Testing for identification in potentially misspecified linear GMM

Frank Kleibergen* Zhaoguo Zhan†

April 23, 2024

Abstract

The difference between the population analogs of the traditional Jacobian rank statistic and the J statistic provides an appropriate identification measure for the pseudo-true value of the continuous updating estimator in misspecified linear GMM whilst the population analog of the widely used traditional Jacobian rank test does not. It is by construction non-negative and if equal to zero, the pseudo-true value of the continuous updating estimator is not identified while the pseudo-true value of the two-step estimator is degenerate and the limiting distribution of its estimator non-standard. The sample analog of the identification measure provides a likelihood ratio statistic for testing no-identification. For the homoskedastic setting, we construct a conditional critical value function to obtain a conditional likelihood ratio test of no-identification. Unlike the Jacobian rank test, it provides an appropriate test of no-identification in possibly misspecified linear GMM. Applying it for empirical linear asset pricing where misspecification is common, it shows that no-identification is often not rejected at the 5% significance level.

JEL Classification: G12

Keywords: misspecification, weak identification, risk factor, asset pricing

*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, for example, 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), Stock and Wright (1999) and Dufour (1997). Inference methods have therefore been developed which remain reliable under weak identification, see *e.g.* Kleibergen (2002, 2005), Moreira (2003), Andrews and Cheng (2012) and Andrews and Mikusheva (2016). Weak identification in correctly specified GMM occurs when the Jacobian is relatively close to a reduced rank value. Tests for a reduced rank value of the Jacobian are therefore commonly employed to determine the strength of identification of the structural parameters of interest, see e.g. Cragg and Donald (1997), Kleibergen and Paap (2006) and Robin and Smith (2000).

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

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

Because the traditional rank statistic identification measure does not properly reflect identification in misspecified linear moment condition models estimated by GMM, the common practice of using its empirical analog to test for (no) identification is inappropriate because it does not test the suitable no identification hypothesis. We develop an appropriate test of (no) identification in misspecified linear moment conditions models. The involved test statistic is the sample analog of the population identification measure so it equals the difference between the sample analogs of the traditional rank statistic and the J-statistic. It coincides with the (quasi-) likelihood ratio (LR) statistic which tests if the moment vector evaluated at a zero value of the structural parameters is absent in a linear combination of the moment vector and its Jacobian evaluated at a zero value of the structural parameters. 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 build up the conditional critical value function stepwise for gradually more challenging (no) identification testing problems. We start from one structural parameter and a known covariance matrix. It allows us to specify the LR (no) identification statistic as a function of two of the three elements of an appropriate specification of the maximal invariant while its third element provides an approximately independent conditioning statistic. This is similar in spirit to the conditional likelihood ratio test of Moreira (2003) albeit that the conditioning statistic and the null distribution under which the conditional critical value function is computed differ. Hereafter, we provide empirically important generalizations which incorporate covariance matrix estimators and multiple structural parameters. Because misspecification allows for population moments which are non-zero at the pseudo-true value of the structural parameter, 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), Hansen and Lee (2021), Lee (2018) and Gospodinov et al. (2014). Alongside the dependence on the conditioning statistic, the conditional critical value function that we provide for empirical settings with multiple structural parameters therefore depends on the sample size at hand and a few consistently estimable nuisance parameters.

The remaining content of the paper is organized as follows. Section 2 introduces the appropriate identification measure for possibly misspecified over-identified linear moment equation models. Section 3 emphasizes its empirical importance showcasing eight well known studies from the asset pricing literature. Section 4 develops the conditional LR (no) identification test build up along the lines alluded to previously. We show that it has good size and power properties. Section 4 also compares the LR (no) identification test with existing tests that test part of the no identification hypotheses or just use elements of the LR (no) identification statistic. We show that while these tests can have superior power for specific settings, they are inadequate in other settings which overall renders them inappropriate for testing the no identification hypothesis. Section 4 also applies the LR (no) identification test to the Fama-French (1993) three factor model using data from Lettau, Ludvigson and Ma (2019). The misspecification J-test signals that the three factor model leads to misspecification because the J-test rejects correct specification at tiny significance levels for both the specification that incorporates the zero-beta return as well as the one without. The traditional identification rank tests indicate strong identification for both of these specifications. On the other hand, the appropriate LR (no) identification test just rejects no identification with 5% significance when the zero-beta return is not incorporated while it does not when the zero-beta return is incorporated. Given the importance of the Fama-French three factor model, it illustrates the empirical relevance of using the appropriate identification test. Section 5 discusses how to incorporate more general covariance matrix settings which is mainly left for future work because the distribution resulting from the null hypothesis involves more parameters. The sixth section draws some conclusions.

2 Identification in over-identified misspecified linear GMM

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

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

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

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

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

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

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

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

Assumption 1 is satisfied under mild conditions, see e.g. White (1984). A further assumption that we use later is the one of a Kronecker product structure (KPS) of the joint covariance matrix V.

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

$$V = (\Omega \otimes Q), \tag{4}$$

with Ω *and* Q *resp.* $(m+1) \times (m+1)$ *and* $Q : k_f \times k_f$ *dimensional matrices.*

[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.

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

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

$$\mu_f(\theta_0) = \mu_f(0) + J(0)\theta_0 = 0. \tag{5}$$

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

$$(\mu_f(0) \vdots J(0)) = D(-\theta_0 \vdots I_m), \tag{6}$$

with D a $k_f \times m$ dimensional matrix. GMM estimation methods differ with respect to how they estimate D and consequently θ. For example, the population objective functions of the abundantly used two steps methods assuming homoskedasticity, like, for example, two stage least squares and two-pass (generalized) least squares risk premia estimators, use $J(0)$ for D :

$$Q_{\text{hom},2s}(y) = \mu_f(y)'Q^{-1}\mu_f(y), \tag{7}$$

for y a m-dimensional vector, while the population objective function of the continuous updating estimator (CUE):

$$Q_{CUE}(y) = \mu_f(y)'\left[\binom{I_{k_f}}{(y\otimes I_{k_f})}'V\binom{I_{k_f}}{(y\otimes I_{k_f})}\right]^{-1}\mu_f(y), \tag{8}$$

however, (implicitly) uses

$$\text{vec}(D_{CUE}(y)) = \text{vec}(J(0)) - \binom{0}{I_{mk_f}}'V\binom{I_{k_f}}{(y\otimes I_{k_f})}\left[\binom{I_{k_f}}{(y\otimes I_{k_f})}'V\binom{I_{k_f}}{(y\otimes I_{k_f})}\right]^{-1}(\mu_f(0) + J(0)y), \tag{9}$$

with $D_{CUE}(\theta)$ a $k_f \times m$ dimensional matrix, for D, see Kleibergen (2007) and Kleibergen and Mavroeidis (2009). Under correct specification, the minimizers of (7) and (8) coincide and equal the true value since it sets the moment condition (5) to zero.

Under misspecification, this is, however, no longer the case because there is no longer a value of θ which sets the moment condition to zero. The object of interest is then the minimizer of the population objective function, the so-called, pseudo-true value, which differs over the estimation methods. Under Assumption 2, an explicit expression of the pseudo-true value results from the k-class expression[2]:

$$\begin{aligned}\theta^* &= -(J(0)'Q^{-1}J(0) - \tau\Omega_{JJ})^{-1}(J(0)'Q^{-1}\mu_f(0) - \tau\omega_{J\mu}) \\ &= -\Omega_{JJ}^{-\frac{1}{2}}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}} - \tau I_m)^{-1}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}\mu_f(0) - \tau\Omega_{JJ}^{-\frac{1}{2}\prime}\omega_{J\mu}),\end{aligned} \tag{10}$$

for $\Omega = \begin{pmatrix} \omega_{\mu\mu} & \vdots & \omega_{\mu J} \\ \omega_{J\mu} & \vdots & \Omega_{JJ} \end{pmatrix}$, with $\omega_{\mu\mu}$, $\omega_{J\mu} = \omega'_{\mu J}$, Ω_{JJ} resp. 1×1, $m \times 1$ and $m \times m$ dimensional matrices. For $\tau = 0$, the k-class expression provides the pseudo-true value of the two stage estimator while τ equal to the smallest root, $\tau_{\min}$, of the characteristic polynomial

$$\left|\tau\Omega - (\mu_f(0) \vdots J(0))'Q^{-1}(\mu_f(0) \vdots J(0))\right| = 0, \tag{11}$$

leads to the pseudo-true value of the CUE.

Under correct specification, the smallest root of (11) equals zero so the pseudo-true values of the two stage estimator and the CUE coincide. Under misspecification and a full rank value of $J(0)$, the smallest root of (11), however, differs from zero so the pseudo-true values of the two stage estimator and the CUE are no longer identical. The difference resembles the behavior of the minimizers of the sample objective functions under correct specification paired with weak identification where two stage estimators can also deviate considerably from the CUE. The difference results since under misspecification $\mu_f(0)$ is not fully spanned by $J(0)$ so $(\mu_f(0) \vdots J(0))$ is not a reduced rank matrix unless $J(0)$ is of reduced rank.

[2]For expository purposes, the expression of the k-class estimator differs from the textbook one for the linear IV regression model in which "k" corresponds with $\tau + 1$, see Hausman (1984). The pseudo-true value of the least squares estimator for the linear IV regression model therefore now results for $\tau = -1$ instead of "$k = 0$" and similary for the other estimators.

Theorem 1 states, under Assumptions 1 and 2, how two stage and continuous updating estimation procedures differ with respect to minimizing the distance between $(\mu_f(0) \vdots J(0))$ and a reduced rank matrix to obtain the respective pseudo-true value/estimator.

Theorem 1: *Under Assumptions 1 and 2, the pseudo-true value of the CUE, θ^*_{CUE}, results from a singular value decomposition (SVD) of $Q^{-\frac{1}{2}}(\mu_f(0) \vdots J(0))\Omega^{-\frac{1}{2}}$:*

$$
\begin{array}{l}
Q^{-\frac{1}{2}}\left(\mu_f(0) \vdots J(0)\right)\Omega^{-\frac{1}{2}} = \mathcal{USV}' = \\
Q^{-\frac{1}{2}} D^*_{CUE}(-\theta^*_{CUE} \vdots I_m)\Omega^{-\frac{1}{2}} + Q^{\frac{1}{2}} D^*_{CUE,\perp}\lambda^*_{CUE}\left(-\theta^*_{CUE} \vdots I_m\right)_{\perp}\Omega^{\frac{1}{2}},
\end{array}
\tag{12}
$$

with $\mathcal{U}$ a $k_f \times k_f$ dimensional orthonormal matrix, $\mathcal{V}$ a $(m+1) \times (m+1)$ dimensional orthonormal matrix, and $\mathcal{S}$ a $k_f \times (m+1)$ dimensional diagonal matrix with the singular values in decreasing order on the main diagonal:

$$
\mathcal{U} = \begin{pmatrix} \mathcal{U}_{11} & \mathcal{U}_{12} \\ \mathcal{U}_{21} & \mathcal{U}_{22} \end{pmatrix}, \ \mathcal{S} = \begin{pmatrix} \mathcal{S}_1 & 0 \\ 0 & \mathcal{S}_2 \end{pmatrix} \ \text{and} \ \mathcal{V} = \begin{pmatrix} \mathcal{V}_{11} & \mathcal{V}_{12} \\ \mathcal{V}_{21} & \mathcal{V}_{22} \end{pmatrix}, \tag{13}
$$

where $\mathcal{U}_{11}$, $\mathcal{S}_1$, $\mathcal{V}_{21}$ are $m \times m$ dimensional matrices; $\mathcal{S}_2$ is a $(k_f - m)$ dimensional vector, $\mathcal{V}'_{11}$, $\mathcal{V}_{22}$ are m dimensional vectors, $\mathcal{U}_{12}$, $\mathcal{U}_{21}$, and $\mathcal{U}_{22}$ are $m \times (k_f - m)$, $(k_f - m) \times m$ and $(k_f - m) \times (k_f - m)$ dimensional matrices and $\mathcal{V}_{12}$ is a scalar. The $k_f \times (k_f - m)$ dimensional matrix $D^*_{CUE,\perp}$ is the orthogonal complement of D^*_{CUE}, $D^{*\prime}_{CUE,\perp}D^*_{CUE} \equiv 0$, $D^{*\prime}_{CUE,\perp}QD^*_{CUE,\perp} \equiv I_{k_f-m}$; and $\left(-\theta^*_{CUE} \vdots I_m\right)_{\perp}$ is the $1 \times (m+1)$ dimensional orthogonal complement of $\left(-\theta^*_{CUE} \vdots I_m\right)$, $\left(-\theta^*_{CUE} \vdots I_m\right)\left(-\theta^*_{CUE} \vdots I_m\right)'_{\perp} \equiv 0$ and $\left(-\theta^*_{CUE} \vdots I_m\right)_{\perp}$ $\Omega\left(-\theta^*_{CUE} \vdots I_m\right)'_{\perp} \equiv 1$, so $\left(-\theta^*_{CUE} \vdots I_m\right)_{\perp} = \left[\binom{1}{\theta^*_{CUE}}'\Omega\binom{1}{\theta^*_{CUE}}\right]^{-\frac{1}{2}}\binom{1}{\theta^*_{CUE}}'$.
The specification of the pseudo-true value of the CUE as a function of the elements of the

SVD is:

$$
\begin{aligned}
\text{CUE:}\quad (\mu_f(0) \vdots J(0)) =&\ D^*_{CUE}(-\theta^*_{CUE} \vdots I_m) + QD^*_{CUE,\perp}\lambda^*_{CUE}(-\theta^*_{CUE} \vdots I_m)_\perp \Omega \\
D^*_{CUE} =&\ Q^{\frac{1}{2}}\mathcal{U}_1 S_1 \mathcal{V}'_{21}\Omega^{\frac{1}{2}}_{JJ} \\
\theta^*_{CUE} =&\ -\Omega^{-1}_{JJ}\omega_{J\mu} - \Omega^{-\frac{1}{2}}_{JJ}\mathcal{V}'^{-1}_{21}\mathcal{V}'_{11}\omega^{\frac{1}{2}}_{\mu\mu.J} = -\Omega^{-1}_{JJ}\omega_{J\mu} + \Omega^{-\frac{1}{2}}_{JJ}\mathcal{V}_{22}\mathcal{V}^{-1}_{12}\omega^{\frac{1}{2}}_{\mu\mu.J} \\
\lambda^*_{CUE} =&\ \left[D^{*\prime}_{CUE,\perp}QD^*_{CUE,\perp}\right]^{-1} D^{*\prime}_{CUE,\perp}(\mu_f(0) \vdots J(0))(\theta^*_{CUE} \vdots I_m)'_\perp \\
&\left[(\theta^*_{CUE} \vdots I_m)_\perp \Omega(\theta^*_{CUE} \vdots I_m)'_\perp\right]^{-1} \\
=&\ (\mathcal{U}_{22}\mathcal{U}'_{22})^{-\frac{1}{2}}\mathcal{U}_{22}S_2\mathcal{V}'_{12}(\mathcal{V}_{12}\mathcal{V}'_{12})^{-\frac{1}{2}}. \\
\tau_{\min} =&\ \lambda^{*\prime}_{CUE}\lambda^*_{CUE} = S'_2 S_2.
\end{aligned}
\tag{14}
$$

The expressions of the pseudo-true values of the two stage and least squares estimators, where the latter results from (10) when $\tau = -1$, as functions of the elements of the SVD read:

$$
\begin{aligned}
\text{2S:}\quad \theta^*_{2s} =&\ -\Omega^{-1}_{JJ}\omega_{J\mu} - \Omega^{-\frac{1}{2}}_{JJ}\left(\mathcal{V}_{21}\mathcal{S}'_1\mathcal{S}_1\mathcal{V}'_{21} + \tau_{\min}\mathcal{V}_{22}\mathcal{V}'_{22}\right)^{-1} \\
&(\mathcal{V}_{21}\mathcal{S}'_1\mathcal{S}_1\mathcal{V}'_{11} + \tau_{\min}\mathcal{V}_{22}\mathcal{V}'_{12})\omega^{\frac{1}{2}}_{\mu\mu.J} \\
\text{LS:}\quad \theta^*_{ls} =&\ -\Omega^{-1}_{JJ}\omega_{J\mu} - \Omega^{-\frac{1}{2}}_{JJ}(\mathcal{V}_{21}\mathcal{S}'_1\mathcal{S}_1\mathcal{V}'_{21} + \tau_{\min}\mathcal{V}_{22}\mathcal{V}'_{22} + I_m)^{-1} \\
&(\mathcal{V}_{21}\mathcal{S}'_1\mathcal{S}_1\mathcal{V}'_{11} + \tau_{\min}\mathcal{V}_{22}\mathcal{V}'_{12})\omega^{\frac{1}{2}}_{\mu\mu.J}.
\end{aligned}
\tag{15}
$$

Proof. see the Appendix, which shows the equivalence of (14) with the k-class expression in (10), and also Kleibergen and Paap (2003). ■

Equations (14) and (15) show how the pseudo-true values of the two stage estimator and CUE differ. The difference occurs since the pseudo-true value of the CUE solely results from the eigenvectors of the singular values while the dependence of the pseudo-true value of the two stage estimator is mixed and depends on a (traditional) strength of identification measure compared to a quantity reflecting the (normalized) amount of misspecification. We reflect the normalized amount of misspecification by the minimal value of the CUE population objective function while the (traditional) strength of identification is reflected by the minimal normalized distance of the Jacobian from a reduced rank value. The sample analog of the former is the J-statistic which tests for misspecification while the sample analog of the latter is the commonly used rank test for identification, see e.g. Cragg and Donald (1997), Kleibergen and Paap (2006) and Robin and Smith (2000). Because the latter results from constrained minimization of the CUE objective function, it is always larger than or equal to the former.

Theorem 2: *The amount of misspecification reflected by the minimal value of the CUE population objective function:*

$$\text{MISS} = \min_{y\in\mathbb{R}^m} Q_{CUE}(y), \tag{16}$$

is at most as large as the traditional measure of the identification strength:

$$\begin{aligned} \text{IS} =& \min_{x\in\mathbb{R}^{m-1}} Q_{IS}(x) \\ Q_{IS}(x) =& \tbinom{1}{x}' J(0)' \left[\left(\tbinom{1}{x}\otimes I_{k_f}\right)' V_{JJ} \left(\tbinom{1}{x}\otimes I_{k_f}\right)\right]^{-1} J(0)\tbinom{1}{x} \\ =& \min_{A\in\mathbb{R}^{k_f\times(m-1)}} Q_{IS}(x,A) \\ Q_{IS}(x,A) =& \left[\text{vec}\left(J(0)-A(-x \vdots I_{m-1}\right)\right]' V_{JJ}^{-1}\left[\text{vec}\left(J(0)-A(-x \vdots I_{m-1}\right)\right] \\ \text{MISS} \leq& \text{ IS}. \end{aligned} \tag{17}$$

Under Assumption 2, MISS *equals the smallest root of the characteristic polynomial (11):*

$$\text{MISS} = \tau_{\min} = \lambda_{CUE}^{*\prime}\lambda_{CUE}^{*} = S_2'S_2, \tag{18}$$

while IS *equals the smallest root of the characteristic polynomial:*

$$\left|\nu\Omega_{JJ} - J(0)'Q^{-1}J(0)\right| = 0, \tag{19}$$

which can similarly be expressed as the smallest eigenvalue of $\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}$.

Proof. see the Appendix. ■

In correctly specified linear GMM, IS is indicative for the identification of the (pseudo-) true value. This is, however, no longer so for the pseudo-true value in misspecified linear GMM. The identification of the pseudo-true value then depends on the difference between IS and MISS which is by construction always larger than or equal to zero. The identification of the pseudo-true value of the CUE breaks down when IS=MISS. Because IS results from constrained optimization of the CUE objective function of which MISS is the global uncontrained minimum, when the constrained and unconstrained minimizers are identical, the global minimum is attained at an infinite value of θ. Hence the pseudo-true value of the CUE is not identified. We use Assumption 2 to further discuss this non-identified setting.

When IS=MISS, the pseudo-true value of the CUE is not identified because the minimal value of the CUE objective function results from the Jacobian. For the KPS covariance matrix setting, this results from the specification of the moments and Jacobian as functions of the elements of the SVD is:[3]

$$\begin{aligned} \mu_f(0) = & \quad J(0)\Omega_{JJ}^{-1}\omega_{J\mu} + Q^{\frac{1}{2}}\left(\mathcal{U}_1\mathcal{S}_1\mathcal{V}_{11}' + \mathcal{U}_2\mathcal{S}_2\mathcal{V}_{12}'\right)\omega_{\mu\mu.J}^{\frac{1}{2}} \\ J(0) = & \quad Q^{\frac{1}{2}}(\mathcal{U}_1\mathcal{S}_1\mathcal{V}_{21}' + \mathcal{U}_2\mathcal{S}_2\mathcal{V}_{22}')\Omega_{JJ}^{\frac{1}{2}}. \end{aligned} \tag{20}$$

The minimal value of the CUE objective function is associated with the smallest singular value, $\mathcal{S}_2$, resulting from the SVD. When it fully results from the Jacobian, $\mathcal{V}_{22}\mathcal{V}_{22}' = 1$ which indicates that the eigenvectors associated with the smallest singular value also result from the Jacobian. Because of the orthonormality of $\mathcal{V}$, we then next have that $\mathcal{V}_{12} = 0$, $\mathcal{V}_{11}\mathcal{V}_{11}' = 1$ and $\mathcal{V}_{21}$ is a singular matrix which lies in the orthogonal complements of $\mathcal{V}_{11}$ and $\mathcal{V}_{22}$, $\mathcal{V}_{11}\mathcal{V}_{21}' = 0$, $\mathcal{V}_{21}'\mathcal{V}_{22} = 0$.[4]

Because $\mathcal{V}_{21}$ is singular and $\mathcal{V}_{12} = 0$, the pseudo-true value of the CUE resulting from

[3] These expressions are obtained in the proof of Theorem 1.

[4] Because of the orthonormality of $\mathcal{V}$, $\mathcal{V}'\mathcal{V} = \mathcal{V}\mathcal{V}' = I_{m+1}$: $\mathcal{V}_{11}'\mathcal{V}_{11} + \mathcal{V}_{21}'\mathcal{V}_{21} = I_m$, $\mathcal{V}_{12}'\mathcal{V}_{12} + \mathcal{V}_{22}'\mathcal{V}_{22} = 1$, $\mathcal{V}_{11}'\mathcal{V}_{12} + \mathcal{V}_{21}'\mathcal{V}_{22} = 0$, $\mathcal{V}_{11}\mathcal{V}_{11}' + \mathcal{V}_{12}\mathcal{V}_{12}' = 1$, $\mathcal{V}_{21}\mathcal{V}_{21}' + \mathcal{V}_{22}\mathcal{V}_{22}' = I_m$, $\mathcal{V}_{21}\mathcal{V}_{11}' + \mathcal{V}_{22}\mathcal{V}_{12}' = 0$, so when $\mathcal{V}_{22}'\mathcal{V}_{22} = 1$, $\mathcal{V}_{12} = 0$, $\mathcal{V}_{11}\mathcal{V}_{11}' = 1$ and $\mathcal{V}_{21}$ is a singular matrix which lies in the orthogonal complements of $\mathcal{V}_{11}$ and $\mathcal{V}_{22}$ because $\mathcal{V}_{21}\mathcal{V}_{21}' + \mathcal{V}_{22}\mathcal{V}_{22}' = I_m \Leftrightarrow \mathcal{V}_{21}\mathcal{V}_{21}'\mathcal{V}_{22} + \mathcal{V}_{22}\mathcal{V}_{22}'\mathcal{V}_{22} = \mathcal{V}_{22} \Leftrightarrow \mathcal{V}_{21}\mathcal{V}_{21}'\mathcal{V}_{22} + \mathcal{V}_{22} = \mathcal{V}_{22}$ (since $\mathcal{V}_{22}'\mathcal{V}_{22} = 1$) so $\mathcal{V}_{21}\mathcal{V}_{21}'\mathcal{V}_{22} = 0$ which implies that $\mathcal{V}_{21}'\mathcal{V}_{22} = 0$. Similarly, $\mathcal{V}_{11}'\mathcal{V}_{11} + \mathcal{V}_{21}'\mathcal{V}_{21} = I_m$, so $\mathcal{V}_{11}'\mathcal{V}_{11} + \mathcal{V}_{21}'\mathcal{V}_{21} = I_m$, $\mathcal{V}_{11}'\mathcal{V}_{11}\mathcal{V}_{11}' + \mathcal{V}_{21}'\mathcal{V}_{21}\mathcal{V}_{11}' = \mathcal{V}_{11}' \Leftrightarrow \mathcal{V}_{11}' + \mathcal{V}_{21}'\mathcal{V}_{21}\mathcal{V}_{11}' = \mathcal{V}_{11}' \Leftrightarrow \mathcal{V}_{21}'\mathcal{V}_{21}\mathcal{V}_{11}' = 0 \Leftrightarrow \mathcal{V}_{21}\mathcal{V}_{11}' = 0$.

(14) is not identified. The pseudo-true value of the two step estimator is, however, identified, see also Andrews (2019):

$$\begin{aligned} \text{2S:} \quad \theta_{2s}^{*} &= -\Omega_{JJ}^{-1}\omega_{J\mu} - \Omega_{JJ}^{-\frac{1}{2}}\mathcal{V}_{22,\perp}\left(B\mathcal{S}_1'\mathcal{S}_1 B'\right)^{-1} B\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}'\omega_{\mu\mu.J}^{\frac{1}{2}} \\ &= -\Omega_{JJ}^{-1}\omega_{J\mu} \qquad\qquad m=1, \end{aligned} \tag{21}$$

which results from (15) and because when IS=MISS=$\tau_{\min}$: $\mathcal{V}_{21} = (\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp})\binom{0}{B}$, with $(\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp})$ an orthonormal $m \times m$ dimensional matrix, $\mathcal{V}_{22,\perp}$ a $m \times (m-1)$ dimensional matrix orthogonal to $\mathcal{V}_{22}$ and B a $(m-1) \times m$ dimensional matrix.[5] When MISS=IS, the smallest singular value only loads on the Jacobian. The expression of the pseudo-true value of the two-stage estimator (21) shows that it is not affected by its loadings. For $m=1$, when just the smallest singular value affects the Jacobian, it does not depend on any element of the SVD and the pseudo-true value equals the bias of the two-step estimator for the correctly identified setting with IS=0 for which MISS=IS=0.

An identical pseudo-true value of the two-stage estimator for $m = 1$ results when IS>MISS and the smallest singular value only happens to load on the (recentered) moment vector so $\mathcal{V}_{12} = 1$, and consequently $\mathcal{V}_{11} = 0$, $\mathcal{V}_{22} = 0$, $\mathcal{V}_{21}\mathcal{V}_{21}' = I_m$. Using (20), we then have that the recentered moment vector, $Q^{-\frac{1}{2}}(\mu_f(0) - J(0)\Omega_{JJ}^{-1}\omega_{J\mu})$, and Jacobian, $Q^{-\frac{1}{2}}(\mu_f(0) - J(0)\Omega_{JJ}^{-1}\omega_{J\mu})$, are orthogonal[6]. The pseudo-true values of the CUE and two

[5] Because $\mathcal{V}_{21}'\mathcal{V}_{22} = 0$, we have that $(\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp})'\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{21}'(\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp}) = \binom{0}{B}\mathcal{S}_1'\mathcal{S}_1\binom{0}{B}'$ so $(\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp})'[\mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m)\mathcal{V}_{21}' + \tau_{\min}I_m](\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp}) = \binom{0}{B}\mathcal{S}_1'\mathcal{S}_1\binom{0}{B}' + \tau_{\min}[I_m - \binom{0}{B}\binom{0}{B}'] = \begin{pmatrix} \tau_{\min} & 0 \\ 0 & \tau_{\min}I_{m-1} + B(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m)B' \end{pmatrix} = \tau_{\min}I_m + \binom{0}{B}(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_{m-1})\binom{0}{B}'$, $(\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp})'\mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m) = \binom{0}{B}(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m) = \begin{pmatrix} 0 \\ B(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m) \end{pmatrix}$, so $\Omega_{JJ}^{-\frac{1}{2}}\left(\mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m)\mathcal{V}_{21}' + \tau_{\min}I_m\right)^{-1}\mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m)\mathcal{V}_{11}'\omega_{\mu\mu.J}^{\frac{1}{2}} = \Omega_{JJ}^{-\frac{1}{2}}(\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp})\begin{pmatrix} 0 \\ (\tau_{\min}I_{m-1} + B(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m)B')^{-1}B(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m) \end{pmatrix}\mathcal{V}_{11}'\omega_{\mu\mu.J}^{\frac{1}{2}}$.

[6] From (20), we have that $(\mu_f(0) - J(0)\Omega_{JJ}^{-1}\omega_{J\mu})'Q^{-1}J(0) = \omega_{\mu\mu.J}^{\frac{1}{2}}\mathcal{V}_{11}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{21}'\Omega_{JJ}^{\frac{1}{2}} + \omega_{\mu\mu.J}^{\frac{1}{2}}\mathcal{V}_{12}\mathcal{S}_2'\mathcal{S}_2\mathcal{V}_{22}'\Omega_{JJ}^{\frac{1}{2}} = 0$ because $\mathcal{V}_{11} = 0$ and $\mathcal{V}_{22} = 0$.

stage estimator (and least squares) are then all identical, see also Andrews (2019):

$$\text{CUE=2S:} \quad \theta^*_{CUE} = -\Omega^{-1}_{JJ}\omega_{J\mu} = \theta^*_{2s}. \tag{22}$$

Equations (21) and (22) show that the same pseudo-true value of the two-stage estimator can result from different settings. In (21), the value of the pseudo-true value of the two-stage estimator merely results since MISS=IS irrespective of by how much the true underlying misspecification exceeds IS. It induces the orthogonality of the recentered moment vector and Jacobian which leads to the pseudo-true value of the two stage estimator. In (22), IS>MISS and the orthogonality of recentered moment vector and the Jacobian is not a mere consequence of the strengths of misspecification and identification. These two settings leading to the same pseudo-true value of the two-stage estimator are thus very different and show that the mapping to obtain the pseudo-true value of the two stage estimator is not injective. Furthermore, as we show later, the large sample distribution of the two-stage estimator when MISS=IS is non-standard and depends on nuisance parameters.

Hence, MISS equal to IS indicates identification issues for both the pseudo-true values of the two stage estimator and the CUE. For the CUE, IS=MISS means that the pseudo-true value is not identified while for the two stage estimator it implies an a priori known pseudo-true value which is identical to one which results from orthogonality of the recentered moment vector and Jacobian when IS>MISS. Tests of identification in misspecified linear GMM thus boil down to testing the significance of the difference between IS and MISS, IS-MISS, while in correctly specified GMM, where MISS=0, it only concerns the significance of IS.

2.1 Misspecification in sample

The moment equations of every over-identified empirical model estimated by GMM are to some extend misspecified. Alongside the behavior of the pseudo-true values of different estimators in population, it is therefore important to understand the behavior of the underlying estimators in sample to see how representative the behavior of the pseudo-true value is. Theorem 3 therefore states the limiting behavior of the two-step estimator for a setting of weak misspecification and identification. Weak misspecification implies that the misspecification for the over-identified model is minor while weak identification indicates that the Jacobian is relatively small. The paired combination of weak misspecification and identification reflects, for example, the difficulty in linear IV regression models to have instruments that are both exogenous and correlated with the right hand side endogenous variable. Similarly in asset pricing, factors are often weakly correlated with asset returns while their betas do not fully span the cross-section of asset returns. This shows the empirical relevance of joint weak misspecification and identification. We model it using the commonly employed drifting sequences that result in parameters that are functions of the sample size N, see e.g. Staiger and Stock (1997).

Theorem 3. *When Assumptions 1 and 2 hold, $m = 1$ and under the weak identification and misspecification assumption:*

$$J = J(0) = \tfrac{1}{\sqrt{N}} Q^{\frac{1}{2}} C \Omega_{JJ}^{\frac{1}{2}}, \ \mu_J = \tfrac{1}{\sqrt{N}} Q^{\frac{1}{2}} a \omega_{\mu\mu.J}^{\frac{1}{2}}, \tag{23}$$

where $\mu_J = \mu(0) - J\Omega_{JJ}^{-1}\omega_{J\mu}$ and $a = \sqrt{N} Q^{-\frac{1}{2}} \mu_J \omega_{\mu\mu.J}^{-\frac{1}{2}}$ and $C = \sqrt{N} Q^{-\frac{1}{2}} J \Omega_{JJ}^{-\frac{1}{2}}$ are k-dimensional vectors of finite non-zero constants, the limiting distribution of the two-stage estimator is characterized by:

$$\hat{\theta}_{2S} - \theta_{2s}^* \underset{d}{\to} -\frac{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi_J^*)'(\psi_\mu^* + \psi_J^* \bar{\theta}_{2s})\omega_{\mu\mu.J}^{\frac{1}{2}}}{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi_J^*)'(C+\psi_J^*)\Omega_{JJ}^{\frac{1}{2}\prime}} - \frac{\Omega_{JJ}^{\frac{1}{2}\prime}\psi_J^{*\prime}(a + C\bar{\theta}_{2s})\omega_{\mu\mu.J}^{\frac{1}{2}}}{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi_J^*)'(C+\psi_J^*)\Omega_{JJ}^{\frac{1}{2}\prime}}, \tag{24}$$

with $\theta^*_{2s} = -\Omega_{JJ}^{-1}\omega_{J\mu} - \Omega_{JJ}^{-\frac{1}{2}}\bar{\theta}_{2s}\omega^{\frac{1}{2}}_{\mu\mu.J}$, $\bar{\theta}_{2s} = -\frac{C'a}{C'C}$, ψ^*_J *and* ψ^*_μ *k-dimensional independent standard normal random vectors, so when* $C'a = 0$, *we have* $\bar{\theta}_{2s} = 0$ *and:*

$$\hat{\theta}_{2S} + \Omega_{JJ}^{-1}\omega_{J\mu} \underset{d}{\rightarrow} -\frac{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi^*_J)'\psi^*_\mu\omega^{\frac{1}{2}}_{\mu\mu.J}}{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi^*_J)'(C+\psi^*_J)\Omega_{JJ}^{\frac{1}{2}}} - \frac{\Omega_{JJ}^{\frac{1}{2}\prime}\psi^{*\prime}_J a\omega^{\frac{1}{2}}_{\mu\mu.J}}{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi^*_J)'(C+\psi^*_J)\Omega_{JJ}^{\frac{1}{2}}}. \tag{25}$$

Proof. see the Appendix. ■

The limiting distribution of the two-stage estimator in (24) shows its behavior in deviation of the pseudo-true value, θ^*_{2s}. It consists of two components. The first component shows the behavior resulting from the weak identification while, because $a + C\bar{\theta}_{2s} = a - C(C'C)^{-1}C'a$, the second component shows the component implied by the weak misspecification. When the latter components exceeds the first one, the limiting behavior of the two-stage estimator primarily results from the misspecification and we cannot conduct inference using the two-stage estimator in the usual manner. When $C'a = 0$, (25) then shows that the conditions for standard/non-standard inference are in line with the identification conditions for the pseudo-true value of the CUE:

$$\begin{aligned} \text{MISS=IS=}C'C &\implies a'a \geq C'C,\ C'a = 0, \\ \text{IS} > \text{MISS} &\implies C'C > (a + C\bar{\theta}_{2s})'(a + C\bar{\theta}_{2s}). \end{aligned} \tag{26}$$

It shows that when IS equals MISS, the limiting behavior of the t-statistic associated with the two-stage estimator is non-standard and we cannot conduct inference using it. On the other hand when IS strongly exceeds MISS, inference using the two stage estimator t-statistic is feasible for which we do have to correct for the misspecification present when constructing standard errors, see e.g. Hansen and Lee (2021), Lee (2018) and Kan, Robotti and Shanken (2013). This all shows that the difference between IS and MISS is as indicative for interpreting the behavior of the two-stage estimator as it is for the (pseudo-true value of the) CUE. We therefore next provide a generalization of the traditional tests for identification in correctly specified GMM, i.e. the sample analog of IS, towards tests for identification in possibly misspecified GMM based on the difference between IS and MISS, IS−MISS.

3 Empirical importance of IS−MISS identification measure

To show the empirical relevance of using IS−MISS to signal identification issues instead of just IS, Figure 1 shows a scatter plot of MISS (=J) and IS statistics for eight well known specifications of linear asset pricing models: Fama and French (1993), Jagannathan and Wang (1996), Yogo (2006), Lettau and Ludvigson (2001), Savov (2011), Adrian, Etula, and Muir (2014), Kroencke (2017) and He, Kelly, and Manela (2017).[7] In line with common practice, we incorporate a, so-called, zero-beta return while the factors and test assets used in the eight different specifications are:

1. Fama and French (1993), the prominent three, so-called Fama-French, factors: the market return R_m, SMB (small minus big), and HML (high minus low). We use quarterly data from Lettau, Ludvigson, and Ma (2019) over 1963Q3 to 2013Q4, so $T = 202$, for the three factors, and the twenty-five size and book-to-market sorted portfolios as test assets.

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

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

[7] 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).

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

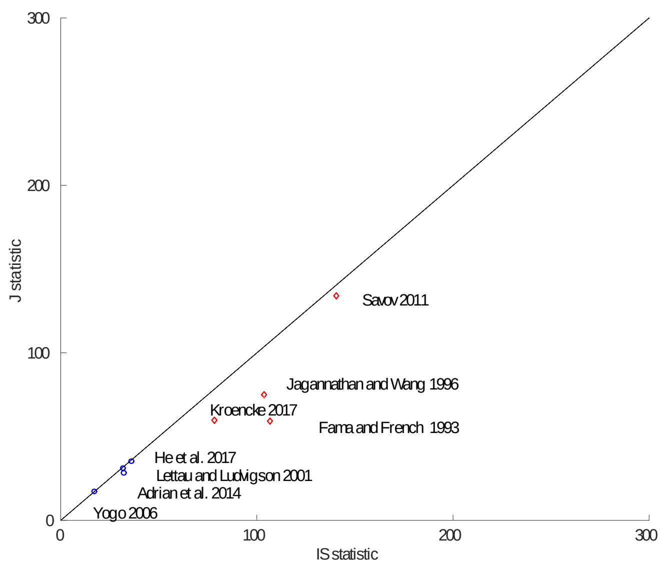
5. Savov (2011), one factor: garbage growth. We use the same annual data, 1960 - 2006, while the test assets are the twenty-five Fama-French portfolios augmented by the ten industry portfolios, as suggested by Lewellen, Nagel, and Shanken (2010).

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

7. Kroencke (2017), one factor: unfiltered annual consumption growth. We use the post-war 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.

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



Notes: Figure 1 shows MISS and IS statistics for eight specifications of linear asset pricing models. Their associated factors are: Fama and French (1993): market, SMB, and HML; Jagannathan and Wang (1996): market, corporate bond yield spread, and per capita labor income growth; Yogo (2006): market, durable and nondurable consumption growth; Lettau and Ludvigson (2001): consumption growth, (lagged) consumption wealth ratio and their interaction; 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.

Because MISS<IS, all scatter points in Figure 1 are below the 45-degree line but their proximity to it is striking. For all points, the distance to the 45-degree line, which equals the IS−MISS identification measure, is much smaller than the value of IS. It shows that IS overstates the identification strength so it is important to use the IS−MISS identification measure instead.

Table 1 shows the values of the MISS (=J) and IS statistics for specifications that incorporate or do no incorporate the zero-beta return, λ_0. Removing it adds to the misspecification MISS but increases the traditional measure of identification IS so the net total effect on the appropriate misspecification measure IS−MISS varies. For some specifications, like, for example, Fama and French (1993) and to a lesser extent Kroencke (2017), the net effect on IS−MISS is substantial but for most others the effect is rather minor which indicates that identification has not improved by removing the zero-beta return.

Table 1: **MISS and IS statistics**

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

	(A) Impose $\lambda_0 = 0$: No		(B) Impose $\lambda_0 = 0$: Yes	
	MISS	IS	MISS	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***

4 Identification testing in weakly identified and misspecified linear GMM

In correctly specified linear GMM, the reliability of traditional inference methods depends on IS which tests identification of the true value. The sample analog of IS is therefore widely used to test for it, see e.g. Cragg and Donald (1997), Kleibergen and Paap (2006) and Robin and Smith (2000). In potentially misspecified linear GMM, it is, however, the difference between IS and MISS that is indicative for identification of the pseudo-true value of the parameter of interest. Figure 1 shows that IS then often overstates the identification strength for well known empirical studies when compared to IS−MISS. It is therefore important to be able to test the hypothesis of no identification in potentially misspecified linear GMM based on IS−MISS instead of just IS. We therefore next develop tests of no identification using IS−MISS. For expository purposes, we do so for increasingly demanding settings of the number of elements m of θ, the strength of misspecification and the specification of the covariance matrix V.

4.1 Homoskedasticity, known covariance, $m = 1$

When IS=MISS, the pseudo-true value θ^*_{CUE} is not identified in misspecified GMM. To cast no-identification into a testable hypothesis for $m = 1$, we use that, see Kleibergen (2007):

$$\text{IS} = \lim\nolimits_{\theta \to \infty} Q_{CUE}(\theta). \tag{27}$$

Under homoskedasticity, the CUE objective function corresponds with (twice) the concentrated log-likelihood that results under normally distributed errors in linear models for $\mu(0)$ and $J(0)$. The sample analog of IS−MISS therefore equals the likelihood ratio (LR) statistic for testing the hypothesis of an infinite value of θ. Because elements of the different test statistics might become ill defined for an infinite value of the parameter of interest, we use the invariance of the CUE population objective function to frame the hypothesis of an infinite value of θ into one of a zero value of the parameter α in an altered moment equation:

$$\left.\begin{array}{c} \mu_f(\theta) = \mu(0) + J(0)\theta \\ \text{H}_0 : \theta = \pm\infty \end{array}\right\} \Leftrightarrow \left\{\begin{array}{c} \eta_f(\alpha) = J(0) + \mu(0)\alpha \\ \text{H}_0 : \alpha = 0 \end{array}\right. \tag{28}$$

with the accompanying CUE population objective function:

$$Q_{CUE}(a) = \eta_f(a)' \left[\left(\begin{smallmatrix} a \otimes I_{k_f} \\ I_{k_f} \end{smallmatrix}\right)' V \left(\begin{smallmatrix} a \otimes I_{k_f} \\ I_{k_f} \end{smallmatrix}\right) \right]^{-1} \eta_f(a), \quad \text{so IS} = Q_{CUE}(a = 0). \tag{29}$$

Hence, the difference between the sample analogs of IS and MISS equals the LR statistic for testing $\mathrm{H}_0 : \alpha = 0$:

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

with $\widehat{\mathrm{IS}}$, $\widehat{\mathrm{MISS}}$ and $\hat{Q}_{CUE}(a)$ the sample analogs of IS, MISS and $Q_{CUE}(a)$ resp..

In correctly specified homoskedastic linear GMM with $m = 1$, the (asymptotic) conditional critical value function for the LR statistic testing a point null hypothesis on the structural parameter against a two-sided alternative, is established by Moreira (2003). Andrews, Moreira and Stock (2006) show that the resulting conditional LR test is optimal. We, however, use the LR statistic for testing the equality of IS and MISS so we establish a conditional critical value function for it under IS=MISS. We do so using the specifications in Theorem 3 and Equation (26) and by respecifying the limit behavior from Assumption 1 into the independent components used for testing $\mathrm{H}_0 : \alpha = 0$ by means of weak identification robust statistics, see e.g. Kleibergen (2005, 2007).

Assumption 1*: *Assumptions 1 and 2 imply that the joint limit behavior of* $\hat{\mu}_J(0) = \hat{\mu}(0) - \hat{J}(0)\Omega_{JJ}^{-1}\Omega_{J\mu}$ *and* $\hat{J}(0)$ *is characterized by:*

$$\begin{aligned} \sqrt{N}\begin{pmatrix} \hat{\mu}_J(0) - \mu_J \\ \mathrm{vec}(\hat{J}(0) - J) \end{pmatrix} & \underset{d}{\rightarrow} \begin{pmatrix} \psi_{\mu.J} \\ \psi_J \end{pmatrix}, \\ \begin{pmatrix} \psi_{\mu.J} \\ \psi_J \end{pmatrix} & \sim N(0, \mathrm{diag}(\omega_{\mu\mu.J}, \Omega_{JJ}) \otimes Q), \end{aligned} \tag{31}$$

with $\psi_{\mu.J} = \psi_\mu - \psi_J\Omega_{JJ}^{-1}\Omega_{J\mu}$, $\omega_{\mu\mu.J} = \Omega_{\mu\mu} - \omega_{\mu J}\Omega_{JJ}^{-1}\omega_{J\mu}$.

Instead of using the asymptotically independent components from Assumption 1* to construct the limiting distribution of LR($\alpha = 0$) under our null hypothesis of no identification:

$$\text{H}_0 : \text{IS=MISS} \Leftrightarrow \text{H}_0 : C'a = 0,\ a'a \geq C'C, \tag{32}$$

we do so instead for the boundary setting:

$$\text{H}_0^* : C'C - a'a = 0,\ C'a = 0. \tag{33}$$

The boundary setting similarly implies no identification of the pseudo-true value. Theorem 4 states the limiting distribution of LR($\alpha = 0$) under H_0^* for known values of Ω and Q.

Theorem 4: *When Assumptions 1, 2, $m = 1$ and the weak identification and misspecification assumptions from Theorem 3 and H_0^* (33) hold:*

$$\begin{aligned}
&\text{LR}(\alpha = 0) = \tfrac{1}{2}\left[\hat{C}'\hat{C} - \hat{a}'\hat{a} + \sqrt{\left(\hat{C}'\hat{C} - \hat{a}'\hat{a}\right)^2 + 4(\hat{C}'\hat{a})^2}\right] \\
&\underset{d}{\rightarrow} \tfrac{1}{2}\left[(\psi_J^* + C)'(\psi_J^* + C) - (\psi_{\mu.J}^* + a)'(\psi_{\mu.J}^* + a) + \right. \\
&\quad \left.\sqrt{\left((\psi_J^* + C)'(\psi_J^* + C) - (\psi_{\mu.J}^* + a)'(\psi_{\mu.J}^* + a)\right)^2 + 4\left((\psi_J^* + C)'(\psi_{\mu.J}^* + a)\right)^2}\right] \\
&\underset{d}{\rightarrow} \tfrac{1}{2}\left(\psi_J^{*\prime}\psi_J^* - \psi_{\mu.J}^{*\prime}\psi_{\mu.J}^*\right) + C'\psi_J^* - a'\psi_{\mu.J}^* + \\
&\quad \tfrac{1}{2}\sqrt{\left(\psi_J^{*\prime}\psi_J^* + 2C'\psi_J^* - \psi_{\mu.J}^{*\prime}\psi_{\mu.J}^* - 2a'\psi_{\mu.J}^*\right)^2 + 4\left(\psi_J^{*\prime}\psi_{\mu.J}^* + a'\psi_J^* + C'\psi_{\mu.J}^*\right)^2},
\end{aligned} \tag{34}$$

with $\hat{a} = \sqrt{N}Q^{-\frac{1}{2}}\hat{\mu}_J(0)\omega_{\mu\mu.J}^{-\frac{1}{2}}$, $\hat{C} = \sqrt{N}Q^{-\frac{1}{2}}\hat{J}(0)\Omega_{JJ}^{-\frac{1}{2}}$, $C'a = 0$, $a'a = C'C$, $\psi_J^* = Q^{-\frac{1}{2}}\psi_J\Omega_{JJ}^{-\frac{1}{2}} \sim N(0, I_{k_f})$, $\psi_{\mu.J}^* = Q^{-\frac{1}{2}}\psi_{\mu.J}\omega_{\mu\mu.J}^{-\frac{1}{2}} \sim N(0, I_{k_f})$ *and independent of* ψ_J^*.

Proof. The first expression results from solving for the smallest root of the (quadratic) characteristic polynomial, see Moreira (2003). For the remaining part, see the Appendix. ■

The functional expression of LR($\alpha = 0$) (34) shows that it just consists of sample analogs of the discrepancy of both elements of our hypothesis of interest (33). The limiting distribution of the LR statistic in Theorem 4 is for the setting of joint weak identification and misspecification stated in Theorem 3 with known values of the covariances Ω and Q. The limiting distribution differs when incorporating the estimation error resulting from estimating these covariances which we discuss later.

The LR statistic is an invariant statistic and therefore, as all invariant statistics are, a function of the maximal invariant which equals the normalized quadratic form of $(\hat{\mu}_J(0) \vdots \hat{J}(0))$, see Andrews et. al. (2006):

$$\text{MAXINV} = N \times \Omega^{-\frac{1}{2}\prime}(\hat{\mu} \vdots \hat{J}(0))' Q^{-1}(\hat{\mu} \vdots \hat{J}(0))\Omega^{-\frac{1}{2}} = (\hat{a} \vdots \hat{C})'(\hat{a} \vdots \hat{C}) = \begin{pmatrix} \hat{a}'\hat{a} & \vdots & \hat{a}'\hat{C} \\ \hat{C}'\hat{a} & \vdots & \hat{C}'\hat{C} \end{pmatrix}. \tag{35}$$

The limiting distribution of the maximal invariant only depends on three population parameters: $a'a$, $C'C$ and $C'a$. In Moreira (2003) and Andrews et. al. (2006), the LR statistic is used to test a null hypothesis which pins down two of the three parameters on which the maximal invariant depends, $C'C = 0$, $a'C = 0$, while a sufficient statistic exists for the remaining parameter, $a'a$, which is also asymptotically independent of the other components of the maximal invariant. They can therefore construct the conditional distribution of the LR statistic given the realized value of this sufficient statistic from which the conditional critical value function for the LR test of their hypothesis of interest results.

We construct a conditional critical value function for our LR statistic for the boundary non-identified setting (33) which equates two of the three parameters of the maximal invariant: $a'a = C'C$, and has the remaining one equal to zero, $a'C = 0$. To obtain a convenient conditioning statistic, we transform the components of the maximal invariant to:

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

which similarly represents the maximal invariant because any invertible transformation of the maximal invariant is also a maximal invariant. The reasons for the transformation are two fold:

1. $\text{LR}(\alpha = 0)$ in (34) is only a function of the first two elements of the maximal invariant

in (36), $\hat{C}'\hat{C}-\hat{a}'\hat{a}$, $\hat{C}'\hat{a}$, and not of the last element, $\hat{a}'\hat{a}+\hat{C}'\hat{C}$.

2. Under H_0^* (33):

$$\begin{array}{rcl} \hat{C}'\hat{C}-\hat{a}'\hat{a} & \underset{d}{\rightarrow} & C'C-a'a+2\binom{C}{-a}'\binom{\psi_J^*}{\psi_{\mu.J}^*}+\binom{\psi_J^*}{-\psi_{\mu.J}^*}'\binom{\psi_J^*}{\psi_{\mu.J}^*} \\ & = & 2\binom{C}{-a}'\binom{\psi_J^*}{\psi_{\mu.J}^*}+\binom{\psi_J^*}{-\psi_{\mu.J}^*}'\binom{\psi_J^*}{\psi_{\mu.J}^*} \\ \hat{C}'\hat{a} & \underset{d}{\rightarrow} & (\psi_J^*+C)'(\psi_{\mu.J}^*+a)=\binom{a}{C}'\binom{\psi_J^*}{\psi_{\mu.J}^*}+\psi_J^{*\prime}\psi_{\mu.J}^* \\ \hat{C}'\hat{C}+\hat{a}'\hat{a} & \underset{d}{\rightarrow} & C'C+a'a+2\binom{C}{a}'\binom{\psi_J^*}{\psi_{\mu.J}^*}+\binom{\psi_J^*}{\psi_{\mu.J}^*}'\binom{\psi_J^*}{\psi_{\mu.J}^*}, \end{array} \tag{37}$$

and $\binom{C}{-a}'\binom{C}{a}=0$, $\binom{a}{C}'\binom{C}{a}=0$, $\binom{a}{C}'\binom{C}{-a}=0$. The dependence between the different components in (37) then only results from their last elements which are uncorrelated but not independent. The first two components of MAXINV (36-37) therefore become independently distributed from the last component for larger values of $a'a=C'C$ when the sample size increases and are approximately independently distributed for smaller values of $a'a=C'C$.

The above two reasons show why

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

is a convenient conditioning statistic for computing a conditional critical value function for LR($\alpha=0$). Because of our different hypothesis of interest, this conditioning statistic is distinct from the one used for the conditional critical value function of Moreira (2003) which is also a sufficient statistic for the remaining population parameter while (38) is not.

The algorithm for computing the conditional critical value function is stated in the Appendix. By sampling $\hat{a}$ and $\hat{C}$ using $\psi_{\mu.J}^*$ and ψ_J^* for a large range of values of a and C that satisfy H_0^*, it computes the conditional distribution of LR($\alpha=0$) given rk$=\hat{a}'\hat{a}+\hat{C}'\hat{C}$. The conditional critical value function for conducting a 5% significance test of H_0^* using LR($\alpha=0$) then corresponds with the 95% percentiles of the computed conditional distribution of LR($\alpha = 0$) given rk.

The figures in Panel 2 show the computed conditional critical value functions for different numbers of moment equations and the resulting rejection frequencies when we use them for a 5% significance LR test of no identification as a function of IS=MISS. It is striking how close these rejection frequencies are to 5%. There is only some very minor overrejection for small values of IS=MISS which could be removed by further calibrating the conditional critical value function. Because of the computational ease of our algorithm and just the very small size distortions it leads to, we, for now, refrain from doing so.

Panel 2. Conditional critical value function for testing H_0^* at the 5% signficance level using LR($\alpha = 0$) as function of rk (38) and the resulting rejection frequencies for $k_f = 3$ (solid), 10 (dash-dotted) and 25 (dashed).

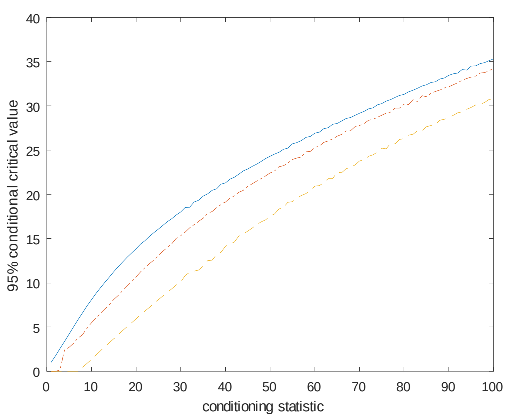


Figure 2.1. 95% conditional critical value function

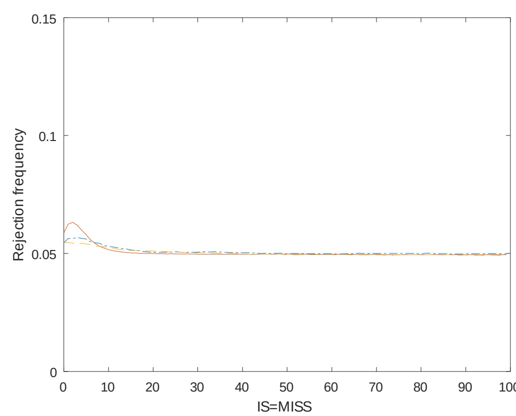


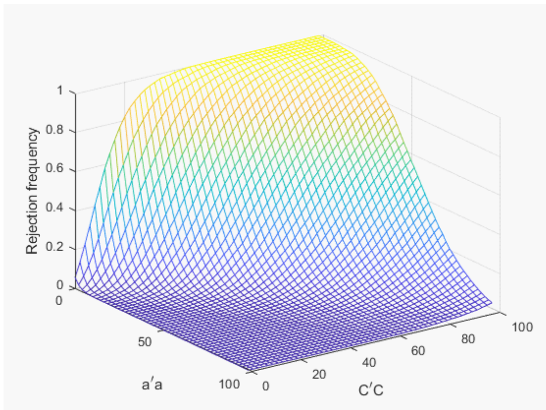
Figure 2.2. Rejection frequencies

The figures in Panel 3 show the power surface of the LR test of no identification. The conditional critical value function is calibrated to the boundary setting of no identification H_0^* (33) but our main hypothesis of interest is (32). The power surfaces in Panel 3 show the rejection frequencies with respect to violation of one of the two components in (32) while the other one is kept at the hypothesized value. Figure 3.1, and its contourlines in Figure 3.3, thus show the power surface of the LR test against violations of $C'C \leq a'a$ while $C'a = 0$, and Figure 3.2 shows it again for violations of $C'a = 0$ while $C'C = a'a$.

Figures 3.1 and 3.3 show that when $a'a$ exceeds $C'C$ that the LR test of no identification hardly rejects and rejects 5% for the boundary setting for which the conditional critical value function is computed. The latter is revealed by the contour line at 5% coinciding with the 45^0 degree line in Figure 3.3. When $C'C$ exceeds $a'a$, Figures 3.1 and 3.3 shows that the LR test has discriminatory power for rejecting these settings. The contour lines in Figure 3.3 are kind of parallel to the 45^0 degree line. On lines orthogonal to the 45^0 degree, $a'a+C'C$ is constant and so is then, approximately, the conditioning statistic. The contourlines therefore show that for constant values of $a'a+C'C$, the LR test clearly discriminates between settings for which $C'C > a'a$, for which it mostly rejects, or $C'C < a'a$, for which it does not reject.

Figure 3.2 similarly shows that the LR test has discriminatory power for detecting values of $C'a$ which differ from zero when $a'a = C'C$.

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



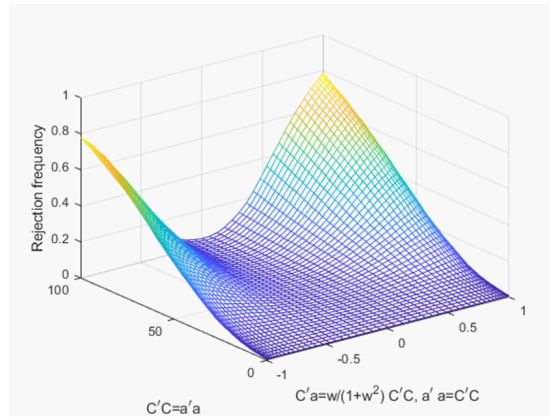


Figure 3.1

$C'a = 0$

Figure 3.2

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

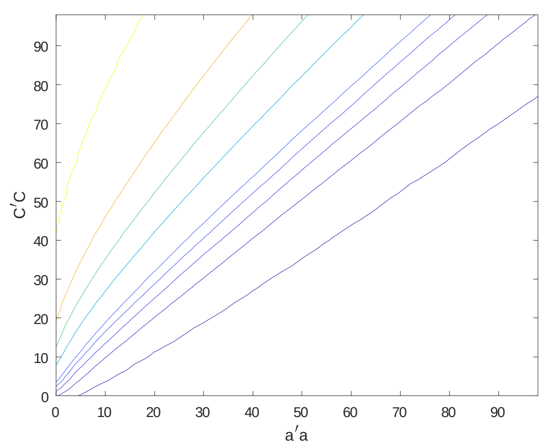


Figure 3.3. contourlines of Figure 3.1 at

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

It is interesting to compare the power of the LR no identification test with other tests that test part of the null hypothesis of no identification. Panel 4 therefore shows the power of: 5% signficance LR test of no identification, 5% IS test which equals the F-statistic for testing $J = 0$ and 5% MISS test which equals the J-statistic.

Figure 4.1 shows power 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. Figure 4.2 has an increased level of misspecification for which the power of the 5% IS test has not changed compared to Figure 4.1. Up to IS equal to 6 there is, however, no identification. The rejection frequency of the MISS test at IS=6, is 17% while that of the IS test is 50% and 5% for the LR test. The IS test therefore overstates the identification strength while the MISS test has difficulty detecting misspecification. This is further shown in Figure 4.3 where the level of misspecification has increased to 10. At MISS=10, the rejection frequency of the IS test is around 80%, while there is no identification, that of the MISS is around 45% and of the LR test is still approximately 5%. Figure 4.3 is interesting because the MISS test still mostly not rejects up to IS=10 at which the IS test rejects around 80%. The combination of these two tests therefore over states the identification strength and under states the misspecification which shows the importance of a proper test for no identification which allows for misspecification like the LR test.

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

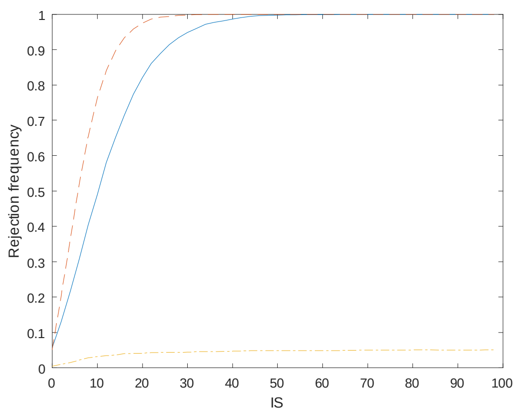


Figure 4.1. MISS=0, $C'a = 0$

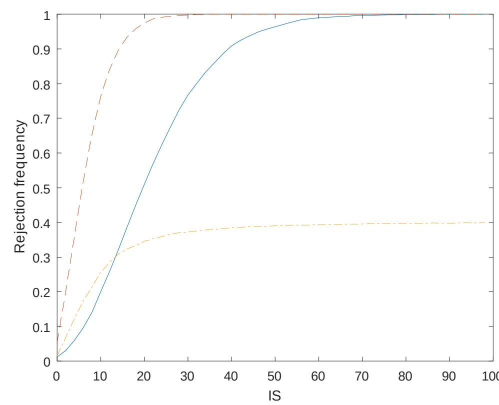


Figure 4.2. MISS=6, $C'a = 0$

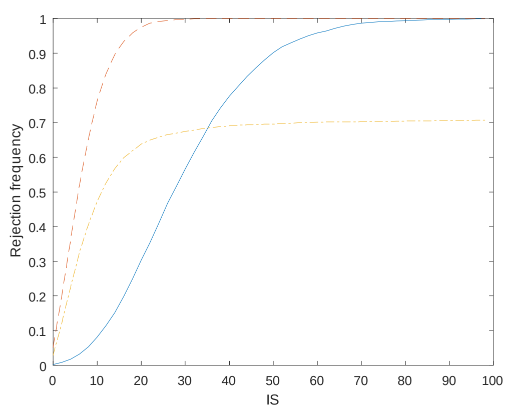


Figure 4.3. MISS=10, $C'a = 0$

Figures 5.1 and 5.2 in Panel 5 show the power surface of the DRLM test proposed in Kleibergen and Zhang (2023). It is a size correct test of $H_0 : \alpha = 0$ when using $\chi^2(m)$ critical values. Since the score of the population objective function is zero when $C'a = 0$, power and size of the DRLM test coincide in Figure 5.1 while it has good power, exceeding that of the LR test, in Figure 5.2. Similar to the IS test in Figure 4.1, it shows that for specific settings, the power of tests of just one component of the composite hypothesis H_0 (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 H_0 (32) gets violated.

Panel 5. Power of the 95% significance DRLM of $H_0 : \alpha = 0$, $k_f = 3$

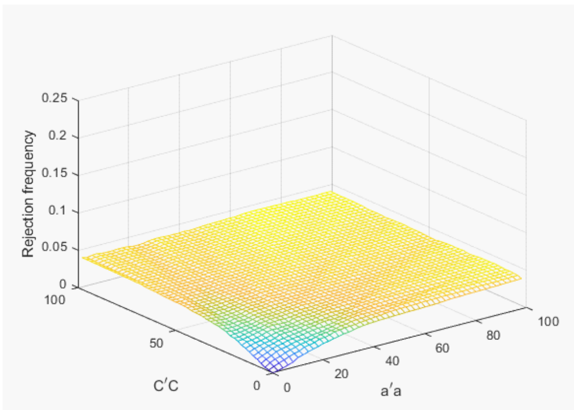


Figure 5.1

$C'a = 0$

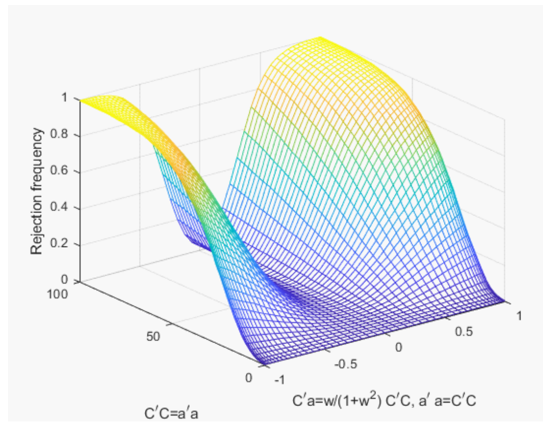


Figure 5.2

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

The figures in Panels 6 and 7 show the power surfaces of 5% significance IS and MISS tests for identical settings as for the LR test in Figures 3.1 and 3.2. They clearly show that both tests do not manage to appropriately test all components of the composite hypothesis H_0 (32).

The power surfaces in Panels 3-7 are for $k_f = 3$ but are representative for other values of the number of moment equations.

Panel 6. Power of 5% significance IS no identification test, $k_f = 3$

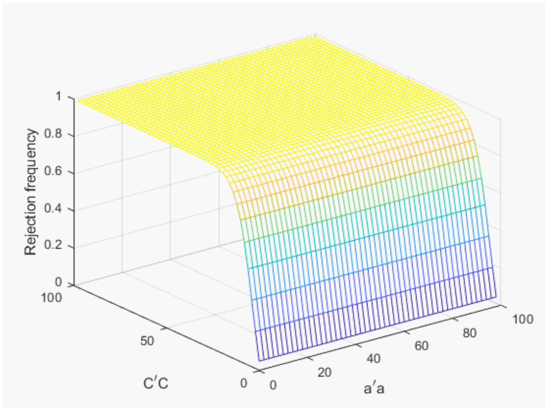


Figure 6.1

$C'a = 0$

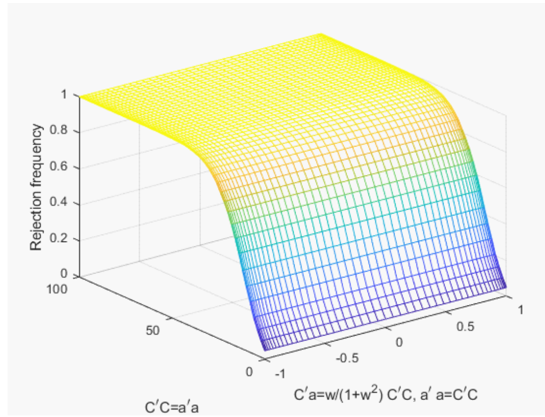


Figure 6.2

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

Panel 7. Power of 5% significance MISS misspecification test, $k_f = 3$

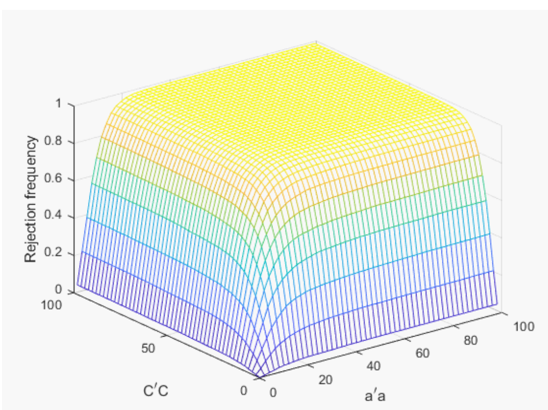


Figure 7.1

$C'a = 0$

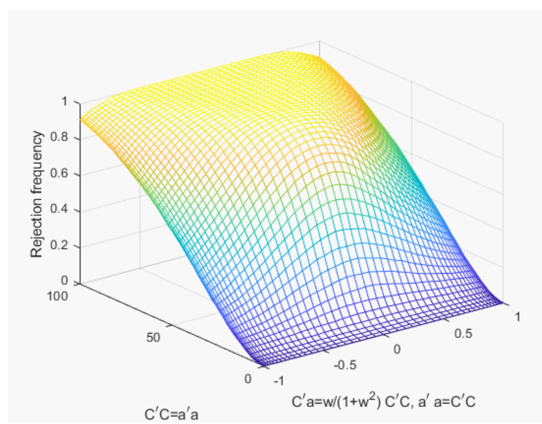


Figure 7.2

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

4.2 Homoskedasticity, unknown covariance, $m = 1$

In empirical settings, the covariance matrices Ω and Q are unknown so we use consistent estimators for them. To obtain a conditional critical value function for the LR test of no identification, we use the estimators:

$$\begin{aligned} \hat{\Omega} = & \ \frac{1}{N}\sum_{i=1}^{N}\binom{u_i}{V_i}\binom{u_i}{V_i}' = \begin{pmatrix} \hat{\omega}_{\mu\mu} & \vdots & \hat{\omega}_{\mu J} \\ \hat{\omega}_{J\mu} & \vdots & \hat{\Omega}_{JJ} \end{pmatrix} \\ \hat{Q} = & \ \frac{1}{N}\sum_{i=1}^{N} Z_i Z_i', \end{aligned} \tag{39}$$

with $\binom{u_i}{V_i}$ and Z_i iid realizations of $m+1$ and k_f dimensional random vectors with mean zero and covariance matrices Ω and Q resp.. We respecify them as quadratic forms with respect to (infeasible) normalized covariance matrix estimators using Cholesky decompositions of Ω and Q :

$$\begin{aligned} \hat{\Omega} = & \ \Omega^{\frac{1}{2}\prime}\dot{\Omega}\Omega^{\frac{1}{2}}, & \hat{\Omega}^{-1} = & \ \Omega^{-\frac{1}{2}}\dot{\Omega}^{-1}\Omega^{-\frac{1}{2}\prime} = \hat{\Omega}^{-\frac{1}{2}\prime}\hat{\Omega}^{-\frac{1}{2}}, & \hat{\Omega}^{-\frac{1}{2}} = & \ \dot{\Omega}^{-\frac{1}{2}}\Omega^{-\frac{1}{2}\prime}, \\ \hat{Q} = & \ Q^{\frac{1}{2}}\dot{Q}Q^{\frac{1}{2}\prime}, & \hat{Q}^{-1} = & \ Q^{-\frac{1}{2}\prime}\dot{Q}^{-1}Q^{-\frac{1}{2}} = \hat{Q}^{-\frac{1}{2}\prime}\hat{Q}^{-\frac{1}{2}}, & \hat{Q}^{-\frac{1}{2}} = & \ \dot{Q}^{-\frac{1}{2}}Q^{-\frac{1}{2}}, \end{aligned} \tag{40}$$

so $\dot{\Omega} = \frac{1}{N}\sum_{i=1}^{N}\binom{\dot{u}_i}{\dot{V}_i}\binom{\dot{u}_i}{\dot{V}_i}' = \begin{pmatrix} \dot{\omega}_{\mu\mu} & \vdots & \dot{\omega}_{\mu J} \\ \dot{\omega}_{J\mu} & \vdots & \dot{\Omega}_{JJ} \end{pmatrix}$, $\dot{Q} = \frac{1}{N}\sum_{i=1}^{N}\dot{Z}_i\dot{Z}_i'$, with $\binom{\dot{u}_i}{\dot{V}_i} = \Omega^{-\frac{1}{2}\prime}\binom{u_i}{V_i}$, $\dot{Z}_i = Q^{-\frac{1}{2}}Z_i$, and hence have identity covariance matrices. We use (40) to express LR($\alpha = 0$) as a function of a, C and normalized components which converge to standardized random variables.

Theorem 5: *When Assumptions 1 and 2, $m = 1$ and the weak identification and misspecification assumptions from Theorem 3 and H_0 (32) apply, LR($\alpha = 0$), which uses the covariance matrix estimators (39)-(40), can be expressed as:*

$$\mathrm{LR}(\alpha = 0) = \tfrac{1}{2}\left[\hat{C}'\hat{C} - \hat{a}'\hat{a} + \sqrt{\left(\hat{C}'\hat{C} - \hat{a}'\hat{a}\right)^2 + 4\left(\hat{C}'\hat{a}\right)^2}\right], \tag{41}$$

with

$$
\begin{aligned}
\hat{a} =& \quad \sqrt{N}\hat{Q}^{-\frac{1}{2}}\left(\hat{\mu}_f(0) - \hat{J}(0)\hat{\Omega}_{JJ}^{-1}\hat{\omega}_{J\mu}\right)\hat{\omega}_{\mu\mu.J}^{-\frac{1}{2}} \\
=& \quad \dot{Q}^{-\frac{1}{2}\prime}\left((a - C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}) + \left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\dot{Z}_i(\dot{u}_i - \dot{V}_i'\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu})\right)\right)\left(\dot{\omega}_{\mu\mu} - \dot{\omega}_{\mu J}\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}\right)^{-\frac{1}{2}} \\
\hat{C} =& \quad \sqrt{N}\hat{Q}^{-\frac{1}{2}}\hat{J}(0)\hat{\Omega}_{JJ}^{-\frac{1}{2}} \\
=& \quad \dot{Q}^{-\frac{1}{2}\prime}\left(C\dot{\Omega}_{JJ}^{-\frac{1}{2}} + \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\dot{Z}_i\dot{V}_i'\dot{\Omega}_{JJ}^{-\frac{1}{2}}\right).
\end{aligned}
\tag{42}
$$

Proof. see the Appendix. ■

The expression for LR($\alpha = 0$) in Theorem 5 shows that it under H$_0$ (32) only depends on the length of a and C, which are identical while they are also orthogonal, and the normalized components: $\dot{\omega}_{\mu\mu}$, $\dot{\omega}_{\mu J}$, $\dot{\Omega}_{JJ}$, $\dot{Q}$ and $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\dot{Z}_i\binom{\dot{u}_i}{\dot{V}_i}'$, which converge to standardized random variables when the sample size N increases. Since the population values of $\dot{\omega}_{\mu\mu}$, $\dot{\Omega}_{JJ}$, $\dot{Q}$ and $\dot{\omega}_{\mu J}$, are, resp., one, the identity matrices of dimensions 1 and k_f and zero resp., $\dot{\omega}_{\mu J}$ converges to zero at rate $\sqrt{N}$ when the sample size N grows. The limiting distribution of LR($\alpha = 0$) therefore changes compared to Theorem 4 when, for example, $C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}$ remains relevant when the sample size gets large. The weak identification and misspecification approximation of the limiting distribution of LR($\alpha = 0$) is then only valid when the length of a and C under H$_0$ (32) are of a smaller order of magnitude than the sample size so $C'C/N = a'a/N \to 0$ when the sample size grows because of which also $C\dot{\Omega}_{JJ}^{-1}\dot{\omega}_{J\mu}$ becomes negligible.

We compute the conditional critical value function using the same conditioning statistic as for the known homoskedastic covariance setting:

$$
\text{rk} = \hat{a}'\hat{a} + \hat{C}'\hat{C}. \tag{43}
$$

The conditional critical value function is computed specifically for the sample size under consideration so all elements that contribute to LR($\alpha = 0$) remain relevant. We calibrate a conditional critical value function for LR($\alpha = 0$) by simulating $\hat{a}$ and $\hat{C}$ for a range of values of the identical lengths of a and C, while they are also orthogonal, and the respective sample size N for the data under consideration. Using simulated standardized iid random variables $\dot{u}_i$, $\dot{V}_i$ and $\dot{Z}_i$ for $i = 1, \ldots, N$, we compute $\hat{a}$ and $\hat{C}$ from (42) and construct the critical value function given rk (43) using the algorithm discussed previously and stated in the Appendix.[8]

Figures 8.1 and 8.2 in Panel 8 show the 95% conditional critical value function of the conditional LR test of (32) given rk (43) for $k_f = 3$, 10, 25 and $N = 250$ and the rejection frequencies of using them for a 5% significance LR test of no identification. Compared to Figure 2.1, Figure 8.1 shows that the 95% conditional critical values are larger for higher values of the conditioning statistic. For these values, the assumption that $C'C/N = a'a/N \to 0$, which validates the critical value function in Figure 2.1, is longer accurate. The rejection frequencies in Figure 8.2 are all close to 5% which shows that the computed conditional critical value function controls the size of the test.

Because power surfaces for the estimation covariance matrix setting are very similar to those for the know covariance matrix setting, we, for reasons of brevity, do not show them.

[8] The conditional critical value function can similarly be computed by resampling the normalized data using the bootstrap. It is also possible to construct the conditional critical value function using a higher order approximation which uses the limiting distributions of the (infeasible) normalized covariance estimators. This higher order approximation equivalently depends on the sample size N.

Panel 8. Conditional critical value function for testing H_0^* at the 5% signficance level using $\mathrm{LR}(\alpha = 0)$ as function of rk (38) and the resulting rejection frequencies for $k = 3$ (solid), 10 (dash-dotted) and 25 (dashed), $N = 250$.

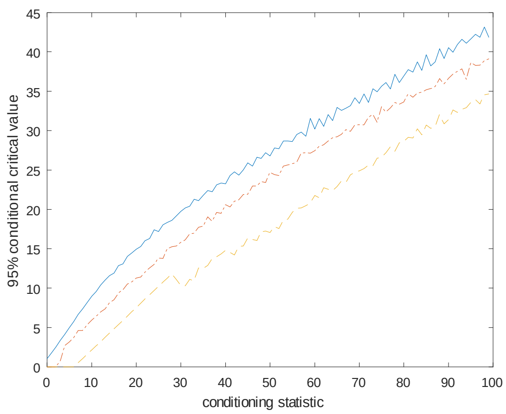


Figure 8.1. 95% conditional critical value function

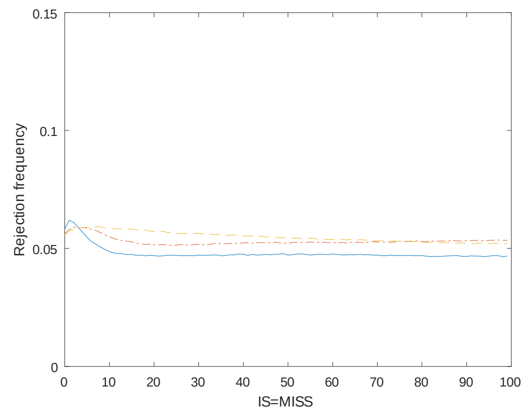


Figure 8.2. Rejection frequencies

4.3 Homoskedasticity, known covariance, $m > 1$

For $m > 1$ and homoskedasticity, which implies the KPS covariance matrix from Assumption 2, the difference between the sample analogs of IS and MISS equals the LR statistic that tests for a zero value of α in the population moment equation:

$$\eta_f(\alpha, \gamma) = J(0)_1 + \mu(0)\alpha + J(0)_2\gamma, \tag{44}$$

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

$$
\begin{aligned}
&\text{LR}(\alpha = 0) = \widehat{\text{IS}} - \widehat{\text{MISS}} = \min_{g \in \mathbb{R}^{m-1}} \hat{Q}_{CUE}(a = 0, g) - \min_{a \in \mathbb{R}} \min_{g \in \mathbb{R}^{m-1}} \hat{Q}_{CUE}(a, g), \\
&\text{for } \hat{Q}_{CUE}(a, g) = \hat{\eta}_f(a, g)' \left[\begin{pmatrix} (a \otimes I_{k_f}) \\ I_{k_f} \\ (g \otimes I_{k_f}) \end{pmatrix}' \hat{V} \begin{pmatrix} (a \otimes I_{k_f}) \\ I_{k_f} \\ (g \otimes I_{k_f}) \end{pmatrix} \right]^{-1} \hat{\eta}_f(a, g), \\
&\quad \hat{\eta}_f(\alpha, \gamma) = \hat{J}(0)_1 + \hat{\mu}(0)\alpha + \hat{J}(0)_2\gamma.
\end{aligned}
\tag{45}
$$

The moment equation in (44) is normalized using the first column of $J(0)$. The LR statistic is invariant with respect to this normalization so an identical value results when normalized using any other column of $J(0)$.

The expression of the LR statistic in (45) is identical to the subset LR statistic for testing one of multiple structural parameters in the homoskedastic linear IV regression model analyzed in Kleibergen (2021). In Kleibergen (2021), a conditional critical value function for it is constructed which makes the subset LR test size correct under weak identification of any of the structural parameters and optimal under strong identification of the partialled out endogenous variables. The null hypothesis in Kleibergen (2021) under which the conditional critical value function is constructed, however, differs from the one here. We use the specification from Theorem 3 to specify our hypothesis of interest as a function of the population parameters to construct a conditional critical value function for the conditional LR test of no identification.

Theorem 6: *Using the specification from Theorem 3, a sufficient condition for the hypothesis of no identification is:*

$$
\text{H}_0 : \text{IS}{=}\text{MISS} \Leftrightarrow \text{H}_0 : U_{C,m}' a = 0 \textit{ and } b'b \geq \text{IS} = s_{C,m}^2, \tag{46}
$$

with $b = U'_{C,2}a$ *and which results from a singular value decomposition (SVD) of* C :

$$C = U_C S_C V'_C, \tag{47}$$

where $U_C = (U_{C,1} \vdots U_{C,m} \vdots U_{C,2})$ *is an orthonormal* $k_f \times k_f$ *dimensional matrix,* $U_{C,1}$: $k_f \times (m-1)$, $U_{C,m}$: $k_f \times 1$, $U_{C,2}$: $k_f \times (k_f - m)$ *dimensional matrices,* V_C *is a* $m \times m$ *dimensional orthonormal matrix and* S_C *is a* $k_f \times m$ *dimensional matrix with the singular values,* $s_{C,1} \ldots s_{C,m}$, *in decreasing order on the main diagonal.*

Proof. see the Appendix. ■

The population value of MISS equals the smallest root of the characteristic polynomial:

$$\begin{array}{rl} \left|\lambda\Omega - N\begin{pmatrix} \mu_f(0) & J(0) \end{pmatrix}' Q^{-1} \begin{pmatrix} \mu_f(0) & J(0) \end{pmatrix}\right| & = 0 \quad \Leftrightarrow \\ \left|\lambda I_{m+1} - \begin{pmatrix} a & C \end{pmatrix}' \begin{pmatrix} a & C \end{pmatrix}\right| & = 0. \end{array} \tag{48}$$

Theorem 6 lays down the conditions for IS, which is the squared smallest singular value of C, $s^2_{C,m}$, to be identical to the smallest root of (48). The boundary condition occurs when $b'b$ equals IS and $U'_{C,m}a = 0$. It results in both the smallest and second smallest characteristic roots of (48) to be equal to IS:

$$\mathrm{H}^*_0 : \lambda_{aC,m} = \lambda_{aC,(m+1)} = \mathrm{IS}, \tag{49}$$

with $\lambda_{aC,(m+1)}$ and $\lambda_{aC,m}$ the smallest and second smallest root of (48). Compared to the boundary setting under $m = 1$ H_0^* (33), the boundary setting for m larger than one is:

Boundary non-identified setting for $m = 1$ compared to $m > 1$:

	$m = 1$	$m > 1$
H_0^* :	$C'C - a'a = 0$	$s_{C,m}^2 - b'b = 0$
	$C'a = 0$	$U_{C,m}'a = 0$
Conditioning statistic	$\hat{C}'\hat{C} + \hat{a}'\hat{a}$	$\widehat{\text{MISS}} + \hat{\lambda}_{aC,m}$
Population value conditioning statistic	$C'C + a'a$	$b'b + \lambda_{aC,m}$

(50)

The LR statistic, $\text{LR}(\alpha = 0)$, equals the difference between IS and MISS:

$$\text{LR}(\alpha = 0) = \widehat{\text{IS}} - \widehat{\text{MISS}}. \tag{51}$$

Under H_0^* and for $m = 1$, the smallest two roots of (48) are $a'a$ and $C'C$, in either order because $C'a = 0$, and they equal IS which is computed just using C, IS$=C'C$. Under H_0^* and for m larger than one, the smallest two roots of (48) are also both equal to IS. The sample analog of their sum is similarly approximately independent of the sample analog of their difference as for $\hat{C}'\hat{C} + \hat{a}'\hat{a}$ compared to $\hat{C}'\hat{C} - \hat{a}'\hat{a}$ (37) which we used for the conditioning statistic when $m = 1$. We therefore use

$$\text{rk} = \widehat{\text{MISS}} + \hat{\lambda}_{aC,m}, \tag{52}$$

with $\hat{\lambda}_{aC,m}$ the second smallest characteristic root of the sample analog of (48), which equals the squared second smallest singular value of $(\hat{a} \vdots \hat{C})$, as conditioning statistic for larger values of m.

We compute the conditional critical value function identical to the one structural parameter setting, $m = 1$. The null distribution accords with Theorem 6 where we also use large values for the second smallest to largest singular values of C. We do so because Kleibergen (2021) shows that the distribution of $\text{LR}(\alpha = 0)$ is a non-decreasing function of them. Figure 9.1 shows the 95% conditional critical value function when $m = 3$ and $k_f = 5$, 10 and 25. Figure 9.2 shows that these conditional critical values control the size of a 5% LR no identification test because the rejection frequencies are all close to 5%. It shows that the conditioning statistic (52) and the resulting conditional critical value function work well. Figure 2 uses large values for the larger singular values of C. Figure 9.3 shows the sensitivity of the LR no identification test using a conditional critical value function computed for large values of the second smallest to largest singular value of C when the actual singular values are (much) smaller. The rejection frequencies in Figure 9.3 are all below 5% so the conditional critical value function also makes the size of the LR no identification test controlled for smaller singular values of C.

Panel 9. Conditional critical value function for testing H_0^* at the 5% signficance level using $\text{LR}(\alpha = 0)$ as function of rk (38) and the resulting rejection frequencies for $k = 5$ (solid), 10 (dash-dotted) and 25 (dashed), $m = 3$.

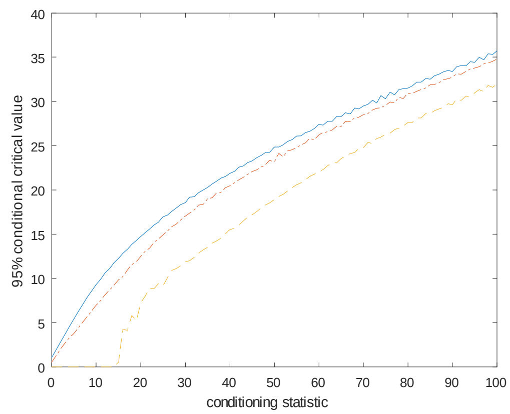


Figure 9.1. 95% conditional critical value function

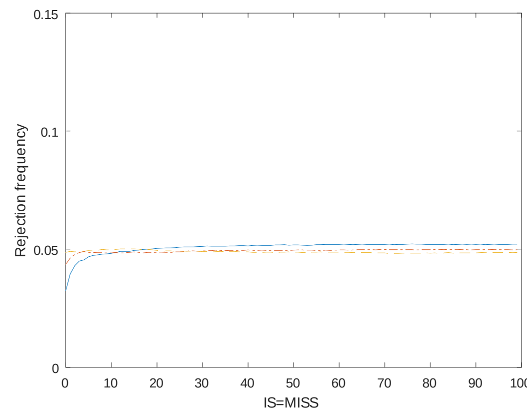


Figure 9.2. Rejection frequencies

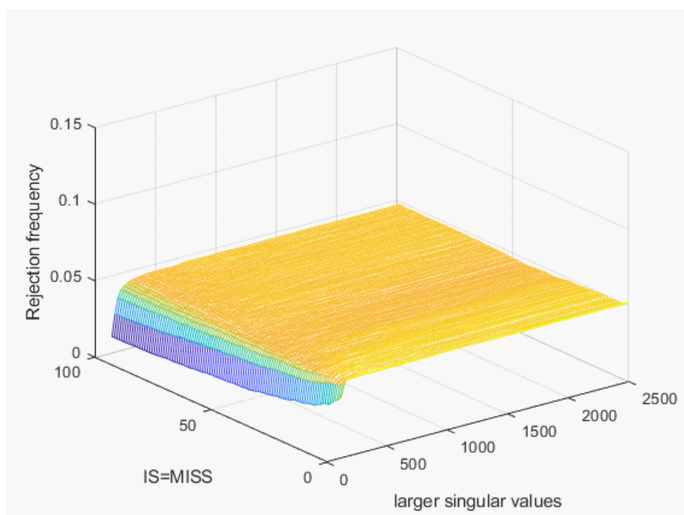


Figure 9.3. Rejection frequency surface as function of larger singular values of C, $m = 3$, $k_f = 10$.

The figures in Panel 10 show the power of a 5% significance LR no identification test. Figure 10.1 shows the power surface when we vary $b'b$ and $s_{C,m}^2$ while $U_{C,m}'a = 0$. Figure 10.2 shows the contour lines of the power surface in Figure 10.1. It shows that the 5% rejection frequency contour line coincides with the 45^0 line. Under H_0, the population value of the conditioning statistic corresponds with $b'b + s_m^2$ whose equivalue lines are orthogonal to the 45^0 line which further explains the choice of this conditioning statistic. It shows that power is (locally) maximized along the line of constant values of the conditioning statistic.

Panel 10. Power of 5% significance conditional LR no identification test, $m = 3,\ k_f = 10$

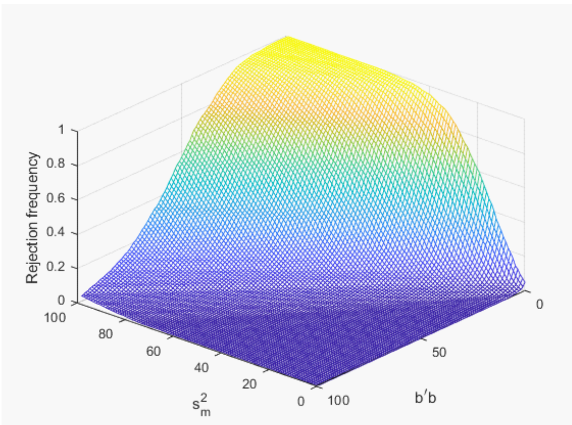


Figure 10.1

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

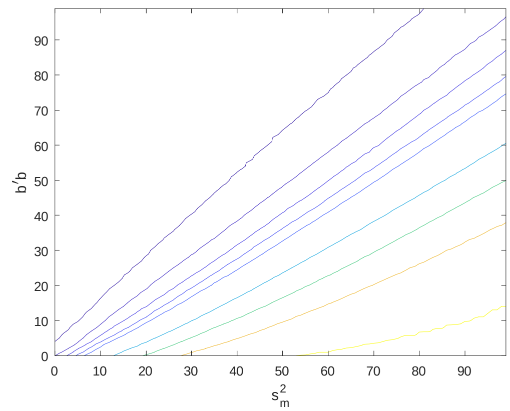


Figure 10.2

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

4.4 Homoskedasticity, unknown covariance, $m > 1$

For homoskedasticity with m larger than one, we use the covariance matrix estimators in (39) which we again specify as quadratic forms. We use (40) to express LR($\alpha = 0$) as a function of a, C and normalized components which converge to standardized random variables.

Theorem 7: *When Assumptions 1 and 2 and the weak identification and misspecification assumptions from Theorem 3 and H_0^* (50) apply, LR($\alpha = 0$), which uses the covariance matrix estimators (39)-(40), can be expressed as:*

$$\text{LR}(\alpha = 0) = \widehat{\text{IS}} - \widehat{\text{MISS}}, \tag{53}$$

where $\widehat{\text{IS}}$ *and* $\widehat{\text{MISS}}$ *are the smallest characteristic roots of resp.:*

$$\begin{aligned} \left|\lambda I_m - \hat{C}'\hat{C}\right| &= 0, \\ \left|\tau I_{m+1} - \left(\hat{a} \vdots \hat{C}\right)' \left(\hat{a} \vdots \hat{C}\right)\right| &= 0, \end{aligned} \tag{54}$$

with

$$\begin{aligned} \hat{a} = & \sqrt{N}\hat{Q}^{-\frac{1}{2}}\left(\hat{\mu}_f(0) - \hat{J}(0)\hat{\Omega}_{JJ}^{-1}\hat{\omega}_{J\mu}\right)\hat{\omega}_{\mu\mu.J}^{-\frac{1}{2}} \\ = & \ddot{Q}^{\frac{1}{2}}\left[(\dot{a} - S_C\ddot{\Omega}_{JJ}^{-1}\ddot{\omega}_{J\mu})\dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}} + \left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N \ddot{Z}_i(\dot{u}_i - \ddot{V}_i\ddot{\Omega}_{JJ}^{-1}\ddot{\omega}_{J\mu})\dot{\omega}_{\mu\mu.J}^{-\frac{1}{2}}\right)\right. \\ \hat{C} = & \sqrt{N}\hat{Q}^{-\frac{1}{2}}\hat{J}(0)\hat{\Omega}_{JJ}^{-\frac{1}{2}} \\ = & \ddot{Q}^{\frac{1}{2}}\left[S_C\ddot{\Omega}_{JJ}^{-\frac{1}{2}} + \left(\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N \ddot{Z}_i\ddot{V}_i'\ddot{\Omega}_{JJ}^{-\frac{1}{2}}\right)\right], \end{aligned} \tag{55}$$

and $\dot{a} = U_C'a$, $\ddot{Z}_i = U_C'\dot{Z}_i$, $\ddot{Q} = U_C'\dot{Q}U_C = \frac{1}{N}\sum_{i=1}^N \ddot{Z}_i\ddot{Z}_i'$, $\ddot{Q}^{-\frac{1}{2}} = \dot{Q}^{-\frac{1}{2}}U_C$, $\ddot{V}_i = V_C'\dot{V}_i$, $\ddot{\Omega}_{JJ} = V_C'\dot{\Omega}_{JJ}V_C$, *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* $\ddot{\omega}_{J\mu} = V_C'\dot{\omega}_{J\mu}$.

Proof. see the Appendix. ■

The expressions for $\hat{a}$ and $\hat{C}$ in (55) are similar to those in (42) and differ only because the normalized covariance matrix estimators involved in (55) are quadratic forms of the eigenvectors of the singular values resulting from the SVD of C in (47) with respect to the normalized covariance matrix estimators used in (42). Because these eigenvectors are all orthonormal, the population value of the normalized covariance matrix estimators in (55) remains the identity matrix.

Theorem 7 shows that the large sample behavior of LR$(\alpha = 0)$ only depends on $\dot{a}$, the m singular values of C present in S_C and the standardized random variables $\dot{\omega}_{\mu\mu.J}$, $\dot{\omega}_{J\mu}$, $\dot{\Omega}_{JJ}$ and $(\frac{1}{\sqrt{N}}\sum_{i=1}^N \ddot{Z}_i(\dot{u}_i \vdots \ddot{V}_i)')$. The boundary setting of no identification H_0^* remains identical to (50).

For the known homoskedastic covariance matrix setting, we obtained a conditional critical value function for the LR test of no identification by setting the larger singular values of C,

$(s_{C,1} \dots s_{C,m-1})$, to (very) large numbers, see Figures 9.1-10.2. The known covariance matrix setting corresponds with values of $\ddot{Q}$, $\ddot{\Omega}_{JJ}$ and $\dot{\omega}_{\mu\mu}$ equal to identity matrices and one resp., and $\ddot{\omega}_{J\mu}$ equal to zero in $\hat{a}$ and $\hat{C}$ in (55). The infeasible covariance estimators converge to these values at rate $\sqrt{N}$ so for the conditional critical values from the known covariance matrix setting to apply, the larger singular values have to be such that $S_C/\sqrt{N}$ becomes negligible and $C/\sqrt{N}$ as well. Hence also the larger singular values have to be small relative to $\sqrt{N}$ for the known covariance matrix setting to result in conditional critical values that lead to a size correct LR test of no identification.

Theorem 7 shows the spill over of the larger singular values of C present in S_C to $\hat{a}$ which did not occur for the known homoskedastic covariance matrix case. To compute the conditional critical value function for the empirical application of interest, we therefore need to estimate the second smallest to largest singular values in S_C and $U'_{C,1}a$, *i.e.* the part of a that is spanned by the eigenvectors of the second smallest to largest singular values. Hence, we compute S_C, $U'_{C,1}a$ and use the specific sample size N. Next, we generate iid mean zero, identity covariance matrix realizations of $\dot{u}_i$, $\ddot{V}_i$ and $\ddot{Z}_i$, $i = 1, \dots, N$, which we use to compute $\hat{a}$ and $\hat{C}$ for a range of values of $b'b = s^2_{C,m}$. We then compute conditional critical values using the algorithm in the Appendix given:

$$\text{rk} = \widehat{\text{MISS}} + \hat{\lambda}_{aC,m}, \tag{56}$$

with $\widehat{\text{MISS}}$ and $\hat{\lambda}_{aC,m}$, the smallest and second smallest characteristic roots of the characteristic polynomial on the bottom line of (54). These are also computed from the simulated realizations.

For the Fama-French (1993) three factor model with and without the zero-β return using data from Lettau et al. (2018), Figure 11.1 in Panel 11 shows the 95% conditional critical value functions that result when incorporating the zero-β return or not. Figure 11.1 shows that these critical value function are basically identical but differ from the one which would result when we ignore the estimation error that results from the covariance matrix estimators. Figure 11.2 in Panel 11 shows the resulting rejection frequencies which are very close to 5% for both specifications. Figure 11.2 also shows that usage of the conditional critical value function that does not incorporate the estimation error from the covariance matrix estimators leads to considerable size distortion.

Panel 11. Conditional critical value function for testing H_0^* at the 5% signficance level using LR($\alpha = 0$) as function of rk (38) calibrated to Fama-French (1993) three factor model with zero-β return (solid), without (dash-dotted), $k_f = 25,\ m = 3$, $N = 201$ (solid) and known covariance (dashed) and resulting rejection frequencies

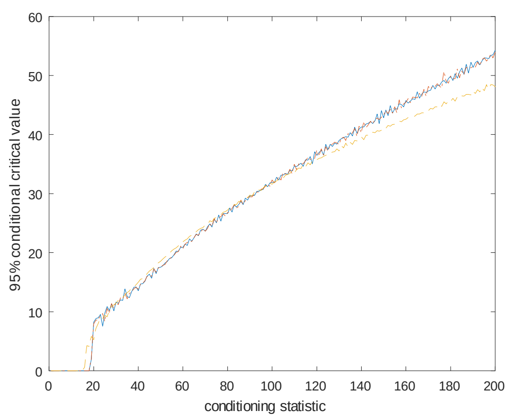


Figure 11.1. 95% conditional critical value function

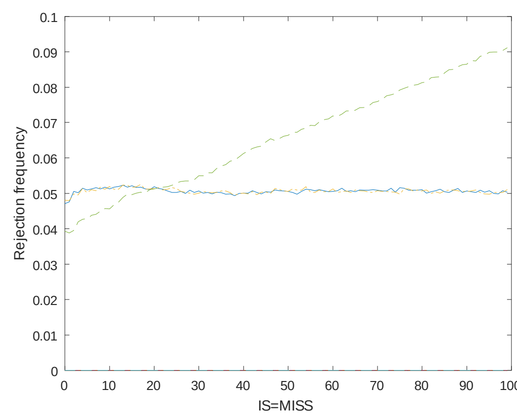


Figure 11.2. Rejection frequencies

Panel 12. Power of 5% significance conditional LR no identification test calibrated to FF93, $m = 3,\ k_f = 25,\ \lambda_0 \neq 0$

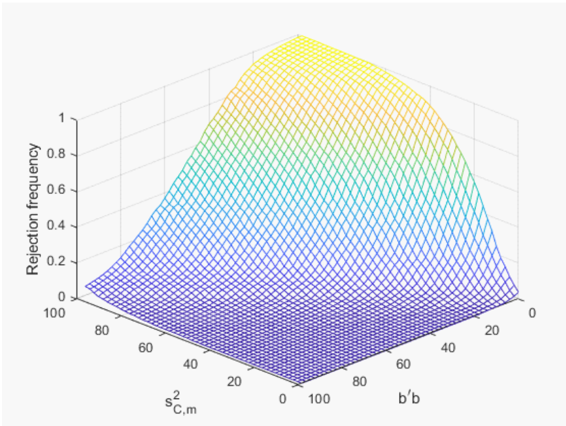


Figure 12.1

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

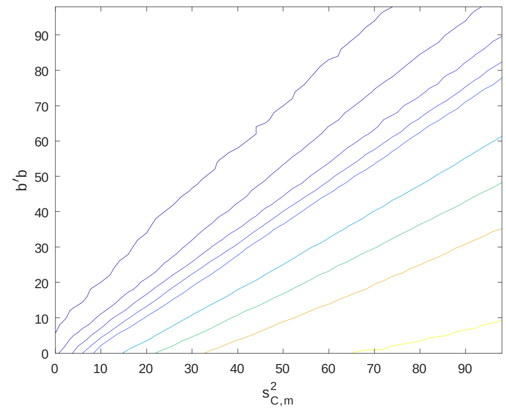


Figure 12.2

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

The figures in Panel 12 illustrate the power of the LR test of no identification calibrated to the Fama and French three factor model with the zero-β return and using the Lettau et al. (2018) data. Figure 12.1 shows the power surface over IS and MISS, while $U'_{C,m}a = 0$, and Figure 12.2 shows the accompanying contour lines. It shows that the contourline at 5% is very close to the 45^0 degree line, where IS=MISS, and the equivalue lines of the population value of the conditioning statistic are orthogonal to it.

Table 2 shows the results of the LR no identification test for the Fama-French three factor model using data from Lettau et al. (2018). Both for the specifications with and without the zero-β return, the IS and MISS statistics are strongly significant which gives the impression that the risk premia are identified in either specification while they are also misspecified. For the specification which includes the zero-β return, the proper LR no-identification test is, however, not significant at the 5% level but just at the 10% level. It shows that we can not reject no identification with 5% significance. For the specification which does not include the zero-β return, the risk premia are well identified as reflected by the very large value of the LR statistic.

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

	$\lambda_0 \neq 0$	$\lambda_0 = 0$
$\widehat{\text{IS}}$	106.8***	974.4***
$\widehat{\text{MISS}}$	59.3***	87.8***
$\text{LR}(\alpha = 0) = \widehat{\text{IS}} - \widehat{\text{MISS}}$	47.5*	887***
Conditioning statistic: $\widehat{\text{MISS}} + \hat{\lambda}_{aC,m}$	182.3	1109
95% conditional critical value	50.3	215.8

5 General covariance, $m = 1$

We next analyze how we can reflect the null hypothesis of equal values of MISS and IS as a function of the population parameters for a setting with a general covariance structure, weak identification and misspecification and one structural parameter of interest, so $m = 1$. We aim to pin down the setting of the parameters in Assumptions 1 and 1* that lead to equality of IS and MISS in order to establish a limiting distribution of the different statistics under our hypothesis of interest. No identification of θ implies that both the derivative and Hessian of the CUE population objective function are equal to zero at $\alpha = 0$. Theorem 6 states them at the hypothesized value of the parameter of interest, $\text{H}_0 : \alpha = 0$.

Theorem 8: *When Assumption 1 holds, $m = 1$ and using an extension of the weak identification and misspecification assumption (23):*

$$\begin{array}{c} J = J(0) = \frac{1}{\sqrt{N}} V_{JJ}^{\frac{1}{2}} C, \;\; \mu_J = \frac{1}{\sqrt{N}} V_{\mu\mu.J}^{\frac{1}{2}} a \qquad \Leftrightarrow \\ C = \sqrt{N} V_{JJ}^{-\frac{1}{2}} J, \;\; a = \sqrt{N} V_{\mu\mu.J}^{-\frac{1}{2}} \mu_J = \sqrt{N} V_{\mu\mu.J}^{-\frac{1}{2}} \mu(0) - V_{\mu\mu.J}^{-\frac{1}{2}} V_{\mu J} V_{JJ}^{-\frac{1}{2}\prime} C \end{array} \quad (57)$$

for $\mu_J = \mu(0) - V_{\mu J}V_{JJ}^{-1}J(0)$, $V_{\mu\mu.J} = V_{\mu\mu} - V_{\mu J}V_{JJ}^{-1}V_{\mu J}'$, $V_{JJ.\mu} = V_{JJ} - V_{\mu J}V_{\mu\mu}^{-1}V_{J\mu}$, *and* C *and* a *finite non-zero* k*-dimensional vectors, we can express the first order condition and (scaled) Hessian at* $H_0 : \alpha = 0$ *as:*

$$\begin{aligned}
\tfrac{N}{2}\tfrac{\partial}{\partial\alpha}Q_{CUE}(\alpha=0) = &\; a'V_{\mu\mu.J}^{\frac{1}{2}\prime}V_{JJ}^{-\frac{1}{2}\prime}C = a'B'C = a^{*\prime}S_BC^* = 0 \\
\tfrac{N}{2}\tfrac{\partial^2}{(\partial\alpha)^2}Q_{CUE}(\alpha=0) = &\; a'V_{\mu\mu.J}^{\frac{1}{2}\prime}V_{JJ}^{-1}V_{\mu\mu.J}^{\frac{1}{2}}a - C'V_{JJ}^{-\frac{1}{2}}V_{\mu\mu.J}V_{JJ}^{-\frac{1}{2}\prime}C - 2a'V_{\mu\mu.J}^{\frac{1}{2}\prime}V_{JJ}^{-1}V_{\mu J}'V_{JJ}^{-\frac{1}{2}\prime}C \\
= &\; a'B'Ba - C'BB'C - 2a'B'D'B'C \\
= &\; a^{*\prime}S_B^2a^* - C^{*\prime}S_B^2C^* - 2a^{*\prime}S_BU_B'V_DS_DU_D'V_BS_BC^* \\
= &\; (S_Ba^* - U_B'V_BS_DU_D'V_BS_BC^*)'(S_Ba^* - U_B'V_BS_DU_D'V_BS_BC^*) - \\
&\; C^{*\prime}S_BV_B'U_D(I_{k_f} + S_D^2)U_D'V_BS_BC^*
\end{aligned} \tag{58}$$

with $B = V_{JJ}^{-\frac{1}{2}}V_{\mu\mu.J}^{\frac{1}{2}} = U_BS_BV_B'$, $D = V_{\mu\mu.J}^{-\frac{1}{2}}V_{\mu J}V_{JJ}^{-\frac{1}{2}} = U_DS_DV_D'$, $a^* = V_B'a$, $C^* = U_B'C$, *with* U_B, U_D, V_B, V_D *orthonormal* $k_f \times k_f$ *dimensional matrices and* S_B *and* S_D *the* $k_f \times k_f$ *dimensional diagonal matrix with the singular values of* B *and* D *in decreasing order on the main diagonal.*

Proof. see the Appendix. ■

The specification of the first order condition and Hessian in Theorem 8 shows the difficulty of pinning down the population parameter setting(s) where the Hessian equals zero when the covariance matrix does not have a KPS structure. When the covariance matrix has a KPS structure, so $V_{\mu\mu} = \omega_{\mu\mu}Q$, $V_{JJ} = \omega_{JJ}Q$, $V_{\mu J} = \omega_{\mu J}Q$, $V_{\mu\mu.J} = \omega_{\mu\mu.J}Q$, $\omega_{\mu\mu.J} = \omega_{\mu\mu} - \omega_{\mu J}\omega_{JJ}^{-1}\omega_{\mu J}'$, the first order condition at $\alpha = 0$, $\left(\frac{\omega_{\mu\mu.J}}{\omega_{JJ}}\right)^{\frac{1}{2}}a'C = 0$, further implies that the Hessian at $\alpha = 0$ corresponds with:

$$\begin{aligned}
\tfrac{N}{2}\tfrac{\partial^2}{(\partial\alpha)^2}Q_{CUE}(\alpha=0) = &\; \tfrac{\omega_{\mu\mu.J}}{\omega_{JJ}}\left(a - \tfrac{\omega_{\mu J}}{\sqrt{\omega_{\mu\mu.J}\omega_{JJ}}}C\right)'\left(a - \tfrac{\omega_{\mu J}}{\sqrt{\omega_{\mu\mu.J}\omega_{JJ}}}C\right) - \\
&\; \tfrac{\omega_{\mu\mu.J}}{\omega_{JJ}}C'C\left(1 + \tfrac{\omega_{\mu J}^2}{\omega_{\mu\mu.J}\omega_{JJ})}\right) \\
= &\; \tfrac{\omega_{\mu\mu.J}}{\omega_{JJ}}(a'a - C'C),
\end{aligned} \tag{59}$$

so a zero value of the derivative and Hessian imply that $a'C = 0$ and $a'a = C'C$. Hence, the first order condition and Hessian at $\alpha = 0$ can be expressed as functions of the small number of parameters on which the maximal invariant depends: $a'C$, $a'a$ and $C'C$, which similarly applies to the condition that IS equals MISS. The population value of the conditioning statistic equals $a'a + C'C$ so it provides an (almost asymptotically independent) estimator of the remaining nuisance parameter, (double) the length of $a'a$ and $C'C$.

For a non-KPS covariance matrix, we have to use the data at hand to obtain the nuisance parameters appearing in (58): B and D which can equivalently be represented by S_B, U_B, V_B and S_D, U_D, V_D. We can then set a value of a^* and solve for the value of C^* which makes (58) hold. Using these, we can simulate data and obtain realized values of $\widehat{\text{MISS}}$ and $\widehat{\text{IS}}$, while the sample analog of

$$\begin{aligned} \text{rk}= \; & (S_B a^* - U_B' V_B S_D U_D' V_B S_B C^*)'(S_B a^* - U_B' V_B S_D U_D' V_B S_B C^*)+ \\ & C^{*\prime} S_B V_B' U_D (I_{k_f} + S_D^2) U_D' V_B S_B C^*, \end{aligned} \tag{60}$$

could serve as a conditioning statistic. This all shows that while the extension to more general covariance matrices is empirically very relevant, it is also faces considerable challenges which we leave for important future work.

6 Conclusions

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

The conditional critical value function is constructed for a setting of homoskedasticity.

In future work, we plan to extend it to more general covariance structure. Because the null distribution then depends on more parameters, for reasons of brevity, we refrained from doing so in the current paper.

References

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

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

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

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

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

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

[7] 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.

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

[9] Hansen, B.E. and S. Lee. Inference for Iterated GMM under Misspecification. *Econometrica*, 2021. Forthcoming.

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

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

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

[13] 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.

[14] F. Kleibergen. Efficient size correct subset inference in homoskedastic linear instrumental variables regression. *Journal of Econometrics*, **221**:346–372, 2021.

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

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

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

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

[19] 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.

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

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

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

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

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

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

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

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

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

[29] Hall White. *Asymptotic Theory for Econometricians*. Academic Press, 1984.

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

Appendix

Proof of Theorem 1: We first specify:

$$\mathcal{U}_1\mathcal{S}_1\mathcal{V}_1' = (\mathcal{U}_1\mathcal{S}_1\mathcal{V}_{11}' \vdots \mathcal{U}_1\mathcal{S}_1\mathcal{V}_{21}') = Q^{-\frac{1}{2}}D^*\left(-\theta^* \vdots I_m\right)\Omega^{-\frac{1}{2}},$$

so using a Cholesky decomposition of $\Omega^{-\frac{1}{2}}$:

$$\begin{pmatrix} \omega_{\mu\mu.J}^{-\frac{1}{2}} & 0 \\ -\Omega_{JJ}^{-1}\omega_{J\mu}\omega_{\mu\mu.J}^{-\frac{1}{2}} & \Omega_{JJ}^{-\frac{1}{2}} \end{pmatrix},$$

with $\omega_{\mu\mu.J} = \omega_{\mu\mu} - \omega_{\mu J}\Omega_{JJ}^{-1}\omega_{J\mu}$, from which the specifications of D^* and θ^* result:

$$\begin{aligned} \mathcal{U}_1\mathcal{S}_1\mathcal{V}_{21}' &= Q^{-\frac{1}{2}}D^*\Omega_{JJ}^{-\frac{1}{2}} & \Leftrightarrow \\ D^* &= Q^{\frac{1}{2}}\mathcal{U}_1S_1\mathcal{V}_{21}'\Omega_{JJ}^{\frac{1}{2}} \\ \mathcal{U}_1\mathcal{S}_1\mathcal{V}_{11}' &= Q^{-\frac{1}{2}}D^*(-\theta^* - \Omega_{JJ}^{-1}\omega_{J\mu})\omega_{\mu\mu.J}^{-\frac{1}{2}} & \Leftrightarrow \\ \mathcal{U}_1\mathcal{S}_1\mathcal{V}_{11}' &= \mathcal{U}_1S_1\mathcal{V}_{21}'\Omega_{JJ}^{\frac{1}{2}}(-\theta^* - \Omega_{JJ}^{-1}\omega_{J\mu})\omega_{\mu\mu.J}^{-\frac{1}{2}} & \Leftrightarrow \\ \Omega_{JJ}^{\frac{1}{2}}(\theta^* + \Omega_{JJ}^{-1}\omega_{J\mu})\omega_{\mu\mu.J}^{-\frac{1}{2}} &= -\mathcal{V}_{21}'^{-1}\mathcal{V}_{11}' & \Leftrightarrow \\ \theta^* &= -\Omega_{JJ}^{-\frac{1}{2}}\mathcal{V}_{21}'^{-1}\mathcal{V}_{11}'\omega_{\mu\mu.J}^{\frac{1}{2}} - \Omega_{JJ}^{-1}\omega_{J\mu} \\ &= -\Omega_{JJ}^{-1}\omega_{J\mu} + \Omega_{JJ}^{-\frac{1}{2}}\mathcal{V}_{22}\mathcal{V}_{12}^{-1}\omega_{\mu\mu.J}^{\frac{1}{2}} \end{aligned}$$

since $\mathcal{V}_{11}'\mathcal{V}_{12} + \mathcal{V}_{21}'\mathcal{V}_{22} = 0$, so $-\mathcal{V}_{21}'^{-1}\mathcal{V}_{11}' = \mathcal{V}_{22}\mathcal{V}_{12}^{-1}$, $-\mathcal{V}_{11}\mathcal{V}_{21}^{-1} = \mathcal{V}_{12}^{-1\prime}\mathcal{V}_{22}'$.

We next specify:

$$\mathcal{U}_2\mathcal{S}_2\mathcal{V}_2' = Q^{\frac{1}{2}}D_\perp^*\lambda\left(-\theta^* \vdots I_m\right)_\perp\Omega^{\frac{1}{2}},$$

so using

$$\mathcal{U} = \begin{pmatrix} \mathcal{U}_{11} & \mathcal{U}_{12} \\ \mathcal{U}_{21} & \mathcal{U}_{22} \end{pmatrix}, \text{ and } V = \begin{pmatrix} \mathcal{V}_{11} & \mathcal{V}_{12} \\ \mathcal{V}_{21} & \mathcal{V}_{22} \end{pmatrix},$$

where $\mathcal{U}_{11}$, $\mathcal{V}_{21}$ are $m \times m$ dimensional matrices; $\mathcal{V}_{11}'$, $\mathcal{V}_{22}$ are $m \times 1$ dimensional vectors, $\mathcal{U}_{12}$, $\mathcal{U}_{21}$, and $\mathcal{U}_{22}$ are $m \times (k_f - m)$, $(k_f - m) \times m$ and $(k_f - m) \times (k_f - m)$ dimensional matrices and $\mathcal{V}_{12}$ is a scalar, $\mathcal{U}_1 = \binom{\mathcal{U}_{11}}{\mathcal{U}_{21}}$, $\mathcal{U}_2 = \binom{\mathcal{U}_{12}}{\mathcal{U}_{22}}$. A convenient specification of $D_\perp^*$, which also satisfies $D_\perp^{*\prime} Q D_\perp^* \equiv I_{k_f-m}$, is:

$$\begin{aligned}
D_\perp^* = \; & Q^{-\frac{1}{2}} \begin{pmatrix} -\mathcal{U}_{11}^{\prime -1}\mathcal{S}_1^{-1}\mathcal{V}_{21}^{-1}\Omega_{JJ}^{-\frac{1}{2}}\Omega_{JJ}^{\frac{1}{2}}\mathcal{V}_{21}\mathcal{S}_1\mathcal{U}_{21}' \\ I_{k_f-m} \end{pmatrix} \\
& \left[\begin{pmatrix} -\mathcal{U}_{11}^{\prime -1}\mathcal{S}_1^{-1}\mathcal{V}_{21}^{-1}\Omega_{JJ}^{-\frac{1}{2}}\Omega_{JJ}^{\frac{1}{2}}\mathcal{V}_{21}\mathcal{S}_1\mathcal{U}_{21}' \\ I_{k_f-m} \end{pmatrix}' \right. \\
& \left. \begin{pmatrix} -\mathcal{U}_{11}^{\prime -1}\mathcal{S}_1^{-1}\mathcal{V}_{21}^{-1}\Omega_{JJ}^{-\frac{1}{2}}\Omega_{JJ}^{\frac{1}{2}}\mathcal{V}_{21}\mathcal{S}_1\mathcal{U}_{21}' \\ I_{k_f-m} \end{pmatrix} \right]^{-\frac{1}{2}} \\
= \; & Q^{-\frac{1}{2}} \begin{pmatrix} -\mathcal{U}_{11}^{\prime -1}\mathcal{U}_{21}' \\ I_{k_f-m} \end{pmatrix} \left[\begin{pmatrix} -\mathcal{U}_{11}^{\prime -1}\mathcal{U}_{21}' \\ I_{k_f-m} \end{pmatrix}' \begin{pmatrix} -\mathcal{U}_{11}^{\prime -1}\mathcal{U}_{21}' \\ I_{k_f-m} \end{pmatrix} \right]^{-\frac{1}{2}} \\
= \; & Q^{-\frac{1}{2}} \begin{pmatrix} \mathcal{U}_{12}\mathcal{U}_{22}^{-1} \\ I_{k_f-m} \end{pmatrix} \left[\begin{pmatrix} \mathcal{U}_{12}\mathcal{U}_{22}^{-1} \\ I_{k_f-m} \end{pmatrix}' \begin{pmatrix} \mathcal{U}_{12}\mathcal{U}_{22}^{-1} \\ I_{k_f-m} \end{pmatrix} \right]^{-\frac{1}{2}} \\
= \; & Q^{-\frac{1}{2}} \begin{pmatrix} \mathcal{U}_{12}\mathcal{U}_{22}^{-1} \\ I_{k_f-m} \end{pmatrix} (I_{k_f-m} + \mathcal{U}_{22}^{-1\prime}\mathcal{U}_{12}'\mathcal{U}_{12}\mathcal{U}_{22}^{-1})^{-\frac{1}{2}} \\
= \; & Q^{-\frac{1}{2}} \begin{pmatrix} \mathcal{U}_{12} \\ \mathcal{U}_{22} \end{pmatrix} \mathcal{U}_{22}^{-1} (\mathcal{U}_{22}^{-1\prime}(\mathcal{U}_{12}'\mathcal{U}_{12} + \mathcal{U}_{22}'\mathcal{U}_{22})\mathcal{U}_{22}^{-1})^{-\frac{1}{2}} \\
= \; & Q^{-\frac{1}{2}} \begin{pmatrix} \mathcal{U}_{12} \\ \mathcal{U}_{22} \end{pmatrix} \mathcal{U}_{22}^{-1} (\mathcal{U}_{22}^{-1\prime}\mathcal{U}_{22}^{-1})^{-\frac{1}{2}} \\
= \; & Q^{-\frac{1}{2}} \begin{pmatrix} \mathcal{U}_{12} \\ \mathcal{U}_{22} \end{pmatrix} \mathcal{U}_{22}^{-1} (\mathcal{U}_{22}\mathcal{U}_{22}')^{\frac{1}{2}} \\
= \; & Q^{-\frac{1}{2}} \mathcal{U}_2 \mathcal{U}_{22}^{-1} (\mathcal{U}_{22}\mathcal{U}_{22}')^{\frac{1}{2}}
\end{aligned}$$

since $\mathcal{U}_{11}'\mathcal{U}_{12} + \mathcal{U}_{21}'\mathcal{U}_{22} = 0$ (because of the orthogonality of $\mathcal{U}$), $\mathcal{U}_{12}\mathcal{U}_{22}^{-1} = -\mathcal{U}_{11}^{\prime -1}\mathcal{U}_{21}'$, and $\mathcal{U}_{12}'\mathcal{U}_{12} + \mathcal{U}_{22}'\mathcal{U}_{22} = I_{k_f-m}$.

Similarly, since $\left(-\theta^* \vdots I_m\right)_\perp = \left[\binom{1}{\theta^*}'\Omega\binom{1}{\theta^*}\right]^{-\frac{1}{2}} \binom{1}{\theta^*}'$, so $\left(-\theta^* \vdots I_m\right)_\perp \Omega \left(-\theta^* \vdots I_m\right)_\perp' = 1$, and using that

$$\Omega^{\frac{1}{2}} = \begin{pmatrix} \omega_{\mu\mu.J}^{\frac{1}{2}} & \omega_{\mu J}\Omega_{JJ}^{-\frac{1}{2}} \\ 0 & \Omega_{JJ}^{\frac{1}{2}} \end{pmatrix},$$

we have that

$$
\begin{aligned}
\left(-\theta^{*} \vdots I_m\right)_{\perp}\Omega^{\frac{1}{2}} &= \left[\tbinom{1}{\theta^*}'\Omega\tbinom{1}{\theta^*}\right]^{-\frac{1}{2}}\tbinom{1}{\theta^*}'\Omega^{\frac{1}{2}} \\
&= \left[\tbinom{1}{\theta^*}'\Omega\tbinom{1}{\theta^*}\right]^{-\frac{1}{2}}(1 \vdots -\omega_{\mu\mu.J}^{\frac{1}{2}}\mathcal{V}_{11}\mathcal{V}_{21}^{-1}\Omega_{JJ}^{-\frac{1}{2}}-\omega_{\mu J}\Omega_{JJ}^{-1}) \\
&\quad \begin{pmatrix} \omega_{\mu\mu.J}^{\frac{1}{2}} & \omega_{\mu J}\Omega_{JJ}^{-\frac{1}{2}} \\ 0 & \Omega_{JJ}^{\frac{1}{2}} \end{pmatrix} \\
&= \left[\tbinom{1}{\theta^*}'\Omega\tbinom{1}{\theta^*}\right]^{-\frac{1}{2}}\omega_{\mu\mu.J}^{\frac{1}{2}}(1 \vdots -\mathcal{V}_{11}\mathcal{V}_{21}^{-1}) \\
&= \left[(1 \vdots -\mathcal{V}_{11}\mathcal{V}_{21}^{-1})(1 \vdots -\mathcal{V}_{11}\mathcal{V}_{21}^{-1})'\right]^{-\frac{1}{2}}(1 \vdots -\mathcal{V}_{11}\mathcal{V}_{21}^{-1}) \\
&= \left[(1 \vdots \mathcal{V}_{12}^{-1\prime}\mathcal{V}_{22}')(1 \vdots \mathcal{V}_{12}^{-1\prime}\mathcal{V}_{22}')'\right]^{-\frac{1}{2}}(1 \vdots -\mathcal{V}_{12}^{-1\prime}\mathcal{V}_{22}') \\
&= (\mathcal{V}_{12}^{-1\prime}(\mathcal{V}_{12}'\mathcal{V}_{12}+\mathcal{V}_{22}'\mathcal{V}_{22})\mathcal{V}_{12}^{-1})^{-\frac{1}{2}}\mathcal{V}_{12}^{-1\prime}\left(\mathcal{V}_{12}' \vdots \mathcal{V}_{22}'\right) \\
&= (\mathcal{V}_{12}^{-1\prime}\mathcal{V}_{12}^{-1})^{-\frac{1}{2}}\mathcal{V}_{12}^{-1\prime}\left(\mathcal{V}_{12}' \vdots \mathcal{V}_{22}'\right)\Omega^{-\frac{1}{2}} \\
&= (\mathcal{V}_{12}\mathcal{V}_{12}')^{\frac{1}{2}}\mathcal{V}_{12}^{-1\prime}\mathcal{V}_{2}'
\end{aligned}
$$

since $\mathcal{V}_{11}'\mathcal{V}_{12}+\mathcal{V}_{21}'\mathcal{V}_{22}=0$, so $-\mathcal{V}_{21}^{\prime -1}\mathcal{V}_{11}'=\mathcal{V}_{22}\mathcal{V}_{12}^{-1}$, $-\mathcal{V}_{11}\mathcal{V}_{21}^{-1}=\mathcal{V}_{12}^{-1\prime}\mathcal{V}_{22}'$, and $\mathcal{V}_{12}'\mathcal{V}_{12}+\mathcal{V}_{22}'\mathcal{V}_{22}=1$, from which it then results that

$$\lambda=(\mathcal{U}_{22}\mathcal{U}_{22}')^{-\frac{1}{2}}\mathcal{U}_{22}S_2\mathcal{V}_{12}'(\mathcal{V}_{12}\mathcal{V}_{12}')^{-\frac{1}{2}}.$$

We show the equivalence with the k-class expression for the pseudo-true value θ^* :

$$
\begin{aligned}
\theta^* &= (J(0)'Q^{-1}J(0)-\tau\Omega_{JJ})^{-1}(J(0)'Q^{-1}\mu_f(0)-\tau\omega_{J\mu}) \\
&= \Omega_{JJ}^{-\frac{1}{2}}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}-\tau I_m)^{-1}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}\mu_f(0)-\tau\Omega_{JJ}^{-\frac{1}{2}\prime}\omega_{J\mu})
\end{aligned}
$$

where $\tau=0$ leads to the two stage estimator and τ equal to the smallest root, $\tau_{\min}$, of the characteristic polynomial

$$\left|\tau\Omega-(\mu_f(0) \vdots J(0))'Q^{-1}(\mu_f(0) \vdots J(0))\right|=0,$$

provides the CUE. The singular value decomposition of $(\mu_f(0) \vdots J(0))$ shows that $\tau_{\min}=\mathcal{S}_2'\mathcal{S}_2=\lambda'\lambda$.

According to the singular value decomposition:

$$\begin{aligned}
Q^{-\frac{1}{2}}\left(\mu_f(0)-J(0)\Omega_{JJ}^{-1}\omega_{J\mu}\right)\omega_{\mu\mu.J}^{-\frac{1}{2}} &= \mathcal{US}\begin{pmatrix}\mathcal{V}_{11} & \mathcal{V}_{12}\end{pmatrix}' && \Leftrightarrow\\
Q^{-\frac{1}{2}}\mu_f(0) &= Q^{-\frac{1}{2}}J(0)\Omega_{JJ}^{-1}\omega_{J\mu}+\mathcal{US}\begin{pmatrix}\mathcal{V}_{11} & \mathcal{V}_{12}\end{pmatrix}'\omega_{\mu\mu.J}^{\frac{1}{2}} && \Leftrightarrow\\
\mu_f(0) &= J(0)\Omega_{JJ}^{-1}\omega_{J\mu}+Q^{\frac{1}{2}}\left(\mathcal{U}_1\mathcal{S}_1\mathcal{V}_{11}'+\mathcal{U}_2\mathcal{S}_2\mathcal{V}_{12}'\right)\omega_{\mu\mu.J}^{\frac{1}{2}}\\
Q^{-\frac{1}{2}}J(0)\Omega_{JJ}^{-\frac{1}{2}} &= \mathcal{US}\begin{pmatrix}\mathcal{V}_{21} & \mathcal{V}_{22}\end{pmatrix}' && \Leftrightarrow\\
&= \mathcal{U}_1\mathcal{S}_1\mathcal{V}_{21}'+\mathcal{U}_2\mathcal{S}_2\mathcal{V}_{22}'\\
\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}} &= \mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{21}'+\mathcal{V}_{22}\mathcal{S}_2'\mathcal{S}_2\mathcal{V}_{22}'\\
&= \mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{21}'+\tau_{\min}\mathcal{V}_{22}\mathcal{V}_{22}' && \Leftrightarrow\\
&= \mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1-\tau_{\min}I_m)\mathcal{V}_{21}'+\tau_{\min}(\mathcal{V}_{21}\mathcal{V}_{21}'+\mathcal{V}_{22}\mathcal{V}_{22}') && \Leftrightarrow\\
&= \mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1-\tau_{\min}I_m)\mathcal{V}_{21}'+\tau_{\min}I_m,\\
\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}+I_m &= \mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1-\tau_{\min}I_m)\mathcal{V}_{21}'+(\tau_{\min}+1)I_m
\end{aligned}$$

since $\mathcal{V}_{21}\mathcal{V}_{21}'+\mathcal{V}_{22}\mathcal{V}_{22}'=I_m$, which shows that when $\mathcal{V}_{21}$ is of reduced rank, like, for example, zero, $\tau_{\min}$ is one of the characteristic roots/eigenvalues of $\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}$.

We first construct the expression for the pseudo-true value of the two stage estimator θ_{2s}^* :

$$\begin{aligned}
J(0)'Q^{-1}\mu_f(0) &= \left[Q^{-\frac{1}{2}}J(0)\right]'\left[Q^{-\frac{1}{2}}\mu_f(0)\right]\\
&= J(0)'Q^{-1}J(0)\Omega_{JJ}^{-1}\omega_{J\mu}+\Omega_{JJ}^{\frac{1}{2}}\left[\mathcal{U}_1\mathcal{S}_1\mathcal{V}_{21}'+\mathcal{U}_2\mathcal{S}_2\mathcal{V}_{22}'\right]'\\
&\quad\left[\mathcal{U}_1\mathcal{S}_1\mathcal{V}_{11}'+\mathcal{U}_2\mathcal{S}_2\mathcal{V}_{12}'\right]\omega_{\mu\mu.J}^{\frac{1}{2}}\\
&= J(0)'Q^{-1}J(0)\Omega_{JJ}^{-1}\omega_{J\mu}+\Omega_{JJ}^{\frac{1}{2}}\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}'\omega_{\mu\mu.J}^{\frac{1}{2}}+\\
&\quad\Omega_{JJ}^{\frac{1}{2}}\mathcal{V}_{22}\mathcal{S}_2'\mathcal{S}_2\mathcal{V}_{12}'\omega_{\mu\mu.J}^{\frac{1}{2}}\\
&= J(0)'Q^{-1}J(0)\Omega_{JJ}^{-1}\omega_{J\mu}+\Omega_{JJ}^{\frac{1}{2}}\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}'\omega_{\mu\mu.J}^{\frac{1}{2}}+\\
&\quad\tau_{\min}\Omega_{JJ}^{\frac{1}{2}}\mathcal{V}_{22}\mathcal{V}_{12}'\omega_{\mu\mu.J}^{\frac{1}{2}}\\
&= J(0)'Q^{-1}J(0)\Omega_{JJ}^{-1}\omega_{J\mu}+\Omega_{JJ}^{\frac{1}{2}}\mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1-\tau_{\min})\mathcal{V}_{11}'\omega_{\mu\mu.J}^{\frac{1}{2}}\\
\theta_{2s}^* &= -(J(0)'Q^{-1}J(0))^{-1}J(0)'Q^{-1}\mu_f(0)\\
&= -\Omega_{JJ}^{-1}\omega_{J\mu}-\\
&\quad\Omega_{JJ}^{-\frac{1}{2}}\left(\mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1-\tau_{\min}I_m)\mathcal{V}_{21}'+\tau_{\min}I_m\right)^{-1}\mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1-\tau_{\min})\mathcal{V}_{11}'\omega_{\mu\mu.J}^{\frac{1}{2}}\\
&= -\Omega_{JJ}^{-1}\omega_{J\mu}-\Omega_{JJ}^{-\frac{1}{2}}\left(\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{21}'+\tau_{\min}\mathcal{V}_{22}\mathcal{V}_{22}'\right)^{-1}\left(\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}'+\tau_{\min}\mathcal{V}_{22}\mathcal{V}_{12}'\right)\omega_{\mu\mu.J}^{\frac{1}{2}}
\end{aligned}$$

since $\mathcal{V}_{21}\mathcal{V}_{11}'+\mathcal{V}_{22}\mathcal{V}_{12}'=0$, so $\mathcal{V}_{22}\mathcal{V}_{12}'=-\mathcal{V}_{21}\mathcal{V}_{11}'$.

The pseudo-true value of CUE can be specified as

$$
\begin{aligned}
\theta^*_{CUE} = &\ -\Omega_{JJ}^{-\frac{1}{2}}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}-\tau_{\min}I_m)^{-1}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}\mu_f(0)-\tau\Omega_{JJ}^{-\frac{1}{2}\prime}\omega_{J\mu}) \\
= &\ -\Omega_{JJ}^{-\frac{1}{2}}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}-\tau_{\min}I_m)^{-1} \\
&\left[\Omega_{JJ}^{-\frac{1}{2}}J(0)'Q^{-1}\left(J(0)\Omega_{JJ}^{-1}\omega_{J\mu}+\mathcal{US}\begin{pmatrix}\mathcal{V}_{11} & \mathcal{V}_{12}\end{pmatrix}'\omega_{\mu\mu.J}^{\frac{1}{2}}\right)-\tau_{\min}\Omega_{JJ}^{-\frac{1}{2}\prime}\omega_{J\mu}\right] \\
= &\ -\Omega_{JJ}^{-\frac{1}{2}}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}-\tau_{\min}I_m)^{-1}\left(\Omega_{JJ}^{-\frac{1}{2}}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}-\tau_{\min}I_m\right)\Omega_{JJ}^{-\frac{1}{2}\prime}\omega_{J\mu}- \\
&\ \Omega_{JJ}^{-\frac{1}{2}}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}-\tau_{\min}I_m)^{-1}\Omega_{JJ}^{-\frac{1}{2}}J(0)'Q^{-\frac{1}{2}}\mathcal{US}\begin{pmatrix}\mathcal{V}_{11} & \mathcal{V}_{12}\end{pmatrix}'\omega_{\mu\mu.J}^{\frac{1}{2}} \\
= &\ -\Omega_{JJ}^{-1}\omega_{J\mu}-\Omega_{JJ}^{-\frac{1}{2}}\left[(\mathcal{U}_1\mathcal{S}_1\mathcal{V}_{21}'+\mathcal{U}_2\mathcal{S}_2\mathcal{V}_{22}')'(\mathcal{U}_1\mathcal{S}_1\mathcal{V}_{21}'+\mathcal{U}_2\mathcal{S}_2\mathcal{V}_{22}')-\tau_{\min}I_m\right]^{-1} \\
&\ (\mathcal{U}_1\mathcal{S}_1\mathcal{V}_{21}'+\mathcal{U}_2\mathcal{S}_2\mathcal{V}_{22}')'(\mathcal{U}_1\mathcal{S}_1\mathcal{V}_{11}'+\mathcal{U}_2\mathcal{S}_2\mathcal{V}_{12}')\omega_{\mu\mu.J}^{\frac{1}{2}} \\
= &\ -\Omega_{JJ}^{-1}\omega_{J\mu}-\Omega_{JJ}^{-\frac{1}{2}}\left[\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{21}'+\mathcal{V}_{22}\mathcal{S}_2'\mathcal{S}_2\mathcal{V}_{22}'-\tau_{\min}I_m\right]^{-1}(\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}'+\mathcal{V}_{22}\mathcal{S}_2'\mathcal{S}_2\mathcal{V}_{12}')\omega_{\mu\mu.J}^{\frac{1}{2}} \\
= &\ -\Omega_{JJ}^{-1}\omega_{J\mu}-\Omega_{JJ}^{-\frac{1}{2}}\left[\mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1-\tau_{\min}I_m)\mathcal{V}_{21}'+\tau_{\min}(\mathcal{V}_{21}\mathcal{V}_{21}'+\mathcal{V}_{22}\mathcal{V}_{22}'-I_m)\right]^{-1} \\
&\ (\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}'+\mathcal{V}_{22}\mathcal{S}_2'\mathcal{S}_2\mathcal{V}_{12}')\omega_{\mu\mu.J}^{\frac{1}{2}} \\
= &\ -\Omega_{JJ}^{-1}\omega_{J\mu}-\Omega_{JJ}^{-\frac{1}{2}}\left[\mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1-\tau_{\min}I_m)\mathcal{V}_{21}'\right]^{-1}(\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}'+\mathcal{V}_{22}\mathcal{S}_2'\mathcal{S}_2\mathcal{V}_{12}')\omega_{\mu\mu.J}^{\frac{1}{2}} \\
= &\ -\Omega_{JJ}^{-1}\omega_{J\mu}-\Omega_{JJ}^{-\frac{1}{2}}\mathcal{V}_{21}^{-1\prime}(\mathcal{S}_1'\mathcal{S}_1-\tau_{\min}I_m)\left(\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}'+\mathcal{V}_{21}^{-1}\mathcal{V}_{22}\mathcal{S}_2'\mathcal{S}_2\mathcal{V}_{12}'\right)\omega_{\mu\mu.J}^{\frac{1}{2}} \\
= &\ -\Omega_{JJ}^{-1}\omega_{J\mu}-\Omega_{JJ}^{-\frac{1}{2}}\mathcal{V}_{21}^{-1\prime}(\mathcal{S}_1'\mathcal{S}_1-\tau_{\min}I_m)\left(\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}'-\tau_{\min}\mathcal{V}_{11}'\mathcal{V}_{12}'^{-1}\mathcal{V}_{12}'\right)\omega_{\mu\mu.J}^{\frac{1}{2}} \\
= &\ -\Omega_{JJ}^{-1}\omega_{J\mu}-\Omega_{JJ}^{-\frac{1}{2}}\mathcal{V}_{21}^{-1\prime}(\mathcal{S}_1'\mathcal{S}_1-\tau_{\min}I_m)(\mathcal{S}_1'\mathcal{S}_1-\tau_{\min}I_m)\mathcal{V}_{11}'\omega_{\mu\mu.J}^{\frac{1}{2}} \\
= &\ -\Omega_{JJ}^{-1}\omega_{J\mu}-\Omega_{JJ}^{-\frac{1}{2}}\mathcal{V}_{21}^{-1\prime}\mathcal{V}_{11}'\omega_{\mu\mu.J}^{\frac{1}{2}}=-\Omega_{JJ}^{-1}\omega_{J\mu}+\Omega_{JJ}^{-\frac{1}{2}}\mathcal{V}_{22}\mathcal{V}_{12}^{-1}\omega_{\mu\mu.J}^{\frac{1}{2}},
\end{aligned}
$$

since $\mathcal{V}_{21}\mathcal{V}_{21}'+\mathcal{V}_{22}\mathcal{V}_{22}'=I_m$ and $\mathcal{V}_{21}\mathcal{V}_{11}'+\mathcal{V}_{22}\mathcal{V}_{12}'=0$, so $\mathcal{V}_{21}^{-1}\mathcal{V}_{22}=-\mathcal{V}_{11}'\mathcal{V}_{12}'^{-1}$.

The pseudo-true value of the least squares estimator, which results from the expression of the k-class estimator (10) when $\tau=-1$, reads:

$$
\begin{aligned}
\theta^*_{ls} = &\ -(J(0)'Q^{-1}J(0)+\Omega_{JJ})^{-1}(J(0)'Q^{-1}\mu_f(0)+\tau\omega_{J\mu}) \\
= &\ -\Omega_{JJ}^{-\frac{1}{2}}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}+I_m)^{-1}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}\mu_f(0)+\Omega_{JJ}^{-\frac{1}{2}\prime}\omega_{J\mu}) \\
= &\ -\Omega_{JJ}^{-\frac{1}{2}}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}+I_m)^{-1}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}+I_m)\Omega_{JJ}^{-\frac{1}{2}\prime}\omega_{J\mu}+ \\
&\ -\Omega_{JJ}^{-\frac{1}{2}}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}+I_m)^{-1}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-\frac{1}{2}}(\mathcal{U}_1\mathcal{S}_1\mathcal{V}_{11}'+\mathcal{U}_2\mathcal{S}_2\mathcal{V}_{12}')\omega_{\mu\mu.J}^{\frac{1}{2}} \\
= &\ -\Omega_{JJ}^{-1}\omega_{J\mu}-\Omega_{JJ}^{-\frac{1}{2}}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}+I_m)^{-1} \\
&\ [\mathcal{U}_1\mathcal{S}_1\mathcal{V}_{21}'+\mathcal{U}_2\mathcal{S}_2\mathcal{V}_{22}']'[\mathcal{U}_1\mathcal{S}_1\mathcal{V}_{11}'+\mathcal{U}_2\mathcal{S}_2\mathcal{V}_{12}']\omega_{\mu\mu.J}^{\frac{1}{2}} \\
= &\ -\Omega_{JJ}^{-1}\omega_{J\mu}-\Omega_{JJ}^{-\frac{1}{2}}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}+I_m)^{-1} \\
&\ (\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}'+\mathcal{V}_{22}\mathcal{S}_2'\mathcal{S}_2\mathcal{V}_{12}')\omega_{\mu\mu.J}^{\frac{1}{2}} \\
= &\ -\Omega_{JJ}^{-1}\omega_{J\mu}-\Omega_{JJ}^{-\frac{1}{2}}(\Omega_{JJ}^{-\frac{1}{2}\prime}J(0)'Q^{-1}J(0)\Omega_{JJ}^{-\frac{1}{2}}+I_m)^{-1} \\
&\ (\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}'+\tau_{\min}\mathcal{V}_{22}\mathcal{V}_{12}')\omega_{\mu\mu.J}^{\frac{1}{2}} \\
= &\ -\Omega_{JJ}^{-1}\omega_{J\mu}-\Omega_{JJ}^{-\frac{1}{2}}(\mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1-\tau_{\min})\mathcal{V}_{21}'+(\tau_{\min}+1)I_m)^{-1}\mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1-\tau_{\min}I_m)\mathcal{V}_{11}'\omega_{\mu\mu.J}^{\frac{1}{2}}
\end{aligned}
$$

When the smallest singular value only loads on the Jacobian: $\mathcal{V}_{22}'\mathcal{V}_{22} = 1$. The orthonormality of $\mathcal{V}$ states that $\mathcal{V}'\mathcal{V} = \mathcal{V}\mathcal{V}'$=$I_{m+1}$: $\mathcal{V}_{11}'\mathcal{V}_{11} + \mathcal{V}_{21}'\mathcal{V}_{21} = I_m$, $\mathcal{V}_{12}'\mathcal{V}_{12} + \mathcal{V}_{22}'\mathcal{V}_{22} = 1$, $\mathcal{V}_{11}'\mathcal{V}_{12} + \mathcal{V}_{21}'\mathcal{V}_{22} = 0$, $\mathcal{V}_{11}\mathcal{V}_{11}' + \mathcal{V}_{12}\mathcal{V}_{12}' = 1$, $\mathcal{V}_{21}\mathcal{V}_{21}' + \mathcal{V}_{22}\mathcal{V}_{22}' = I_m$, $\mathcal{V}_{21}\mathcal{V}_{11}' + \mathcal{V}_{22}\mathcal{V}_{12}' = 0$, so when $\mathcal{V}_{22}'\mathcal{V}_{22} = 1$, $\mathcal{V}_{12} = 0$, since $\mathcal{V}_{12}$ is a scalar which next implies $\mathcal{V}_{11}\mathcal{V}_{11}' = 1$ and that $\mathcal{V}_{21}$ is a singular matrix which lies in the orthogonal complements of $\mathcal{V}_{11}$ and $\mathcal{V}_{22}$ because $\mathcal{V}_{21}\mathcal{V}_{21}' + \mathcal{V}_{22}\mathcal{V}_{22}' = I_m \Leftrightarrow \mathcal{V}_{21}\mathcal{V}_{21}'\mathcal{V}_{22} + \mathcal{V}_{22}\mathcal{V}_{22}'\mathcal{V}_{22} = \mathcal{V}_{22} \Leftrightarrow \mathcal{V}_{21}\mathcal{V}_{21}'\mathcal{V}_{22} + \mathcal{V}_{22} = \mathcal{V}_{22}$ (since $\mathcal{V}_{22}'\mathcal{V}_{22} = 1$) so $\mathcal{V}_{21}\mathcal{V}_{21}'\mathcal{V}_{22} = 0$ which implies that $\mathcal{V}_{21}'\mathcal{V}_{22} = 0$. Similarly, $\mathcal{V}_{11}'\mathcal{V}_{11} + \mathcal{V}_{21}'\mathcal{V}_{21} = I_m$, so $\mathcal{V}_{11}'\mathcal{V}_{11} + \mathcal{V}_{21}'\mathcal{V}_{21} = I_m$, $\mathcal{V}_{11}'\mathcal{V}_{11}\mathcal{V}_{11}' + \mathcal{V}_{21}'\mathcal{V}_{21}\mathcal{V}_{11}' = \mathcal{V}_{11}' \Leftrightarrow \mathcal{V}_{11}' + \mathcal{V}_{21}'\mathcal{V}_{21}\mathcal{V}_{11}' = \mathcal{V}_{11}' \Leftrightarrow \mathcal{V}_{21}'\mathcal{V}_{21}\mathcal{V}_{11}' = 0 \Leftrightarrow \mathcal{V}_{21}\mathcal{V}_{11}' = 0$.

We next define the orthonormal $m \times m$ dimensional matrix $(\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp})$, with $\mathcal{V}_{22,\perp}$ a $m \times (m-1)$ dimensional matrix orthogonal to $\mathcal{V}_{22}$ and B a $(m-1) \times m$ dimensional matrix such that $\mathcal{V}_{21} = (\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp}) \binom{0}{B}$. Because $\mathcal{V}_{21}'\mathcal{V}_{22} = 0$, we then have that $(\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp})'\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{21}'(\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp}) = \binom{0}{B} \mathcal{S}_1'\mathcal{S}_1 \binom{0}{B}'$ and $(\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp})'[\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{21}' + \tau_{\min}\mathcal{V}_{22}\mathcal{V}_{22}'](\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp}) = \binom{0}{B} \mathcal{S}_1'\mathcal{S}_1 \binom{0}{B}' + \tau_{\min} \binom{1}{0} \binom{1}{0}'] = \begin{pmatrix} \tau_{\min} & 0 \\ 0 & B\mathcal{S}_1'\mathcal{S}_1B' \end{pmatrix}$.

Also $(\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp})'(\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}' + \tau_{\min}\mathcal{V}_{22}\mathcal{V}_{12}')\omega_{\mu\mu.J}^{\frac{1}{2}} = \binom{0}{B} \mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}'\omega_{\mu\mu.J}^{\frac{1}{2}} + \tau_{\min}\binom{1}{0}\mathcal{V}_{12}'\omega_{\mu\mu.J}^{\frac{1}{2}}$.

Combining, we obtain:

$$\begin{aligned}
&[\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{21}' + \tau_{\min}\mathcal{V}_{22}\mathcal{V}_{22}']^{-1}(\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}' + \tau_{\min}\mathcal{V}_{22}\mathcal{V}_{12}') \\
&= (\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp}) \left((\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp})'[\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{21}' + \tau_{\min}\mathcal{V}_{22}\mathcal{V}_{22}'](\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp}) \right)^{-1} \\
&\quad (\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp})'(\mathcal{V}_{21}\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}' + \tau_{\min}\mathcal{V}_{22}\mathcal{V}_{12}') \\
&= (\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp}) \begin{pmatrix} \tau_{\min} & 0 \\ 0 & B\mathcal{S}_1'\mathcal{S}_1B' \end{pmatrix}^{-1} \left(\binom{0}{B} \mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}' + \tau_{\min}\binom{1}{0}\mathcal{V}_{12}' \right) \\
&= \mathcal{V}_{22}\mathcal{V}_{12}' + \mathcal{V}_{22,\perp}\left(B\mathcal{S}_1'\mathcal{S}_1B'\right)^{-1} B\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}' \\
&= \mathcal{V}_{22,\perp}\left(B\mathcal{S}_1'\mathcal{S}_1B'\right)^{-1} B\mathcal{S}_1'\mathcal{S}_1\mathcal{V}_{11}'
\end{aligned}$$

because $\mathcal{V}_{12} = 0$, since $\mathcal{V}_{22}'\mathcal{V}_{22} = 1$, and $\mathcal{V}_{22}'\mathcal{V}_{21} = 0$, which leads to the expression of θ_{2s}^* (21).

Also $(\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp})'\mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m) = \binom{0}{B}(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m) = \begin{pmatrix} 0 \\ B(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m) \end{pmatrix}$, so $\Omega_{JJ}^{-\frac{1}{2}}\left(\mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m)\mathcal{V}_{21}' + \tau_{\min}I_m\right)^{-1}\mathcal{V}_{21}(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m)\mathcal{V}_{11}'\omega_{\mu\mu.J}^{\frac{1}{2}} = \Omega_{JJ}^{-\frac{1}{2}}(\mathcal{V}_{22} \vdots \mathcal{V}_{22,\perp}) \begin{pmatrix} \tau_{\min} & 0 \\ 0 & B\mathcal{S}_1'\mathcal{S}_1B' \end{pmatrix}^{-1} \Big(B$
$\Omega_{JJ}^{-\frac{1}{2}}\mathcal{V}_{22,\perp}(B\mathcal{S}_1'\mathcal{S}_1B')^{-1}B(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m)\mathcal{V}_{11}'\omega_{\mu\mu.J}^{\frac{1}{2}}$. Hence

$$\theta_{LS}^* = -\Omega_{JJ}^{-1}\omega_{J\mu} - \Omega_{JJ}^{-\frac{1}{2}}\mathcal{V}_{22,\perp}(B\mathcal{S}_1'\mathcal{S}_1B')^{-1}B(\mathcal{S}_1'\mathcal{S}_1 - \tau_{\min}I_m)\mathcal{V}_{11}'\omega_{\mu\mu.J}^{\frac{1}{2}}.$$

Proof of Theorem 2: Can be adapted from other paper, see also Kleibergen and Zhan (2023).

The CUE population objective function:

$$\begin{aligned} Q_{CUE}(y) = \;& \mu_f(y)' \left[\binom{I_{k_f}}{(y \otimes I_{k_f})}' V \binom{I_{k_f}}{(y \otimes I_{k_f})} \right]^{-1} \mu_f(y), \\ & (\mu_f(0) + J(0)y)' \left[\binom{I_{k_f}}{(y \otimes I_{k_f})}' V \binom{I_{k_f}}{(y \otimes I_{k_f})} \right]^{-1} (\mu_f(0) + J(0)y) \end{aligned}$$

results from a step-wise optimization of the generalized reduced rank objective function:

$$Q_{CUE}(y, D) = \left[\text{vec}\left(\mu_f(0) \vdots J(0)) + D \left(-y \vdots I_m \right) \right) \right]' \Sigma^{-1} \left[\text{vec}\left(\left(\mu_f(0) \vdots J(0) \right) + D \left(-y \vdots I_m \right) \right) \right],$$

with $\Sigma = \text{Var}\left(\sqrt{T} \left(\hat{\mu}_f(0)' \vdots \text{vec}(\hat{J}(0))' \right)' \right)$, and

$$Q_{CUE}(y) = \min_{D \in \mathbb{R}^{k_f \times m}} Q_{CUE}(y, D).$$

This results since

$$\begin{aligned} D(y) = \;& \arg\min_{D \in \mathbb{R}^{k_f \times m}} Q_{CUE}(y, D) \\ = \;& - \left[\left(\left(-y \vdots I_m \right)' \otimes I_{k_f} \right)' \Sigma^{-1} \left(\left(-y \vdots I_m \right)' \otimes I_{k_f} \right) \right]^{-1} \\ & \left(\left(-y \vdots I_m \right)' \otimes I_{k_f} \right)' \Sigma^{-1} \text{vec}\left(\mu_f(0) \vdots J(0) \right), \end{aligned}$$

so when we substitute $D(y)$ in $Q_{CUE}(y, D)$:

$$\begin{aligned} Q_{CUE}(y, D(y))) = \;& \left[\text{vec}\left(\mu_f(0) \vdots J(0)) + D(y) \left(-y \vdots I_m \right) \right) \right]' \Sigma^{-1} \\ & \left[\text{vec}\left(\mu_f(0) \vdots J(0)) + D(y) \left(-y \vdots I_m \right) \right) \right] \\ = \;& \left(\Sigma^{-\frac{1}{2}} \text{vec}\left(\mu_f(0) \vdots J(0) \right) \right)' M_{\Sigma^{-\frac{1}{2}} \left(\left(-y \vdots I_m \right)' \otimes I_{k_f} \right)} \left(\Sigma^{-\frac{1}{2}} \text{vec}\left(\mu_f(0) \vdots J(0) \right) \right) \\ = \;& \left(\Sigma^{-\frac{1}{2}} \text{vec}\left(\mu_f(0) \vdots J(0) \right) \right)' P_{\Sigma^{\frac{1}{2}} \left(\left(1 \vdots y \right)' \otimes I_{k_f} \right)} \left(\Sigma^{-\frac{1}{2}} \text{vec}\left(\mu_f(0) \vdots J(0) \right) \right) \\ = \;& (\mu_f(0) + J(0)y)' \left[\binom{I_{k_f}}{(y \otimes I_{k_f})}' V \binom{I_{k_f}}{(y \otimes I_{k_f})} \right]^{-1} (\mu_f(0) + J(0)y), \end{aligned}$$

with $M_H = I_N - P_H$, $P_H = H(H'H)^{-1}H'$, for H a $k_f \times n$ dimensional full rank matrix and we used that for $H_\perp : k_f \times (k_f - m)$ dimensional defined such that $H_\perp' H \equiv 0,\ M_H = P_{H_\perp}$. For $H = \Sigma^{-\frac{1}{2}}\left(\left(-y \vdots I_m\right)' \otimes I_{k_f}\right)$, a specification for $H_\perp$ is:

$$H_\perp = \Sigma^{\frac{1}{2}}\left(\left(1 \vdots y\right)' \otimes I_{k_f}\right),$$

which we substituted in the above.

The minimal value of $Q_{CUE}(y, D)$ when minimizing with respect to $(y,\ D)$ equals the minimal value of

$$Q_{CUE}(g, A) = \left[\text{vec}\left(\left(\mu_f(0) \vdots J(0)\right) + A(I_m \vdots -g)\right)\right]' V^{-1}\left[\text{vec}\left(\left(\mu_f(0) \vdots J(0)\right) + A(I_m \vdots -g)\right)\right],$$

when minimizing with respect to (g, A), with A $k_f \times m$ matrix and g $m \times 1$. This results since $D(-y \vdots I_m)$ and $A(I_m \vdots -g)$ are equivalent representations of an $k_f \times (m+1)$ dimensional matrix of rank m (except for a measure zero space). Restricting the top element of $g = (g_1 \vdots g_2')'$, with g_1 a scalar and g_2 a $(m-1)$-dimensional vector, to zero, so $g_1 = 0$, does not decrease the minimal value of the above function. The resulting restricted specification reads

$$\begin{aligned}
Q_{CUE}(g_1 = 0, g_2, A) = & \left[\text{vec}\left(\left(\mu_f(0) \vdots J(0)\right) + A(I_m \vdots -g)\right)\right]' \\
& \begin{pmatrix} V_{\mu\mu.J}^{-1} & -V_{\mu\mu.J}^{-1}V_{\mu J}V_{JJ}^{-1} \\ -V_{JJ}^{-1}V_{J\mu}V_{\mu\mu.J}^{-1} & V_{JJ}^{-1} + V_{JJ}^{-1}V_{J\mu}V_{\mu\mu.J}^{-1}V_{\mu J}V_{JJ}^{-1} \end{pmatrix} \\
& \left[\text{vec}\left(\left(\mu_f(0) \vdots J(0)\right) + A(I_m \vdots -g)\right)\right] \\
= & \left[\mu_f(0) + a_1 - V_{\mu J}V_{JJ}^{-1}\text{vec}\left(J(0) + A_2\left(I_{m-1} \vdots -g_2\right)\right)\right]' V_{\mu\mu.J}^{-1} \\
& \left[\mu_f(0) + a_1 - V_{\mu J}V_{JJ}^{-1}\text{vec}\left(J(0) + A_2\left(I_{m-1} \vdots -g_2\right)\right)\right] + \\
& \left[\text{vec}\left(J(0) + A_2\left(I_{m-1} \vdots -g_2\right)\right)\right]' V_{JJ}^{-1}\left[\text{vec}\left(\beta + A_2\left(I_{m-1} \vdots -g_2\right)\right)\right],
\end{aligned}$$

where we used that $A = (a_1 \vdots A_2)$, a_1 $N \times 1$, A_2 $N \times (K-1)$ matrices and the partitioned inverse of V :

$$V = \begin{pmatrix} V_{\mu\mu} & V_{\mu J} \\ V_{J\mu} & V_{JJ} \end{pmatrix},\ V^{-1} = \begin{pmatrix} V_{\mu\mu.J}^{-1} & -V_{\mu\mu.J}^{-1}V_{\mu J}V_{JJ}^{-1} \\ -V_{JJ}^{-1}V_{J\mu}V_{\mu\mu.J}^{-1} & V_{JJ}^{-1} + V_{JJ}^{-1}V_{J\mu}V_{\mu\mu.J}^{-1}V_{\mu J}V_{JJ}^{-1} \end{pmatrix},$$

with $V_{\mu\mu.J} = V_{\mu\mu} - V_{\mu J}V_{JJ}^{-1}V_{J\mu}$.

Stepwise minimization of $Q_{CUE}(g_1 = 0, g_2, A) = Q_{CUE}(g_1 = 0, g_2, a_1, A_2)$ now results in:

$$\begin{aligned} a_1(g_1 = 0, g_2, A_2) = & \ \arg\min_{a_1 \in \mathbb{R}^N} Q_{CUE}(g_1 = 0, g_2, a_1, A_2) \\ = & \ -\left[\mu_f(0) - V_{\mu J}V_{JJ}^{-1}\text{vec}\left(J(0) + A_2\left(I_{m-1} \vdots -g_2\right)\right)\right] \end{aligned}$$

so

$$\begin{aligned} Q_{CUE}(g_1 = 0, g_2, A_2) = & \ \min_{a_1 \in \mathbb{R}^{k_f}} Q_{CUE}(g_2, a_1, A_2) \\ = & \ \left[\text{vec}\left(J(0) + A_2\left(I_{m-1} \vdots -g_2\right)\right)\right]' V_{JJ}^{-1}\left[\text{vec}\left(\beta + A_2\left(I_{m-1} \vdots -g_2\right)\right)\right] \\ = & \ \left[\text{vec}(J(0)) + \left(\left(I_{m-1} \vdots -g_2\right)' \otimes I_N\right)\text{vec}(A_2)\right]' \Sigma_{\hat{\beta}\hat{\beta}}^{-1} \\ & \ \left[\text{vec}(\beta) + \left(\left(I_{m-1} \vdots -g_2\right)' \otimes I_N\right)\text{vec}(A_2)\right]. \end{aligned}$$

When we next optimize $Q_{CUE}(g_1 = 0, g_2, A_2)$ over A_2 :

$$\begin{aligned} \text{vec}(A_2(g_1 = 0, g_2)) = & \ \arg\min_{A_2 \in \mathbb{R}^{k_f \times (m-1)}} Q_{CUE}(g_1 = 0, g_2, A_2) \\ = & \ -\left[\left(\left(I_{m-1} \vdots -g_2\right)' \otimes I_{k_f}\right)' V_{JJ}^{-1}\left(\left(I_{m-1} \vdots -g_2\right)' \otimes I_{k_f}\right)\right]^{-1} \\ & \ \left(\left(I_{m-1} \vdots -g_2\right)' \otimes I_{k_f}\right)' V_{JJ}^{-1}\text{vec}(J(0)), \end{aligned}$$

so we obtain:

$$\begin{aligned} Q_{CUE}(g_1 = 0, g_2) = & \ \min_{A_2 \in \mathbb{R}^{k_f \times (m-1)}} Q_{CUE}(g_2, A_2) \\ = & \ \left[\text{vec}(J(0)) + \left(\left(I_{m-1} \vdots -g_2\right)' \otimes I_{k_f}\right)\text{vec}(A_2(g_1 = 0, g_2))\right]' V_{JJ}^{-1} \\ & \ \left[\text{vec}(J(0)) + \left(\left(I_{m-1} \vdots -g_2\right)' \otimes I_{k_f}\right)\text{vec}(A_2(g_1 = 0, g_2))\right] \\ = & \ \text{vec}(J(0))'V_{JJ}^{-\frac{1}{2}} M_{V_{JJ}^{-\frac{1}{2}}\left(\left(I_{m-1} \vdots -g_2\right)' \otimes I_{k_f}\right)} V_{JJ}^{-\frac{1}{2}}\text{vec}(J(0)) \\ = & \ \binom{g_2}{1}' J(0)' \left[\left(\binom{g_2}{1} \otimes I_{k_f}\right)' V_{JJ}\left(\binom{g_2}{1} \otimes I_{k_f}\right)\right]^{-1} J(0)\binom{g_2}{1}, \end{aligned}$$

which used our previous result that $M_H = P_{H_\perp}$. The specification of $Q_{CUE}(g_1 = 0, g_2)$ is identical to that of $Q_{IS}(\varphi)$ in (17). IS equals the minimal value of $Q_{CUE}(g_1 = 0, g_2)$ with respect to g_2 or $Q_{IS}(\varphi)$ with respect to φ :

$$IS = \min_{g_2 \in \mathbb{R}^{(m-1)}} Q_{CUE}(g_1 = 0, g_2) = \min_{\varphi \in \mathbb{R}^{m-1}} Q_{IS}(\varphi).$$

The minimizer of the population generalized reduced rank objective function $Q_{CUE}(y, D)$ and $Q_{CUE}(g, A)$ is invariant with respect to the specification of the involved lower rank matrix. Optimizing over either (y, D) or (g, A), we then obtain $D^*(-y^* \vdots I_m) = A^*(I_m \vdots -g^*)$ which implies $D^* = A^* \left(\begin{smallmatrix} 0 \\ I_{m-1} \end{smallmatrix} \vdots - g^* \right)$ so, since $-D^* y^* = a_1^*$, $-A^* \left(\begin{smallmatrix} 0 \\ I_{m-1} \end{smallmatrix} \vdots - g^* \right) y^* = a_1^*$, which can be written as $A^*(g_1^* l_K^* \vdots -y_1^* + g_2^* y_m^* \vdots \dots \vdots - y_{m-1}^* + g_m^* y_m^*)' = a_1^*$. This can be solved for when $g_1^* \neq 0$. Since IS equals the minimized value of the population objective function when g_1^* is restricted to zero, IS equals the minimal value of the population objective function when θ^* is not identified. Hence, when IS and the minimal value of the population objective function coincide, $g_1^* = 0$ so θ^* is not identified. The difference between IS and the minimal value of the population objective function thus provides a measure of the identification of θ^*.

Proof of Theorem 3: We construct the limit behavior of the two-stage estimator under homoskedasticity, weak misspecification and identification for $m = 1$. It results from the joint limit behavior of its two different elements:

$$\sqrt{N} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{J} - J \end{pmatrix} \underset{d}{\rightarrow} \begin{pmatrix} \psi_\mu \\ \psi_J \end{pmatrix},$$

where $\hat{\mu} = \hat{\mu}(0)$, $\mu = \mu(0)$, $\hat{J} = \hat{J}(0)$, $J = J(0)$ and $(\psi_\mu' \vdots \psi_J')' \sim N(0, \Omega \otimes Q)$. To focus on a setting where both the misspecification and Jacobian are small and perhaps just borderline significant, we use the weak identification and misspecification assumption:

$$J = J_N = \tfrac{1}{\sqrt{N}} Q^{\frac{1}{2}} C \Omega_{JJ}^{\frac{1}{2}}, \ \mu_J = \tfrac{1}{\sqrt{N}} Q^{\frac{1}{2}} a \omega_{\mu\mu.J}^{\frac{1}{2}},$$

with $\mu_J = \mu - J\Omega_{JJ}^{-1}\omega_{J\mu}$, C and a are k-dimensional vectors of constants. Under the small misspecification and Jacobian assumption, the limit behavior of the least squares estimator $\hat{J}$ and $\hat{\mu}_J = \hat{\mu} - \hat{J}\Omega_{JJ}^{-1}\omega_{J\mu}$ are characterized by:

$$\sqrt{N}\hat{J} \underset{d}{\rightarrow} Q^{\frac{1}{2}}(C + \psi_J^*)\Omega_{JJ}^{\frac{1}{2}}, \ \sqrt{N}\hat{\mu}_J \underset{d}{\rightarrow} Q^{\frac{1}{2}}(a + \psi_\mu^*)\omega_{\mu\mu.J}^{\frac{1}{2}},$$

where $\psi_J^* = Q^{-\frac{1}{2}}\psi_J\Omega_{JJ}^{-\frac{1}{2}} \sim N(0, I_k)$, $\psi_\mu^* = Q^{-\frac{1}{2}}(\psi_\mu - \psi_J\Omega_{JJ}^{-1}\omega_{J\mu})\omega_{\mu\mu.J}^{-\frac{1}{2}} \sim N(0, I_k)$, and independent of ψ_J^*.

We next characterize the behavior of the two-stage estimator:

$$\begin{aligned}\hat\theta_{2S} = & -\frac{\hat J'\hat Q^{-1}\hat\mu}{\hat J'\hat Q^{-1}\hat J} = -\frac{\hat J'\hat Q^{-1}(\hat\mu_J + \hat J\Omega_{JJ}^{-1}\omega_{J\mu})}{\hat J' Q^{-1}\hat J} = -\Omega_{JJ}^{-1}\omega_{J\mu} - \frac{\hat J'\hat Q^{-1}\hat\mu_J}{\hat J'\hat Q^{-1}\hat J} \\ = & -\Omega_{JJ}^{-1}\omega_{J\mu} - \frac{\hat J'\hat Q^{-1}[\hat\mu_J - \mu_J + \mu_J]}{\hat J'\hat Q^{-1}\hat J}\end{aligned}$$

so for small values of the Jacobian and misspecification:

$$\begin{aligned}\hat\theta_{2S} + \Omega_{JJ}^{-1}\omega_{J\mu} = & -\frac{(\sqrt N\hat J)'\hat Q^{-1}\left[\sqrt N\hat\mu_J\right]}{(\sqrt N\hat J)'\hat Q^{-1}(\sqrt N\hat J)} \\ \underset{d}{\to} & \ \Omega_{JJ}^{-\frac{1}{2}}\bar\theta_{2s}\omega_{\mu\mu.J}^{\frac{1}{2}} - \frac{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi_J^*)'\left[(C+\psi_J^*)\Omega_{JJ}^{\frac{1}{2}}\Omega_{JJ}^{-\frac{1}{2}}\bar\theta_{2s} + \psi_{\mu.J}^* + a\right]\omega_{\mu\mu.J}^{\frac{1}{2}}}{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi_J^*)'(C+\psi_J^*)\Omega_{JJ}^{\frac{1}{2}\prime}} \\ = & \ \Omega_{JJ}^{-\frac{1}{2}}\bar\theta_{2s}\omega_{\mu\mu.J}^{\frac{1}{2}} - \frac{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi_J^*)'\left[\psi_\mu^* + a + (C+\psi_J^*)\bar\theta_{2s}\right]\omega_{\mu\mu.J}^{\frac{1}{2}}}{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi_J^*)'(C+\psi_J^*)\Omega_{JJ}^{\frac{1}{2}\prime}} \\ = & \ \Omega_{JJ}^{-\frac{1}{2}}\bar\theta_{2s}\omega_{\mu\mu.J}^{\frac{1}{2}} - \frac{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi_J^*)'(\psi_\mu^* + \psi_J^*\bar\theta_{2s})\omega_{\mu\mu.J}^{\frac{1}{2}}}{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi_J^*)'(C+\psi_J^*)\Omega_{JJ}^{\frac{1}{2}\prime}} - \frac{\Omega_{JJ}^{\frac{1}{2}\prime}\psi_J^{*\prime}(a + C\bar\theta_{2s})\omega_{\mu\mu.J}^{\frac{1}{2}}}{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi_J^*)'(C+\psi_J^*)\Omega_{JJ}^{\frac{1}{2}\prime}},\end{aligned}$$

with $\bar\theta_{2s} = -\frac{C'a}{C'C}$ so $C'(a + C\bar\theta_{2s}) = 0$. When $C'a = 0$, we have $\bar\theta_{2s} = 0$ and:

$$\hat\theta_{2S} + \Omega_{JJ}^{-1}\omega_{J\mu} \underset{d}{\to} \ -\frac{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi_J^*)'\psi_\mu^*\omega_{\mu\mu.J}^{\frac{1}{2}}}{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi_J^*)'(C+\psi_J^*)\Omega_{JJ}^{\frac{1}{2}}} - \frac{\Omega_{JJ}^{\frac{1}{2}\prime}\psi_J^{*\prime}a\omega_{\mu\mu.J}^{\frac{1}{2}}}{\Omega_{JJ}^{\frac{1}{2}\prime}(C+\psi_J^*)'(C+\psi_J^*)\Omega_{JJ}^{\frac{1}{2}}}.$$

Proof of Theorem 4: The expression of the LR statistic follows along the lines of Moreira (2003). Imposing $C'a = 0$, $C'C = a'a$ then yields:

$$\begin{aligned}&\mathrm{LR}(\alpha = 0) = \\ &\tfrac{1}{2}\Big[N\Omega_{JJ}^{-1}\hat J(0)'Q^{-1}\hat J(0) - N\omega_{\mu\mu.J}^{-1}\hat\mu_J(0)'Q^{-1}\hat\mu_J(0) + \\ &\sqrt{\left(N\Omega_{JJ}^{-1}\hat J(0)'Q^{-1}\hat J(0) - N\omega_{\mu\mu.J}^{-1}\hat\mu_J(0)'Q^{-1}\hat\mu_J(0)\right)^2 + 4\left(N\Omega_{JJ}^{-\frac{1}{2}}\hat J(0)'Q^{-1}\hat\mu_J(0)\omega_{\mu\mu.J}^{-\frac{1}{2}}\right)^2}\Big] \\ &\underset{d}{\to} \tfrac{1}{2}\Big[(\psi_J^* + C)'(\psi_J^* + C) - (\psi_{\mu.J}^* + a)'(\psi_{\mu.J}^* + a) + \\ &\sqrt{\left((\psi_J^* + C)'(\psi_J^* + C) - (\psi_{\mu.J}^* + a)'(\psi_{\mu.J}^* + a)\right)^2 + 4\left((\psi_J^* + C)'(\psi_{\mu.J}^* + a)\right)^2}\Big] \\ &\underset{d}{\to} \tfrac{1}{2}\Big[\psi_J^{*\prime}\psi_J^* + 2C'\psi_J^* - \psi_{\mu.J}^{*\prime}\psi_{\mu.J}^* - 2a'\psi_{\mu.J}^* + \\ &\sqrt{\left(\psi_J^{*\prime}\psi_J^* + 2C'\psi_J^* - \psi_{\mu.J}^{*\prime}\psi_{\mu.J}^* - 2a'\psi_{\mu.J}^*\right)^2 + 4\left(\psi_J^{*\prime}\psi_{\mu.J}^* + a'\psi_J^* + C'\psi_{\mu.J}^*\right)^2}\Big] \\ &\underset{d}{\to} \tfrac{1}{2}\left(\psi_J^{*\prime}\psi_J^* - \psi_{\mu.J}^{*\prime}\psi_{\mu.J}^*\right) + C'\psi_J^* - a'\psi_{\mu.J}^* + \\ &\tfrac{1}{2}\sqrt{\left(\psi_J^{*\prime}\psi_J^* + 2C'\psi_J^* - \psi_{\mu.J}^{*\prime}\psi_{\mu.J}^* - 2a'\psi_{\mu.J}^*\right)^2 + 4\left(\psi_J^{*\prime}\psi_{\mu.J}^* + a'\psi_J^* + C'\psi_{\mu.J}^*\right)^2}\end{aligned}$$

with $\psi_J^* = Q^{-\frac{1}{2}}\psi_J\Omega_{JJ}^{-\frac{1}{2}}$, $\psi_{\mu.J}^* = Q^{-\frac{1}{2}}\psi_{\mu.J}\omega_{\mu\mu.J}^{-\frac{1}{2}}$.

Algorithm to compute initial estimate of the conditional critical value function of the LR test of H_0^* The algorithm to compute the critical value function of the LR test of H_0^* (33) uses the entier function, [..], and the first and second columns of I_k indicated by $e_{1,k}$ and $e_{2,k}$ resp.:

- Set all elements of the array "sum" to zero and for a range of values of c from $0...c_{\max}$, $a = \sqrt{c}e_{1,k}$, $C = \sqrt{c}e_{2,k}$ 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$= \hat{a}'\hat{a} + \hat{C}'\hat{C}$ 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

- $cv(r, \alpha)$ equals $(1 - \alpha) \times 100$-th percentile of sorted $Z(r, :)$

Proof of Theorem 5: We specify the covariance matrix estimators $\hat{\Omega}$ and $\hat{Q}$ as

$$\hat{\Omega} = \begin{pmatrix} \hat{\omega}_{\mu\mu} & \hat{\omega}_{\mu\mu} \\ \hat{\omega}_{V\mu} & \hat{\Omega}_{VV} \end{pmatrix} = \Omega^{\frac{1}{2}\prime}\dot{\Omega}\Omega^{\frac{1}{2}}, \qquad \hat{\Omega}^{-1} = \Omega^{-\frac{1}{2}}\dot{\Omega}^{-1}\Omega^{-\frac{1}{2}\prime} = \hat{\Omega}^{-\frac{1}{2}\prime}\hat{\Omega}^{-\frac{1}{2}}, \qquad \hat{\Omega}^{-\frac{1}{2}} = \dot{\Omega}^{-\frac{1}{2}}\Omega^{-\frac{1}{2}\prime},$$

$$\hat{Q} = Q^{\frac{1}{2}}\dot{Q}Q^{\frac{1}{2}\prime}, \qquad \hat{Q}^{-1} = Q^{-\frac{1}{2}\prime}\dot{Q}^{-1}Q^{-\frac{1}{2}} = \hat{Q}^{-\frac{1}{2}\prime}\hat{Q}^{-\frac{1}{2}}, \qquad \hat{Q}^{-\frac{1}{2}} = \dot{Q}^{-\frac{1}{2}}Q^{-\frac{1}{2}},$$

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

$J(0)\Omega_{JJ}^{-1}\omega_{J\mu} = \frac{1}{\sqrt{N}}Q^{\frac{1}{2}}a\omega_{\mu\mu.J}^{\frac{1}{2}}$. The different elements of the covariance matrix estimator $\hat{\Omega}$ can then be expressed as:

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

For

$$\begin{aligned}
\dot{\Omega} =& \ \begin{pmatrix} \dot{\omega}_{\mu\mu} & \dot{\omega}_{\mu J} \\ \dot{\omega}_{J\mu} & \dot{\Omega}_{JJ} \end{pmatrix} \\
=& \ \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} \dot{u}_i \\ \dot{V}_i \end{pmatrix}\begin{pmatrix} \dot{u}_i \\ \dot{V}_i \end{pmatrix}' \\
\dot{Q} =& \ \frac{1}{N}\sum_{i=1}^{N}\dot{Z}_i\dot{Z}_i',
\end{aligned}$$

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

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

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

Proof of Theorem 6: Our hypothesis of interest corresponds with a value of IS equal to MISS. Using the specfication from Theorem 3, IS and MISS are defined by:

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

To specify the hypothesis of no identification, H_0 :IS=MISS, more explicitly as a function of a and C, we first conduct a SVD of C :

$$C = U_C S_C V_C',$$

with U_C and V_C $k_f \times k_f$ and $m \times m$ dimensional orthonormal matrices and S_C a $k_f \times m$ dimensional matrix with the singular values in decreasing order on the main diagonal. The SVD enables us to respecify the characteristic polynomial of which MISS is the smallest root:

$$\begin{aligned}
\left|\lambda I_m - (a \vdots C)'(a \vdots C)\right| &= 0 \quad \Leftrightarrow \\
\left|\lambda I_m - \begin{pmatrix} a'a & a'U_C S_C V_C' \\ V_C S_C' U_C' a & V_C S_C' S_C V_C' \end{pmatrix}\right| &= 0 \quad \Leftrightarrow \\
\left|\lambda I_m - \begin{pmatrix} 1 & 0 \\ 0 & V_C \end{pmatrix}' \begin{pmatrix} a'a & a'U_C S_C \\ S_C' U_C' a & S_C' S_C \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & V_C \end{pmatrix}\right| &= 0 \quad \Leftrightarrow \\
\left|\tau I_m - \begin{pmatrix} a'a & a'U_C S_C \\ S_C' U_C' a & S_C' S_C \end{pmatrix}\right| &= 0.
\end{aligned}$$

When IS=MISS, IS, which equals the squared smallest singular value of C, $s_{C,m}^2$, is also the smallest root of the above characteristic polynomial so for $U_C = (U_{C,1} \vdots U_{C,m} \vdots U_{C,2})$, with $U_{C,1} : k_f \times (m-1)$, $U_{C,m} : k_f \times 1$, $U_{C,2} : k_f \times (k_f - m)$, it implies that:

$U_{C,m}' a = 0$ and the smallest characteristic root of: $\left|\rho I_m - \begin{pmatrix} a'a & a'U_{C,1} S_{C,1} \\ S_{C,1}' U_{C,1}' a & S_{C,1}' S_{C,1} \end{pmatrix}\right| = 0,$

is larger than or equal to $s_{C,m}^2$ (=IS) for $S_C = \begin{pmatrix} S_{C,1} & 0 \\ 0 & s_{C,m} \\ 0 & 0 \end{pmatrix}$,

with $S_{C,1} : (m-1) \times (m-1)$, $s_{C,m} : 1 \times 1$. We note that, because $U_{C,m}' a = 0$, the matrix in the above characteristic polynomial comprises of the matrix in the previous characteristic polynomial with the rows and columns associated with the smallest singular value of C, s_{cm}, removed.

We next construct a bounding expression for the smallest root of the above characteristic polynomial. This smallest root is larger than or equal to the one which results when we replace all singular values in $S_{C,1}$ by its smallest element on the diagonal, so $S_{C,1} = s_{C,(m-1)} \binom{I_{m-1}}{0}$, with $s_{C,(m-1)}$ the $(m-1)$-st element on the main diagonal of $S_{C,1}$, see

Theorem 7 in Kleibergen (2007). The resulting characteristic polynomial reads:

$$\left|\rho I_m - \begin{pmatrix} a'a & a'U_{C,1}\binom{I_{m-1}}{0}s_{C,(m-1)} \\ s_{C,m-1}\binom{I_{m-1}}{0}'U_{C,1}'a & s_{C,(m-1)}^2 I_{m-1}\end{pmatrix}\right| =$$
$$(\rho - s_{C,(m-1)}^2)^{m-2}(\rho^2 - \tau(a^{*\prime}a^* + b'b + s_{C,(m-1)}^2) + s_{C,(m-1)}^2 b'b),$$

with $a^* = U_{C,1}'a$, $b = U_{C,2}'a$. Kleibergen (2007) shows that the smallest root results from the second part of the above polynomial which is quadratic so we have the closed form solution:

$$\rho_{\min} = \tfrac{1}{2}\left(a^{*\prime}a^* + b'b + s_{C,(m-1)}^2 - \sqrt{(s_{C,(m-1)}^2 + a^{*\prime}a^* - b'b)^2 + 4a^{*\prime}a^* b'b}\right) \geq b'b.$$

Hence, $b'b$ is a sharp lower bound on the smallest root of the characteristic polynomial $\left|\rho I_m - \begin{pmatrix} a'a & a'U_{C,1}S_{C,1} \\ S_1'U_{C,1}'a & S_{C,1}'S_{C,1}\end{pmatrix}\right| = 0$ so a sufficient condition for IS=MISS is:

$$U_{C,m}'a = 0 \text{ and } b'b \geq s_{C,m}^2 = \text{IS}.$$

Proof of Theorem 7: Using the proof of Theorems 5 and the SVD from Theorem 6:

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

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

Proof of Theorem 8: For $m = 1$, the population CUE objective function is:

$$Q_{CUE}(\alpha) = (J+\mu\alpha)'\left[V_{JJ} + \alpha(V_{\mu J} + V_{\mu J}') + \alpha^2 V_{\mu\mu}\right]^{-1}(J+\mu\alpha),$$

with $J = J(0)$ and $\mu = \mu(0)$. The derivative of $Q_{CUE}(\alpha)$ then becomes:

$$\begin{aligned}
\tfrac{1}{2}\tfrac{\partial}{\partial\alpha}Q_{CUE}(\alpha) = \; & \mu'\left[V_{JJ} + \alpha(V_{\mu J} + V_{\mu J}') + \alpha^2 V_{\mu\mu}\right]^{-1}(J+\mu\alpha) - \\
& (J+\mu\alpha)'\left[V_{JJ} + \alpha(V_{\mu J} + V_{\mu J}') + \alpha^2 V_{\mu\mu}\right]^{-1}\left[\tfrac{1}{2}(V_{\mu J} + V_{\mu J}') + \alpha V_{\mu\mu}\right] \\
& \left[V_{JJ} + \alpha(V_{\mu J} + V_{\mu J}') + \alpha^2 V_{\mu\mu}\right]^{-1}(J+\mu\alpha) \\
= \; & \Big(\mu - \left[V_{\mu J} + \alpha V_{\mu\mu}\right]\left[V_{JJ} + \alpha(V_{\mu J} + V_{\mu J}') + \alpha^2 V_{\mu\mu}\right]^{-1}(J+\mu\alpha)\Big)' \\
& \left[V_{JJ} + \alpha(V_{\mu J} + V_{\mu J}') + \alpha^2 V_{\mu\mu}\right]^{-1}(J+\mu\alpha).
\end{aligned}$$

To construct the Hessian, we use that:

$$\begin{aligned}
& \tfrac{\partial}{\partial\alpha}\Big(\mu - \left[V_{\mu J} + \alpha V_{\mu\mu}\right]\left[V_{JJ} + \alpha(V_{\mu J} + V_{\mu J}') + \alpha^2 V_{\mu\mu}\right]^{-1}(J+\mu\alpha)\Big) \\
= \; & -\left[V_{\mu J} + \alpha V_{\mu\mu}\right]\left[V_{JJ} + \alpha(V_{\mu J} + V_{\mu J}') + \alpha^2 V_{\mu\mu}\right]^{-1}\mu - \\
& V_{\mu\mu}\left[V_{JJ} + \alpha(V_{\mu J} + V_{\mu J}') + \alpha^2 V_{\mu\mu}\right]^{-1}(J+\mu\alpha) + \\
& \left[V_{\mu J} + \alpha V_{\mu\mu}\right]\left[V_{JJ} + \alpha(V_{\mu J} + V_{\mu J}') + \alpha^2 V_{\mu\mu}\right]^{-1}\left[(V_{\mu J} + V_{\mu J}') + 2\alpha V_{\mu\mu}\right] \\
& \left[V_{JJ} + \alpha(V_{\mu J} + V_{\mu J}') + \alpha^2 V_{\mu\mu}\right]^{-1}(J+\mu\alpha)
\end{aligned}$$

so the Hessian becomes:

$$\begin{aligned}
\tfrac{1}{2}\tfrac{\partial^2}{(\partial\alpha)^2}Q_{CUE}(\alpha) = \quad & \Big\{-\left[V_{\mu J}+\alpha V_{\mu\mu}\right]\left[V_{JJ}+\alpha(V_{\mu J}+V_{\mu J}')+\alpha^2 V_{\mu\mu}\right]^{-1}\mu- \\
& V_{\mu\mu}\left[V_{JJ}+\alpha(V_{\mu J}+V_{\mu J}')+\alpha^2 V_{\mu\mu}\right]^{-1}(J+\mu\alpha)+ \\
& \left[V_{\mu J}+\alpha V_{\mu\mu}\right]\left[V_{JJ}+\alpha(V_{\mu J}+V_{\mu J}')+\alpha^2 V_{\mu\mu}\right]^{-1}\left[(V_{\mu J}+V_{\mu J}')+2\alpha V_{\mu\mu}\right] \\
& \left[V_{JJ}+\alpha(V_{\mu J}+V_{\mu J}')+\alpha^2 V_{\mu\mu}\right]^{-1}(J+\mu\alpha)\Big\}' \\
& \left[V_{JJ}+\alpha(V_{\mu J}+V_{\mu J}')+\alpha^2 V_{\mu\mu}\right]^{-1}(J+\mu\alpha)- \\
& \Big(\mu-\left[V_{\mu J}+\alpha V_{\mu\mu}\right]\left[V_{JJ}+\alpha(V_{\mu J}+V_{\mu J}')+\alpha^2 V_{\mu\mu}\right]^{-1}(J+\mu\alpha)\Big)' \\
& \left[V_{JJ}+\alpha(V_{\mu J}+V_{\mu J}')+\alpha^2 V_{\mu\mu}\right]^{-1}\left[(V_{\mu J}+V_{\mu J}')+2\alpha V_{\mu\mu}\right] \\
& \left[V_{JJ}+\alpha(V_{\mu J}+V_{\mu J}')+\alpha^2 V_{\mu\mu}\right]^{-1}(J+\mu\alpha)+ \\
& \Big(\mu-\left[V_{\mu J}+\alpha V_{\mu\mu}\right]\left[V_{JJ}+\alpha(V_{\mu J}+V_{\mu J}')+\alpha^2 V_{\mu\mu}\right]^{-1}(J+\mu\alpha)\Big)' \\
& \left[V_{JJ}+\alpha(V_{\mu J}+V_{\mu J}')+\alpha^2 V_{\mu\mu}\right]^{-1}\mu.
\end{aligned}$$

For the minimum of $Q_{CUE}(\alpha)$ to be at $\alpha=0$, the first order condition has to apply at $\alpha=0$ so:

$$\begin{aligned}
\tfrac{1}{2}\tfrac{\partial}{\partial\alpha}Q_{CUE}(\alpha=0)=0 \quad & \Leftrightarrow \\
\left(\mu-V_{\mu J}V_{JJ}^{-1}J\right)'V_{JJ}^{-1}J=0 \quad & \Leftrightarrow \\
\mu_J'V_{JJ}^{-1}J=0, &
\end{aligned}$$

with $\mu_J=\mu-V_{\mu J}V_{JJ}^{-1}J$. The Hessian at $\alpha=0$ is:

$$\begin{aligned}
\tfrac{1}{2}\tfrac{\partial^2}{(\partial\alpha)^2}Q_{CUE}(\alpha=0) = \quad & \left\{-V_{\mu J}V_{JJ}^{-1}\mu-V_{\mu\mu}V_{JJ}^{-1}J+V_{\mu J}V_{JJ}^{-1}\left[V_{\mu J}+V_{\mu J}'\right]V_{JJ}^{-1}J\right\}'V_{JJ}^{-1}J- \\
& \left(\mu-V_{\mu J}V_{JJ}^{-1}J\right)'V_{JJ}^{-1}(V_{\mu J}+V_{\mu J}')V_{JJ}^{-1}J+\left(\mu-V_{\mu J}V_{JJ}^{-1}J\right)'V_{JJ}^{-1}\mu \\
= \quad & \left\{-V_{\mu J}V_{JJ}^{-1}(\mu-V_{\mu J}V_{JJ}^{-1}J)-(V_{\mu\mu}-V_{\mu J}V_{JJ}^{-1}V_{\mu J}')V_{JJ}^{-1}J\right\}'V_{JJ}^{-1}J- \\
& \left(\mu-V_{\mu J}V_{JJ}^{-1}J\right)'V_{JJ}^{-1}V_{\mu J}'V_{JJ}^{-1}J+\left(\mu-V_{\mu J}V_{JJ}^{-1}J\right)'V_{JJ}^{-1}(\mu-V_{\mu J}V_{JJ}^{-1}J) \\
= \quad & \mu_J'V_{JJ}^{-1}\mu_J-J'V_{JJ}^{-1}(V_{\mu\mu}-V_{\mu J}V_{JJ}^{-1}V_{\mu J}')V_{JJ}^{-1}J-2\mu_J'V_{JJ}^{-1}V_{\mu J}'V_{JJ}^{-1}J.
\end{aligned}$$

If we next define

$$J=J(0)=\tfrac{1}{\sqrt{N}}V_{JJ}^{\frac{1}{2}}C,\ \ \mu_J=\tfrac{1}{\sqrt{N}}V_{\mu\mu.J}^{\frac{1}{2}}a,$$

for $V_{\mu\mu.J}=V_{\mu\mu}-V_{\mu J}V_{JJ}^{-1}V_{\mu J}'$, we can express the first order condition and (scaled) Hessian as:

$$\begin{aligned}
\tfrac{N}{2}\tfrac{\partial}{\partial\alpha}Q_{CUE}(\alpha=0) = \quad & a'V_{\mu\mu.J}^{\frac{1}{2}\prime}V_{JJ}^{-1}V_{JJ}^{\frac{1}{2}}C=a'V_{\mu\mu.J}^{\frac{1}{2}\prime}V_{JJ}^{-\frac{1}{2}\prime}C \\
\tfrac{N}{2}\tfrac{\partial^2}{(\partial\alpha)^2}Q_{CUE}(\alpha=0) = \quad & a'V_{\mu\mu.J}^{\frac{1}{2}\prime}V_{JJ}^{-1}V_{\mu\mu.J}^{\frac{1}{2}}a-C'V_{JJ}^{\frac{1}{2}\prime}V_{JJ}^{-1}V_{\mu\mu.J}V_{JJ}^{-1}V_{JJ}^{\frac{1}{2}}C-2a'V_{\mu\mu.J}^{\frac{1}{2}\prime}V_{JJ}^{-1}V_{\mu J}'V_{JJ}^{-1}V_{JJ}^{\frac{1}{2}}C \\
= \quad & a'V_{\mu\mu.J}^{\frac{1}{2}\prime}V_{JJ}^{-1}V_{\mu\mu.J}^{\frac{1}{2}}a-C'V_{JJ}^{-\frac{1}{2}}V_{\mu\mu.J}V_{JJ}^{-\frac{1}{2}\prime}C-2a'V_{\mu\mu.J}^{\frac{1}{2}\prime}V_{JJ}^{-1}V_{\mu J}'V_{JJ}^{-\frac{1}{2}\prime}C.
\end{aligned}$$