA5. Expansion for computing conditional critical value function for $m > 1$.

Using the notation of the proof of Theorem 3 and the SVD of C from Theorem 4:

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

Because U_C and V_C are orthonormal $k \times k$ and $m \times m$ dimensional matrices, $\dot{u}_i$, $\ddot{V}_i$, $\ddot{Z}_i$ are standardized 1, m and k dimensional mean zero random vectors with identity covariance matrix.

The algorithm for computing the conditional distribution of the LR statistic below accounts for multiple parameters and higher-order elements:

1. Compute $\dot{a}_1 = \hat{U}_{C_1}'\hat{a} : (m-1) \times 1$, $\hat{s}_{C,11} \ldots \hat{s}_{C,(m-1)(m-1)}$ from the data under consideration.

2. Simulate N standardized independent realizations of $\dot{u}_{ii} : 1 \times 1$, $\ddot{V}_i : m \times 1$, $\ddot{Z}_i : k \times 1$.

3. Use the N standardized realizations from **2** to compute: $\ddot{\Omega}_{JJ} = \frac{1}{N}\sum_{i=1}^N \ddot{V}_i\ddot{V}_i'$, $\ddot{\omega}_{J\mu} = \frac{1}{N}\sum_{i=1}^N \ddot{V}_i\dot{u}_i$, $\dot{\omega}_{\mu\mu.J} = \frac{1}{N}\sum_{i=1}^N \dot{u}_i^2 - \ddot{\omega}_{J\mu}'\ddot{\Omega}_{JJ}^{-1}\ddot{\omega}_{J\mu}$, $\ddot{\Sigma} = \frac{1}{N}\sum_{i=1}^N \ddot{Z}_i\ddot{Z}_i'$, and $\frac{1}{\sqrt{N}}\sum_{i=1}^N \ddot{Z}_i(\dot{u}_i - \ddot{V}_i\ddot{\Omega}_{JJ}^{-1}\ddot{\omega}_{J\mu})\dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}}$, $\frac{1}{\sqrt{N}}\sum_{i=1}^N \ddot{Z}_i\ddot{V}_i'\ddot{\Omega}_{JJ}^{-\frac{1}{2}}$.

4. For a range of values of $d \ll s_{C,(m-1)(m-1)}$, set $\dot{a} = (\dot{a}_1' \vdots 0 \vdots d \vdots 0 \ldots 0)' : k \times 1$, $S_C = \text{diag}(\hat{s}_{C,11} \ldots \hat{s}_{C,(m-1)(m-1)}\ d) : k \times m$.

5. Compute $(\hat{a} \vdots \hat{C})$ from the above expressions using the generated variables over the range of values of d.

6. Compute LR$(\alpha = 0)$ and the conditioning statistic rk equal to the sum of the two squared smallest singular values of $(\hat{a} \vdots \hat{C})$.

7. Construct the conditional distribution of LR$(\alpha = 0)$ given the conditioning statistic rk.

SA6. Additional empirical applications. Like the Fama and French (1993) three-factor model whose empirical application is presented in the paper, the specifications for the models presented below are described in Appendix A1-A2.

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%,*.

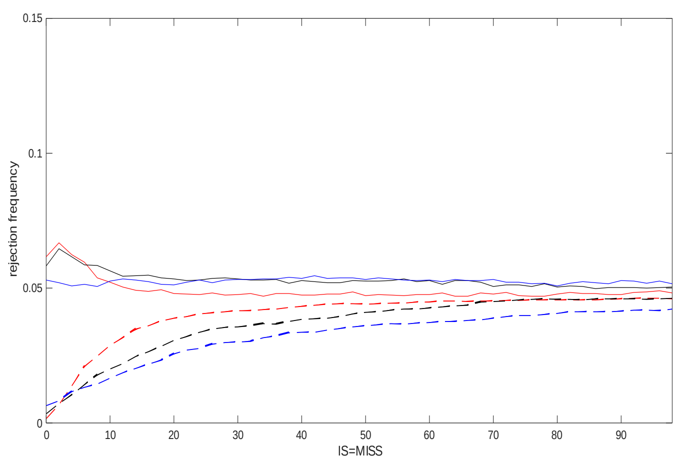
	$\lambda_0 \neq 0$	$\lambda_0 = 0$
$\widehat{\text{IS}}$	78.47^{***}	102.77^{***}
$\widehat{\text{MISS}}$	59.84^{***}	60.03^{***}
$\text{LR}(\alpha = 0) = \widehat{\text{IS}} - \widehat{\text{MISS}}$	18.63	42.74
Conditioning statistic: $\widehat{\text{MISS}} + \hat{\lambda}_{aC,m}$	160.74	226.73
95%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

SA7. Simulation and application using the bounding distribution.

Figure A1: Rejection frequencies for testing $\text{H}^*_{\text{non-ind}}$ at the 5% significance level using $\frac{\text{LR}(\alpha=0)}{\sqrt{\text{rk}}}$ and the bounding distribution for $k = 3$ (red), 10 (black) and 25 (blue). Dashed rejection frequencies result from the bounding critical value. Solid rejection frequencies result from the conditional critical value.



Equation (40) in the paper provides a bounding distribution of the standardized LR statistic, $\frac{\text{LR}(\alpha=0)}{\sqrt{\text{rk}}}$. The bounding distribution $\psi_1 + \sqrt{\psi_1^2 + \psi_2^2}$ provides an easy-to-use critical value: at the 5% significance level, the critical value is around 3.6. By using this fixed critical value, the dashed lines in Figure A1 show the simulated rejection frequencies when we conduct a 5% significance LR test of $\mathrm{H}^*_{\text{non-ind}}$. In contrast, the solid lines in Figure A1 are the counterparts that use the conditional critical value; see also Figure 2 in the paper. Because (40) provides a bounding distribution, Figure A1 shows that the rejection frequencies in dashed lines are all below 5%, but their close proximity makes the bounding distribution useful. For practitioners who do not want to simulate conditional critical values, the fixed critical value can serve as a rule of thumb.

Table A8 corresponds to Table 2 in the paper. It augments the results of the LR no-identification test for the Fama and French (1993) three-factor model, by using the bounding critical value. In addition, the rk statistic is also refined to account for higher-order elements: rk$= \widehat{\text{MISS}} + \hat{\lambda}_{aC,m} + \frac{1\frac{1}{2}}{N}\left[\left(\widehat{\text{MISS}}\right)^2 + \left(\hat{\lambda}_{aC,m}\right)^2\right]$; see also the proof in SA2. For the specification which does not include the zero-β return, Table A8 shows that using the bounding distribution for the critical value leads to the rejection of no-identification. On the other hand, no-identification is not rejected for the specification which includes the zero-β return. All these findings are consistent with those using the conditional critical values.

Table A8: LR test of no-identification for Fama and French (1993) with R_m, 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***
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
LR$(\alpha = 0)/\sqrt{\text{rk}}$	2.64	9.38
95% bounding critical value	3.6	3.6