Kleibergen, Frank; Zhan, Zhaoguo

Article

Double robust inference for continuous updating GMM

Quantitative Economics

Provided in Cooperation with:
The Econometric Society

Suggested Citation: Kleibergen, Frank; Zhan, Zhaoguo (2025) : Double robust inference for continuous updating GMM, Quantitative Economics, ISSN 1759-7331, The Econometric Society, New Haven, CT, Vol. 16, Iss. 1, pp. 295-327, https://doi.org/10.3982/QE2347

This Version is available at:
https://hdl.handle.net/10419/320333

Quantitative Economics 16 (2025), 295–327 1759-7331/20250295

Double robust inference for continuous updating GMM

Frank Kleibergen
Amsterdam School of Economics, University of Amsterdam

Zhaoguo Zhan
Coles College of Business, Kennesaw State University

We propose the double robust Lagrange multiplier (DRLM) statistic for testing hypotheses specified on the minimizer of the population continuous updating objective function. The (bounding) χ^2 limiting distribution of the DRLM statistic is robust to both misspecification and weak identification, hence its name. The minimizer is the so-called pseudo-true value, which equals the true value of the structural parameter under correct specification. To emphasize its importance for applied work where misspecification and weak identification are common, we use the DRLM test to analyze: the risk premia in Adrian et al. (2014) and He et al. (2017); the structural parameters in a nonlinear asset pricing model with constant relative risk aversion.

Keywords. Weak identification, misspecification, robust inference, Lagrange multiplier.

JEL classification. C12, C18, G12.

1. Introduction

A little more than 20 years ago, inference procedures for analyzing possibly weakly identified structural parameters using the generalized method of moments (GMM) of Hansen (1982) were mostly lacking. Since then, huge progress has been made to develop such procedures; see, for example, Staiger and Stock (1997), Dufour (1997), Stock and Wright (2000), Kleibergen (2002, 2005, 2009), Moreira (2003), Andrews and Cheng (2012), Andrews and Mikusheva (2016a, 2016b), and Han and McCloskey (2019). At present, we therefore have a variety of so-called weak identification robust inference methods. Given the prevalence of weak identification in applied work, a lot of emphasis has also been put in raising awareness among practitioners; see, for example, Kleibergen and Mavroeidis (2009), Beaulieu, Dufour, and Khalaf (2013), Mavroeidis, Plagborg-Moller, and Stock (2014), Andrews, Stock, and Sun (2019), and Kleibergen and Zhan (2020).

When there is no so-called true value of the structural parameters where the GMM moment conditions exactly hold, the structural model is rendered misspecified, and GMM estimators provide inconsistent estimates of the true value. Early research on

Frank Kleibergen: f.r.kleibergen@uva.nl
Zhaoguo Zhan: zzhan@kennesaw.edu
The research of Frank Kleibergen has been funded partly by the NWO (Dutch Research Council) Grant 401.21.EB.002: "Double robust inference for structural economic models (DRISEM)".

 https://doi.org/10.3982/QE2347

misspecification focused on inference on the true value by characterizing expressions for the bias and standard errors of inconsistent estimators; see, for example, Maasoumi and Phillips (1982) and Maasoumi (1990). Instead of focusing on the unattainable true value, recent work on misspecification analyzes the so-called pseudo-true value, which is the minimizer of the population objective function; see, for example, White (1994). For practical relevance of the pseudo-true value see, for example, Kan, Robotti, and Shanken (2013) who use the pseudo-true value for evaluating asset pricing models. The pseudo-true value depends on the population objective function at hand, so different objective functions can lead to distinct pseudo-true values. Hall and Inoue (2003), for example, develop inference methods for the pseudo-true value of the two-step GMM estimator, while Hansen and Lee (2021)do so for an iterated GMM estimator. We use the minimizer of the population continuous updating estimator (CUE) objective function of Hansen, Heaton, and Yaron (1996) as the pseudo-true value because of its invariance properties and since weak identification robust inference procedures lead to inference that is centered around it.

In case of misspecification, weak identification robust inference procedures for testing hypotheses specified on the pseudo-true value however become size distorted for just small amounts of misspecification. This would not sound as much of a problem if it was possible to efficiently detect such misspecification. This is not so since misspecification tests, like the Sargan–Hansen test (Sargan (1958) and Hansen (1982)), are virtually powerless in settings of joint misspecification and weak identification; see Gospodinov, Kan, and Robotti (2017). Weak identification robust inference procedures thus came about to overcome the general critique of nonrobustness of traditional inference procedures to varying identification strengths (see, e.g., Staiger and Stock (1997) and Dufour (1997)), but are similarly nonrobust to misspecification.

While weak identification robust inference procedures are size distorted when misspecification is present, the misspecification robust inference procedures proposed by, for example, Hall and Inoue (2003), Kan, Robotti, and Shanken (2013), Lee (2018), and Hansen and Lee (2021), provide misspecification robust covariance matrix estimators to conduct Wald-based inference. Because the covariance matrix estimators are finite by construction, the resulting Wald-based tests cannot lead to unbounded confidence sets, which as shown by Dufour (1997), is a necessary condition for size correct testing in settings with potential identification failure. Hence, except for certain specifications of the structural parameter (see, e.g., Gospodinov, Kan, and Robotti (2014)),[1] these misspecification robust tests are potentially size distorted under weak identification.

One of the first to emphasize the empirical relevance of misspecification in the presence of weak (or no) identification were Kan and Zhang (1999). With the surge in applied work on structural estimation, awareness of misspecification has grown further. In asset pricing models, for example, it is now generally accepted that misspecification, alongside weak identification, is an important empirical issue; see, for example, Kan, Robotti,

[1] Gospodinov, Kan, and Robotti (2014) show that the Wald/t test for a zero risk premium in a linear asset pricing model with useless partialled out factors, which uses their adjusted standard error, is bounded by a χ^2 distribution and, therefore, size correct for the hypothesis of interest.

and Shanken (2013) and Kleibergen and Zhan (2020). Kan, Robotti, and Shanken (2013) therefore developed misspecification robust t-statistics for the Fama–MacBeth (1973) two-pass estimator, that is, the most commonly used estimator of the risk premia in linear asset pricing models. Similarly, in case of heterogeneous treatment effects, the local average treatment effects that result for different instruments can be distinct (see Imbens and Angrist (1994)), making the overidentified linear instrumental variables regression model that uses all instruments misspecified. Evdokimov and Kolesar (2018) and Lee (2018) therefore analyze misspecification robust tests on the treatment effect resulting from using such multiple instruments. These misspecification robust tests are however not robust to weak identification, so identical to the weak identification robust inference procedures, they cannot deal with the empirically relevant setting of both misspecification and weak identification. Yet the aforementioned weak identification and misspecification literature highlight that these two issues have to be taken into account in order to conduct valid inference on the structural parameters of interest.

We extend the weak identification robust score or Lagrange multiplier (KLM) test from Kleibergen (2002, 2005, 2009) to a double robust Lagrange multiplier (DRLM) test, which can also be interpreted as the misspecification robust version of the KLM test. The DRLM test is size correct and robust to both misspecification and weak identification, hence its name. The DRLM statistic is a quadratic form of the score function, which equals zero at all stationary points of the CUE sample objective function. This is also the case for the KLM statistic and explains the power problems of the KLM test; see, for example, Andrews, Moreira, and Stock (2006). To overcome the power problems of the KLM test, the KLM statistic can be combined in a conditional or unconditional manner with the Anderson–Rubin (1949, AR) statistic; see, for example, Moreira (2003) and Andrews (2016). Andrews, Moreira, and Stock (2006) show that the conditional likelihood ratio (LR) test of Moreira (2003) provides the optimal manner of combining these statistics for the homoskedastic linear instrumental variables regression model with one included endogenous variable. In case of misspecification, it is however not obvious how to improve the power of the DRLM test by such combination arguments, since the statistics with which the DRLM statistic is to be combined to improve power, have non-central limiting distributions with parameters that cannot be consistently estimated under misspecification. We therefore show that the power of the DRLM test can be improved by exploiting the derivative property of the DRLM statistic.

The rest of the paper is organized as follows. In Section 2, we discuss continuous updating GMM with misspecification and propose the DRLM test. To illustrate the size correctness of the DRLM test, we conduct a simulation experiment using linear moment equations in Section 3. A power study of the DRLM test and weak identification robust tests is also presented in Section 3. It shows that weak identification robust tests on the pseudo-true value of the structural parameters are size distorted for just small amounts of misspecification while the DRLM test is not. It also proposes power improvement and shows that the resulting test has generally good power. Section 4 applies the DRLM test to risk premia using asset pricing data from Adrian, Etula, and Muir (2014) and He, Kelly, and Manela (2017). Section 5 extends the DRLM test to subvector inference, strong misspecification, which is dealt with by an additional component in the weight matrix,

and nonlinear moment equations from a nonlinear asset pricing model with constant relative risk aversion. Section 6 concludes. Technical details and additional results are relegated to the Supplemental Appendix (Kleibergen and Zhan (2024b)) .

2. DRLM test for possibly misspecified GMM

2.1 *Setup*

We analyze the $m \times 1$ parameter vector $\theta = (\theta_1 \dots \theta_m)'$ whose parameter region is the $\mathbb{R}^m$. The $k_f \times 1$ dimensional function $f(\theta, X_t)$ is a continuously differentiable function of the parameter vector θ and a Borel measurable function of a data vector X_t, which is observed for time/individual t. Since we focus on misspecification, the model is overidentified, that is, there are more moment equations than structural parameters so $k_f > m$. The population moment function of $f(\theta, X_t)$ equals $\mu_f(\theta)$:

$$E_X\big(f(\theta, X_t)\big) = \mu_f(\theta), \tag{1}$$

with $\mu_f(\theta)$ a k_f-dimensional continuously differentiable function. Unlike regular GMM (see Hansen (1982)) , we do not request that there is a specific value of θ, say θ_0, at which $\mu_f(\theta_0) = 0$. Our analysis thus differs from a recent one proposed by Cheng, Dou, and Liao (2022), who construct a model selection procedure for evaluating potentially misspecified models with possibly weakly identified structural parameters, which explicitly uses a set of base moments contained in all considered models that are guaranteed to hold.

We analyze θ using the continuous updating setting of Hansen, Heaton, and Yaron (1996). We use it because of its invariance properties and since it leads to inference based on identification robust statistics in standard GMM; see, for example, Stock and Wright (2000) and Kleibergen (2005). For asset pricing studies, in particular, Peñaranda and Sentana (2015) recommend the continuous updating setting over two-step or iterated GMM procedures. The accompanying population continuous updating objective function is

$$Q_p(\theta) = \mu_f(\theta)' V_{ff}(\theta)^{-1} \mu_f(\theta), \tag{2}$$

with $V_{ff}(\theta)$ the covariance matrix of the sample moment:[2] $f_T(\theta, X) = \frac{1}{T}\sum_{t=1}^{T} f(\theta, X_t)$,

$$V_{ff}(\theta) = \lim_{T\to\infty} E\big[T\big(f_T(\theta, X) - \mu_f(\theta)\big)\big(f_T(\theta, X) - \mu_f(\theta)\big)'\big], \tag{3}$$

so $f_T(\theta, X)$ is the sample analog of $\mu_f(\theta)$ for a data set of T observations: $X_t, t = 1, \dots, T$. We define the pseudo-true value of θ, θ^* as the minimizer of the population objective function:

$$\theta^* = \arg\min_{\theta\in\mathbb{R}^m} Q_p(\theta). \tag{4}$$

[2]Throughout the paper, we use recentered covariance matrices while the continuous updating estimator is identical under a recentered or uncentered version of the same covariance matrix estimator; see Hansen and Lee (2021).

Thus, the pseudo-true value θ^* equals the true value of θ in correctly specified models. In misspecified models, the interpretation of θ^* depends on the context: for example, in asset pricing studies, Kan, Robotti, and Shanken (2013) interpret the pseudo-true value of risk premia as the value that minimizes pricing errors.

The sample objective function for the CUE is

$$\hat{Q}_s(\theta) = f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X), \tag{5}$$

with $\hat{V}_{ff}(\theta)$ a consistent estimator of $V_{ff}(\theta)$, $\hat{V}_{ff}(\theta) \underset{p}{\to} V_{ff}(\theta)$, so the CUE, $\hat{\theta}$ is

$$\hat{\theta} = \arg\min_{\theta \in \mathbb{R}^m} \hat{Q}_s(\theta). \tag{6}$$

At the CUE, the score or derivative of the CUE objective function equals zero:[3]

$$\hat{s}(\theta) = \frac{1}{2}\frac{\partial}{\partial \theta'}\hat{Q}_s(\theta) = f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta), \tag{7}$$

where

$$\hat{D}(\theta) = q_T(\theta, X) - \big[\hat{V}_{q_1 f}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X) \ldots \hat{V}_{q_m f}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X)\big], \tag{8}$$

with $q_T(\theta, X) = \frac{\partial f_T(\theta, X)}{\partial \theta'}|_\theta = (q_{1T}(\theta) \ldots q_{mT}(\theta))$, $J(\theta) = \frac{\partial}{\partial \theta'}\mu_f(\theta) = (J_1(\theta) \ldots J_m(\theta))$, $\hat{V}_{q_i f}(\theta)$ is a consistent estimator of $V_{q_i f}(\theta) = \lim_{T\to\infty} E[T(q_{iT}(\theta) - J_i(\theta))(f_T(\theta, X) - \mu_f(\theta))']$, $i = 1, \ldots, m$. Correspondingly, the population counterpart of the sample score $\hat{s}(\theta)$ in (7) is

$$s(\theta) = \frac{1}{2}\frac{\partial}{\partial \theta'}Q_p(\theta) = \mu_f(\theta)'V_{ff}(\theta)^{-1}D(\theta), \tag{9}$$

where $D(\theta)$ is the recentered Jacobian:

$$D(\theta) = J(\theta) - \big[V_{q_1 f}(\theta)V_{ff}(\theta)^{-1}\mu_f(\theta) \ldots V_{q_m f}(\theta)V_{ff}(\theta)^{-1}\mu_f(\theta)\big]. \tag{10}$$

Our proposed double robust Lagrange multiplier statistic on the pseudo-true value equals a quadratic form of the sample score $\hat{s}(\theta)$ in (7), which involves the product of the sample moment and a recentered estimator of its Jacobian. While the sample moment $f_T(\theta, X)$ is driven by the magnitude of misspecification, the recentered Jacobian estimator $\hat{D}(\theta)$ reflects the strength of identification. Thus, both misspecification and weak identification are to be accounted for when we develop the explicit expression of the double robust Lagrange multiplier statistic based on $\hat{s}(\theta)$ in (7).

We focus on weak (local) misspecification in the sense that at the pseudo-true value θ^*, there exists

$$\tilde{\mu}_f(\theta^*) = \lim_{T\to\infty} \sqrt{T}\mu_f(\theta^*), \tag{11}$$

[3]The construction of the derivative is in Kleibergen (2005) and also in Lemmas 3 and 4 in the Supplemental Appendix.

for $\tilde{\mu}_f(\theta^*)$ a finite valued k_f-dimensional vector, which could equal zero. Thus, $\mu_f(\theta^*)$ is allowed to depend on T with the dependence kept implicit for the purpose of notational simplicity. This treatment thus differs from a strong misspecification case for which $\mu_f(\theta^*)$ is considered a fixed nonzero vector. We discuss the strong misspecification case later in Section 5.2 as an extension. The proposed double robust Lagrange multiplier test (for a properly specified weight matrix) applies to both of these cases.

Analogous to the treatment of $\mu_f(\theta^*)$ in (11), the weak identification literature widely uses the so-called weak instrument asymptotics (see, e.g., Staiger and Stock (1997)), which in our context amounts to using a local to zero sequence for $D(\theta^*)$:

$$\tilde{D}(\theta^*) = \lim_{T\to\infty} \sqrt{T} D(\theta^*), \tag{12}$$

for $\tilde{D}(\theta^*)$ a finite valued $k_f \times m$ dimensional matrix, which could be of lower rank. Although we initially use (12) for ease of exposition, it is worth noting that the double robust Lagrange multiplier test is equally applicable under strong identification as we show when simultaneously discussing strong misspecification in Section 5.2. Put differently, whether $D(\theta^*)$ is local to zero or fixed makes no difference to the double robust Lagrange multiplier statistic, so we allow for all strengths of identification.

2.2 *Example: The linear IV regression model*

For illustration, we use a linear instrumental variables (IV) regression model:[4]

$$\begin{aligned} y_t =& \beta' x_t + \varepsilon_t, \\ x_t =& \Pi' z_t + v_t, \end{aligned} \tag{13}$$

where β and Π are $m \times 1$ and $k_f \times m$ dimensional matrices containing unknown parameters, y_t and x_t are the scalar and $m \times 1$ dimensional endogenous variables, z_t is the $k_f \times 1$ dimensional instrumental variables, ε_t and v_t are the scalar and $m \times 1$ dimensional errors.

Its population moment function reads

$$\begin{aligned} \mu_f(\beta) &= \sigma_{zy} - \Sigma_{zx}\beta \\ &= \mu_f(0) + J(0)\theta, \end{aligned} \tag{14}$$

where, assuming that the observations over the individuals are i.i.d., $\sigma_{zy} = E((z_t - \mu_z)(y_t - \mu_y))$, $\Sigma_{zx} = E((z_t - \mu_z)(x_t - \mu_x)') = Q_{zz}\Pi$, $Q_{zz} = E((z_t - \mu_z)(z_t - \mu_z)')$, $\mu_y = E(y_t)$, $\mu_x = E(x_t)$, $\mu_z = E(z_t)$. The last line of (14) uses the GMM notation from (1) and (10) so $\mu_f(0) = \sigma_{zy}$, $J(0) = -\Sigma_{zx}$, and $\theta = \beta$. Misspecification occurs when the structural error ε_t is correlated with the instruments z_t. There is then no value of β at which the

[4]For expository purposes, we only discuss a simplified version of the linear IV regression model without so-called, included exogenous variables. We specify the linear IV regression model, and also the linear asset pricing model discussed later, using the notation generically used in the literature, and also show their connection to our earlier generic GMM notation.

population moment function (14) is equal to zero. On the other hand, weak identification occurs when instruments z_t are only weakly correlated with the endogenous x_t, so the full rank condition of Σ_{zx} is at risk.

The population objective function of the CUE for the linear IV model with homoskedastic errors is

$$Q_p(\beta) = \frac{1}{\omega_{uu} - 2\omega_{uv}\beta + \beta'\Omega_{vv}\beta}(\sigma_{zy} - \Sigma_{zx}\beta)'Q_{zz}^{-1}(\sigma_{zy} - \Sigma_{zx}\beta), \tag{15}$$

where $\Omega_{uv} = \left(\begin{smallmatrix}\omega_{uu} & \omega_{uv}\\ \omega_{vu} & \Omega_{vv}\end{smallmatrix}\right) = \operatorname{cov}(u_t, v_t)$, for $u_t = \varepsilon_t + v_t'\beta$. The CUE then corresponds with the so-called, limited information maximum likelihood (LIML) estimator, so we can use the well-known k-class notation of the LIML estimator (see, e.g., Hausman (1983) and Andrews (2019)) to express the pseudo-true value as

$$\beta^* = \arg\min_{\beta\in\mathbb{R}^m} Q_p(\beta) = \left(\Sigma_{zx}'Q_{zz}^{-1}\Sigma_{zx} - \tau_{\min}\Omega_{vv}\right)^{-1}\left(\Sigma_{zx}'Q_{zz}^{-1}\sigma_{zy} - \tau_{\min}\omega_{vu}\right), \tag{16}$$

with $\tau_{\min} = \min_{\beta\in\mathbb{R}^m} Q_p(\beta)$.[5] We note that different from the usual k-class notation for $\tau_{\min} = 0$ the pseudo-true value corresponds with the pseudo-true value of the two-stage least squares estimator; similarly for $\tau_{\min} = -1$, the pseudo-true value corresponds with that of the least squares estimator.

For the linear IV regression model, the double robust Lagrange multiplier test is designed for testing hypotheses on β^*, and it addresses weak identification and misspecification simultaneously. The test is based on the score or derivative of $Q_p(\beta)$, which equals zero at β^*. The quadratic form of the sample score constitutes the test statistic, whose explicit expression will be provided using the generic GMM notation in Section 2.4.

2.3 *Assumptions*

We state the assumptions needed for constructing the large sample behavior of test statistics centered around the CUE for the generic GMM setting. These assumptions concern the components of the sample score $\hat{s}(\theta) = f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)$.

We first make Assumption 1 as in Kleibergen (2005) except that it concerns the large sample behavior of the sample moments and their derivative at the pseudo-true value θ^* instead of the true value.

ASSUMPTION 1. *The $k_f \times 1$ dimensional derivative of $f_t(\theta) = f(\theta, X_t)$ with respect to θ_i,*

$$q_{it}(\theta) = \frac{\partial f_t(\theta)}{\partial \theta_i} : k_f \times 1, \quad i = 1, \ldots, m \tag{17}$$

is such that the joint limit behavior of the sums of the series $\bar{f}_t(\theta) = f_t(\theta) - E(f_t(\theta))$ and $\bar{q}_t(\theta) = (\bar{q}_{1t}(\theta)' \ldots \bar{q}_{mt}(\theta)')'$, with $\bar{q}_{it}(\theta) = q_{it}(\theta) - E(q_{it}(\theta))$, accords with the central limit

[5] The minimal value $\tau_{\min}$ equals the smallest root of $|\tau\Omega_{uv} - (\sigma_{zy}\vdots\Sigma_{zx})'Q_{zz}^{-1}(\sigma_{zy}\vdots\Sigma_{zx})| = 0$.

theorem at $\theta = \theta^*$, where θ^* is the minimizer of the continuous updating population objective function:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix}\bar{f}_t(\theta^*)\\ \bar{q}_t(\theta^*)\end{pmatrix} \underset{d}{\to} \begin{pmatrix}\psi_f(\theta^*)\\ \psi_q(\theta^*)\end{pmatrix} \sim N\big(0, V(\theta^*)\big), \tag{18}$$

where $\psi_f : k_f \times 1$, $\psi_q : k_\theta \times 1$, $k_\theta = mk_f$, and $V(\theta^*)$ is a positive semidefinite symmetric $(k_f + k_\theta) \times (k_f + k_\theta)$ matrix,

$$V(\theta^*) = \begin{pmatrix}V_{ff}(\theta^*) & V_{fq}(\theta^*)\\ V_{qf}(\theta^*) & V_{qq}(\theta^*)\end{pmatrix}, \tag{19}$$

with $V_{qf}(\theta^*) = V_{fq}(\theta^*)' = (V_{q_1 f}(\theta^*)' \ldots V_{q_m f}(\theta^*)')'$, $V_{qq}(\theta^*) = (V_{q_i q_j}(\theta^*)) : i, j = 1, \ldots, m$; $V_{ff}(\theta^*)$, $V_{q_i f}(\theta^*)$, $V_{q_i q_j}(\theta^*)$ are $k_f \times k_f$ dimensional matrices for $i, j = 1, \ldots, m$, and

$$V(\theta^*) = \lim_{T\to\infty} \operatorname{var}\left(\sqrt{T}\begin{pmatrix}f_T(\theta^*, X)\\ \operatorname{vec}\big(q_T(\theta^*, X)\big)\end{pmatrix}\right). \tag{20}$$

Assumption 1 requests a joint central limit theorem to hold at the pseudo-true value for the sample moments and their derivative. It is satisfied under mild conditions, which are listed in Kleibergen (2005) like, for example, finite rth moments for $r > 2$ in case of i.i.d. data, mixing conditions for the sample moments in case of time-series data. Allowing for a positive semidefinite covariance matrix $V(\theta^*)$ is important for applications like, for example, dynamic linear panel data models. We next also use Assumption 2 from Kleibergen (2005), which concerns the convergence of the covariance matrix estimator.

Assumption 2. *The convergence behavior of the covariance matrix estimator $\hat{V}(\theta)$ toward $V(\theta)$ is such that (i) $\hat{V}(\theta) \underset{p}{\to} V(\theta)$, and (ii) $\frac{\partial \operatorname{vec}(\hat{V}_{ff}(\theta))}{\partial \theta'} \underset{p}{\to} \frac{\partial \operatorname{vec}(V_{ff}(\theta))}{\partial \theta'}$.*

Assumption 2(i) requests a consistent covariance estimator for $V(\theta)$. For Assumption 2(ii), the covariance matrix estimator $\hat{V}_{ff}(\theta)$ has to be such that the $\hat{V}_{qf}(\theta)$, which results from the derivative of $\hat{V}_{ff}(\theta)$ with respect to θ, is also a consistent estimator of $V_{qf}(\theta)$. When using the same covariance matrix estimator for all elements of $V(\theta)$, Assumption 2 holds under the conditions for consistency of (heteroskedastic autocorrelation consistent) covariance matrix estimators; see, for example, White (1980), Newey and West (1987). Lemma 11 in the Supplemental Appendix provides the low-level regularity conditions under which Assumption 2 holds in the leading i.i.d. case.

The score of the CUE objective function in (7) factorizes as the product of the sample moment and a recentered estimator of its Jacobian (8). Under Assumptions 1 and 2, the limit behaviors of the sample moment and this recentered Jacobian estimator are independent at the pseudo-true value θ^*:

$$\begin{aligned}\sqrt{T}\big(f_T(\theta^*, X) - \mu_f(\theta^*)\big) &\underset{d}{\to} \psi_f(\theta^*) \sim N\big(0, V_{ff}(\theta^*)\big),\\ \sqrt{T}\operatorname{vec}\big(\hat{D}(\theta^*) - D(\theta^*)\big) &\underset{d}{\to} \psi_\theta(\theta^*) \sim N\big(0, V_{\theta\theta}(\theta^*)\big),\end{aligned} \tag{21}$$

with $\psi_f(\theta^*)$ independent of $\psi_\theta(\theta^*)$, and $V_{\theta\theta}(\theta^*) = V_{qq}(\theta^*) - V_{qf}(\theta^*)V_{ff}(\theta^*)^{-1}V_{fq}(\theta^*)$. This asymptotic independence is an extension of Lemma 1 in Kleibergen (2005) and further stated as Lemma 5 in the Supplemental Appendix. It provides the base for the weak identification robust statistics and similarly for our misspecification robust score or LM statistic.

Because of the misspecification, the sample moment is however not centered at zero at the pseudo-true value. Consequently, we cannot use any of the weak identification robust statistics that request the sample moment to converge to zero, like the score, GMM-Anderson–Rubin and extensions of the conditional likelihood ratio statistic of Moreira (2003); see Stock and Wright (2000), Kleibergen (2005), Andrews (2016), and Andrews and Mikusheva (2016a, 2016b). The population score is equal to zero at the pseudo-true value though. When combined with the asymptotic independence of the sample moment and its recentered Jacobian, we then have that the (appropriately scaled) expected value of the limit of the score (7) equals zero at the pseudo-true value.[6] We can therefore use the quadratic form of the sample score to construct a double robust score or LM statistic.

2.4 *DRLM test for weak misspecification*

We focus on the case of weak misspecification, that is, the magnitude of $\mu_f(\theta^*)$ is small such that there exists $\tilde{\mu}_f(\theta^*) = \lim_{T\to\infty}\sqrt{T}\mu_f(\theta^*)$ as in (11). This treatment is analogous to the weak instrument asymptotics (see, e.g., Staiger and Stock (1997)) with $\tilde{D}(\theta^*) = \lim_{T\to\infty}\sqrt{T}D(\theta^*)$ as in (12), which leads to weak identification robust tests. Later in Section 5.2, we discuss potential strong misspecification.

Definition 1. The double robust Lagrange multiplier (DRLM) statistic for testing H_0 : $\theta^* = \theta_0^*$ with θ_0^* the hypothesized pseudo-true value is

$$\text{DRLM}(\theta_0^*) = T \times \hat{s}(\theta_0^*)\hat{W}(\theta_0^*)^{-1}\hat{s}(\theta_0^*)', \tag{22}$$

with

$$\begin{aligned}\hat{W}(\theta_0^*) = {} & \big(I_m \otimes \hat{V}_{ff}(\theta_0^*)^{-1} f_T(\theta_0^*, X)\big)'\hat{V}_{\theta\theta}(\theta_0^*)\big(I_m \otimes \hat{V}_{ff}(\theta_0^*)^{-1} f_T(\theta_0^*, X)\big) \\ & + \hat{D}(\theta_0^*)'\hat{V}_{ff}(\theta_0^*)^{-1}\hat{D}(\theta_0^*),\end{aligned} \tag{23}$$

where $\hat{V}_{\theta\theta}(\theta) = \hat{V}_{qq}(\theta) - \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)'$, $\hat{V}_{qf}(\theta) = (\hat{V}_{q_1 f}(\theta)' \ldots \hat{V}_{q_m f}(\theta)')'$ and $\hat{V}_{qq}(\theta)$ is a consistent estimator of $V_{qq}(\theta) = \lim_{T\to\infty} E[T\,\text{vec}(q_T(\theta) - J(\theta))\,\text{vec}(q_T(\theta) - J(\theta))']$.

The DRLM statistic in (22) is the quadratic form of the sample score in (7) and not just of the moment function. In this respect, the DRLM statistic differs from the GMM Anderson–Rubin statistic, which results from the quadratic form of the moment function. The DRLM statistic provides an extension of the KLM statistic from Kleibergen (2005), which tests hypotheses on θ in correctly specified GMM. The KLM and DRLM

[6] We formally show this in Lemma 6 of the Supplemental Appendix.

statistics are both quadratic forms of the sample score in (7), but differ with respect to the involved weight matrix. The weight matrix of DRLM in (23) explicitly accounts for misspecification, while the weight matrix of KLM does not since the KLM test assumes correct model specification. More specifically, the weight matrix of DRLM in (23) equals the sum of two components, where the last of these two, $\hat{D}(\theta_0^*)'\hat{V}_{ff}(\theta_0^*)^{-1}\hat{D}(\theta_0^*)$, constitutes the weight matrix of the KLM statistic. Thus, the DRLM statistic has an additional component, $(I_m \otimes \hat{V}_{ff}(\theta_0^*)^{-1} f_T(\theta_0^*, X))'\hat{V}_{\theta\theta}(\theta_0^*)(I_m \otimes \hat{V}_{ff}(\theta_0^*)^{-1} f_T(\theta_0^*, X))$, in the weight matrix to account for misspecification. In case of correct model specification, the sample moment $f_T(\theta_0^*, X)$ asymptotically reduces to zero; in contrast, under misspecification, $f_T(\theta_0^*, X)$ is not centered at zero, so the $(I_m \otimes \hat{V}_{ff}(\theta_0^*)^{-1} f_T(\theta_0^*, X))'\hat{V}_{\theta\theta}(\theta_0^*)(I_m \otimes \hat{V}_{ff}(\theta_0^*)^{-1} f_T(\theta_0^*, X))$ component is to explicitly take misspecification into account.

Hall and Inoue (2003) provide an expression for deriving the covariance of the GMM estimator, which also contains a separate part that is associated with misspecification. Yet Hall and Inoue (2003)'s focus is on the GMM estimator, whose asymptotic normal distribution relies on strong identification. In contrast, our interest lies in the sample score, and the validity of the resulting DRLM test does not require strong identification. On the one hand, the DRLM statistic is built on the score, whose population value under the null equals zero regardless of the strength of identification. On the other hand, the variance of the score, as in Definition 1, explicitly accounts for misspecification. Thus, by construction, the DRLM statistic is robust to weak identification and misspecification, as shown in Theorem 1.

THEOREM 1. *When Assumptions 1 and 2 hold and for finite valued k_f and $k_f \times m$ dimensional continuously differentiable functions of θ^*,*

$$\begin{aligned} \tilde{\mu}_f(\theta^*) &= \lim_{T\to\infty} \sqrt{T}\mu_f(\theta^*), \\ \tilde{D}(\theta^*) &= \lim_{T\to\infty} \sqrt{T}D(\theta^*), \end{aligned} \tag{24}$$

the limit behavior of $\mathrm{DRLM}(\theta^*)$ *under $H_0 : \theta^* = \theta_0^*$, for θ^* the minimizer of the population continuous updating objective function, is bounded according to*

$$\lim_{T\to\infty} \Pr\left[\mathrm{DRLM}(\theta_0^*) > cv_{\chi^2(m)}(\alpha)\right] \leq \alpha, \tag{25}$$

with $cv_{\chi^2(m)}(\alpha)$ the $(1-\alpha)\times 100\%$ critical value for the $\chi^2(m)$ distribution.

PROOF. See the Supplemental Appendix. □

3. SIZE AND POWER OF DRLM IN LINEAR MODELS

To illustrate the size and power of the DRLM test as well as its empirical relevance, in addition to the previously discussed linear IV regression model, we also consider the linear asset pricing model in this section. A nonlinear example is discussed in Section 5.3. For more examples of misspecified models that are potentially weakly identified see, for example, Hall and Inoue (2003) and Hansen and Lee (2021).

3.1 *Size of DRLM in the linear asset pricing or IV model*

The linear asset pricing model describes the return process by

$$R_t = \mu_R + \beta(F_t - \mu_F) + u_t, \quad t = 1, \ldots, T, \tag{26}$$

where a k_f-dimensional vector of asset returns R_t is related to m risk factors present in the m-dimensional vector F_t, with $\mu_R = E(R_t)$, $\mu_F = E(F_t)$, $\beta = \operatorname{cov}(R_t, F_t)\operatorname{var}(F_t)^{-1}$, and u_t is a k_f-dimensional vector of mean zero errors. The linear asset pricing model has the mean of R_t spanned by the betas of m risk factors in F_t. It can be reflected by the population moment function:

$$\begin{aligned}\mu_f(\lambda_F) &= \mu_R - \beta\lambda_F \\ &= \mu_f(0) + J(0)\theta,\end{aligned} \tag{27}$$

where λ_F is the risk premium. The last line of (27) again uses the GMM notation from (1) and (10), so $\mu_f(0) = \mu_R$, $J(0) = -\beta$, and $\theta = \lambda_F$.

The linear asset pricing model is correctly specified when there is a value of the risk premium λ_F, say $\lambda_{F,0}$, for which $\mu_R = \beta\lambda_{F,0}$ so the population moment function equals zero. Under misspecification, there is no such value and our pseudo-true value λ_F^* of the risk premium results from minimizing the population continuous updating objective function associated with (27).[7] Weak identification occurs when the full rank condition of β is at risk.

As indicated by (14) and (27), the linear IV regression model and the linear asset pricing model have similar linear moment functions. Under homoskedasticity, the covariance matrix for the sample moment and its derivative in Assumption 1 has a Kronecker product structure (KPS):

$$V(\theta^*) = \left(\begin{array}{c:c} V_{ff}(\theta^*) & V_{fq}(\theta^*) \\ \hdashline V_{qf}(\theta^*) & V_{qq}(\theta^*) \end{array}\right) = \left(\begin{pmatrix} 1 & 0 \\ \theta^* & I_m \end{pmatrix}' \Sigma \begin{pmatrix} 1 & 0 \\ \theta^* & I_m \end{pmatrix} \otimes \Omega\right),$$

with Ω and Σ $k_f \times k_f$ and $(m+1)\times(m+1)$ dimensional matrices, respectively. For a KPS covariance matrix, the expression of the DRLM statistic for testing $\mathrm{H}_0 : \theta^* = \theta_0^*$ simplifies to

$$\mathrm{DRLM}(\theta_0^*) = \hat{\mu}(\theta_0^*)^{*\prime}\hat{D}(\theta_0^*)^*\left[\hat{\mu}(\theta_0^*)^{*\prime}\hat{\mu}(\theta_0^*)^* I_m + \hat{D}(\theta_0^*)^{*\prime}\hat{D}(\theta_0^*)^*\right]^{-1}\hat{D}(\theta_0^*)^{*\prime}\hat{\mu}(\theta_0^*)^*, \tag{28}$$

where

$$\hat{\mu}(\theta_0^*)^* = \sqrt{T}\hat{V}_{ff}(\theta_0^*)^{-\frac{1}{2}} f_T(\theta_0^*, X) = \sqrt{T}\hat{\Omega}^{-\frac{1}{2}}\hat{\mu}(\theta_0^*)\left(\begin{pmatrix} 1 \\ \theta_0^* \end{pmatrix}' \hat{\Sigma} \begin{pmatrix} 1 \\ \theta_0^* \end{pmatrix}\right)^{-\frac{1}{2}}$$

[7] A more general specification of the linear asset pricing model is often used, which incorporates a so-called, zero-beta return, λ_0. Its population moment function is $\mu_f(\lambda_0, \lambda_F) = E(\mathcal{R}_t) - \iota_{k_f+1}\lambda_0 - \mathcal{B}\lambda_F$, with $\mathcal{R}_t$ a (k_f+1)-dimensional vector of asset returns, ι_{k_f+1} a (k_f+1)-dimensional vector of ones, $\mathcal{B} = \operatorname{cov}(\mathcal{R}_t, F_t)\operatorname{var}(F_t)^{-1}$. The population moment function in (27) then results by removing the zero-beta return, which can, for example, be accomplished by subtracting a base asset return, say, the (k_f+1)-th, to which our results are invariant: $R_t = (\mathcal{R}_{1t} \ldots \mathcal{R}_{k_f t})' - \iota_{k_f}\mathcal{R}_{(k_f+1)t}$, $\beta = (\mathcal{B}_1' \ldots \mathcal{B}_{k_f}')' - \iota_{k_f}\mathcal{B}_{k_f+1}$, for $\mathcal{R}_t = (\mathcal{R}_{1t} \ldots \mathcal{R}_{(k_f+1)t})'$, $\mathcal{B} = (\mathcal{B}_1' \ldots \mathcal{B}_{k_f+1}')'$.

is the normalized version of the sample moment, and similarly, $\hat{D}(\theta_0^*)^*$ is the normalized sample recentered Jacobian with

$$\hat{D}(\theta_0^*)^* = \sqrt{T}\hat{\Omega}^{-\frac{1}{2}}\hat{D}(\theta_0^*) \times \left(\begin{pmatrix}0\\ I_m\end{pmatrix}'\hat{\Sigma}\begin{pmatrix}0\\ I_m\end{pmatrix} - \begin{pmatrix}0\\ I_m\end{pmatrix}'\hat{\Sigma}\begin{pmatrix}1\\ \theta_0^*\end{pmatrix}\left(\begin{pmatrix}1\\ \theta_0^*\end{pmatrix}'\hat{\Sigma}\begin{pmatrix}1\\ \theta_0^*\end{pmatrix}\right)^{-1}\begin{pmatrix}1\\ \theta_0^*\end{pmatrix}'\hat{\Sigma}\begin{pmatrix}0\\ I_m\end{pmatrix}\right)^{-\frac{1}{2}},$$

and $\hat{\Omega}$, $\hat{\Sigma}$ are consistent estimators of Ω and Σ, respectively.

The simplified expression of the DRLM statistic in (28) facilitates the joint discussion of the size of the DRLM test for the linear IV regression and linear asset pricing models in Corollary 1 below.

Corollary 1. *When Assumptions 1 and 2 hold, and under i.i.d. homoskedastic errors, the limit behavior of the DRLM statistic under $H_0 : \beta^* = \beta_0^*$ in the linear IV regression model, or under $H_0 : \lambda_F^* = \lambda_{F,0}^*$ in the linear asset pricing model is dependent on $\bar{\mu}$ and $\bar{D}$:*

$$\begin{aligned}\text{DRLM} \underset{d}{\to} &\left[\psi_f'(\bar{D}+\Psi_\theta)+\bar{\mu}'\Psi_\theta\right]\left[(\bar{\mu}+\psi_f)'(\bar{\mu}+\psi_f)I_m+(\bar{D}+\Psi_\theta)'(\bar{D}+\Psi_\theta)\right]^{-1}\\ &\times\left[(\bar{D}+\Psi_\theta)'\psi_f+\Psi_\theta'\bar{\mu}\right]\\ &\preceq \chi^2(m),\end{aligned} \tag{29}$$

with ψ_f and Ψ_θ independent standard normal k_f and $k_f \times m$ dimensional random vectors/matrices, $\bar{\mu}'\bar{D} \equiv 0$, and for linear IV: $\bar{\mu} = Q_{zz}^{-\frac{1}{2}}\tilde{\mu}_f(\beta^*)\sigma_{\varepsilon\varepsilon}^{-\frac{1}{2}}$, $\sigma_{\varepsilon\varepsilon} = \omega_{uu} - 2\omega_{uv}\beta^* + \beta^{*\prime}\Omega_{vv}\beta^*$, $\bar{D} = Q_{zz}^{-\frac{1}{2}}\tilde{D}(\beta^*)\Sigma_{vv.\varepsilon}^{-\frac{1}{2}}$, $\Sigma_{vv.\varepsilon} = \Omega_{vv} - (\omega_{vu} - \Omega_{vv}\beta^*)\sigma_{\varepsilon\varepsilon}^{-1}(\omega_{vu} - \Omega_{vv}\beta^*)'$; while for linear asset pricing: $\bar{\mu} = \Omega^{-\frac{1}{2}}\tilde{\mu}_f(\lambda_F^*)(1 + \lambda_F^{*\prime}Q_{\bar{F}\bar{F}}^{-1}\lambda_F^*)^{-\frac{1}{2}}$, $\bar{D} = \Omega^{-\frac{1}{2}}\tilde{D}(\lambda_F^*)(Q_{\bar{F}\bar{F}} + \lambda_F^*\lambda_F^{*\prime})^{\frac{1}{2}}$, with $Q_{\bar{F}\bar{F}} = \text{var}(F_t)$, $\Omega = \text{var}(u_t)$. "$\preceq$" indicates stochastic dominance, that is, for a continuous nonnegative scalar random variable $u \preceq \chi^2(m)$: $\Pr[u > cv_{\chi^2(m)}(\alpha)] \leq \alpha$ as in Theorem 1.

The limit behavior of the DRLM statistic in Corollary 1 for linear moment equations and a homoskedastic setting shows that it depends on two parameters: the length of $\bar{\mu}$ and the $k_f \times m$ matrix $\bar{D}$, which is orthogonal to $\bar{\mu}$. It is easily seen that the resulting limiting distribution is dominated by the $\chi^2(m)$ distribution, which we further illustrate by simulation in the next subsection.

3.2 *Simulated size of DRLM*

Figure 1 reports the rejection frequencies of 5% significance DRLM tests using $\chi^2(1)$ critical values and the limiting distribution from Corollary 1 as a function of the lengths of $\bar{\mu}$ and $\bar{D}$ for the one included endogenous variable/factor setting, so $m = 1$, and $k_f = 25$.[8]

[8]When $m = 1$, $\bar{D}$ reduces to a vector with the same dimension as $\bar{\mu}$, so their lengths result from the inner products of the elements in each vector. We use $k_f = 25$, since it corresponds with the twenty-five Fama–

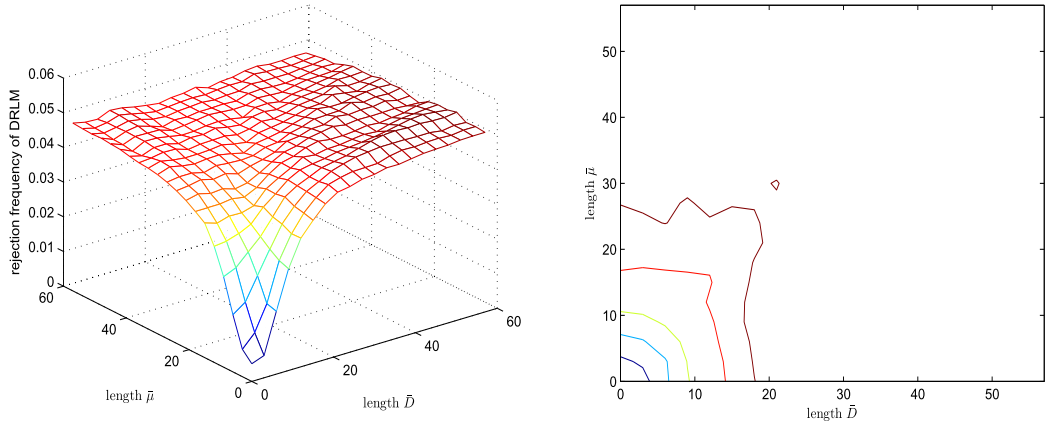


FIGURE 1. Rejection frequency of 5% significance DRLM tests of $H_0 : \theta^* = \theta_0^*$ using a 95% $\chi^2(1)$ critical value as a function of the lengths of $\bar{\mu}$ and $\bar{D}$, $m = 1$, $k_f = 25$.

Figure 1 shows that the DRLM test is size correct, since its rejection frequency does not exceed 5% for any length of $\bar{\mu}$ and $\bar{D}$. For comparison, Figure 2 presents the rejection frequencies of the KLM test (see Kleibergen (2005)), as a function of the lengths of $\bar{\mu}$ and $\bar{D}$. It shows that the KLM test is only size correct when there is no misspecification, so $\bar{\mu} = 0$, and can be severely size distorted for small values of the length of $\bar{\mu}$, especially when paired with small values of the length of $\bar{D}$.

Figure 1 also shows that the DRLM test is conservative when the lengths of both $\bar{\mu}$ and $\bar{D}$ are small. The conservativeness of the DRLM test at these low values can be reduced substantially when we calibrate a feasible conditional critical value function

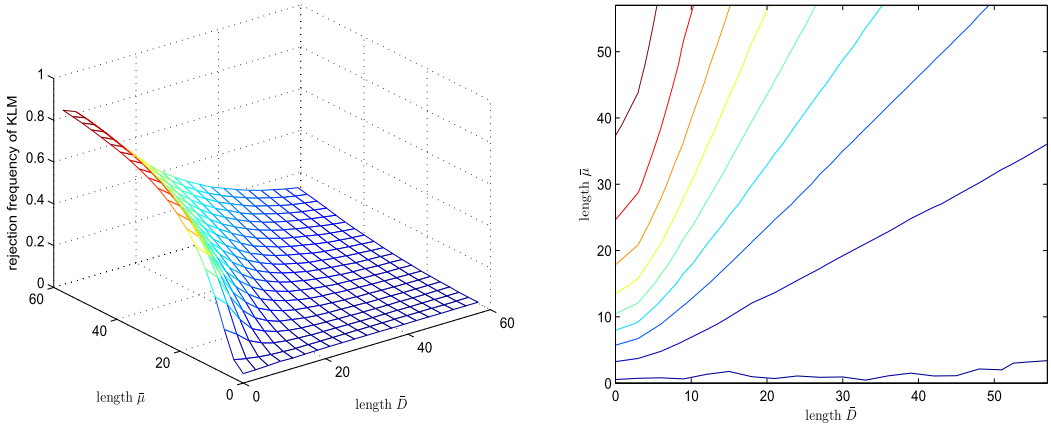


FIGURE 2. Rejection frequency of 5% significance KLM tests of $H_0 : \theta^* = \theta_0^*$ using a 95% $\chi^2(1)$ critical value as a function of the lengths of $\bar{\mu}$ and $\bar{D}$, $m = 1$, $k_f = 25$.

French size and book-to-market sorted portfolios, which are the default in the asset pricing literature; see Fama and French (1993).

for the DRLM test; see the discussion in Section II.E of the Supplemental Appendix and Guggenberger, Kleibergen, and Mavroeidis (2019).

3.3 *Power of DRLM in the linear model with $m = 1$*

For expository purposes, we analyze the power of the DRLM test for a single structural parameter, so $m = 1$, with linear moment equations and homoskedasticity. It leads to the KPS covariance matrix also used in (28). To simplify further, we use that Σ equals the identity matrix, so $\Sigma = I_2$, which is without loss of generality.

We analyze the power of the DRLM test for testing $H_0 : \theta^* = \theta_0^* = 0$. Therefore, for the simulation presented later, we use a data generating process for which the pseudo-true value equals θ^* while we test for a zero value of it. The rejection frequencies of the DRLM test using a range of values for θ^* then map out its power curve for testing H_0. Under $H_0 : \theta^* = 0$, the recentered Jacobian $D(0)$ coincides with its uncentered counterpart $J(0)$.

THEOREM 2. *For testing $H_0 : \theta^* = 0$, with $m = 1$, linear moment equations and homoskedastic errors where $\Sigma = I_2$, the limit behaviors of $\hat{\mu}(0)$ and $\hat{D}(0)$ when the involved pseudo-true value equals θ^* are characterized by*

$$\begin{aligned}\sqrt{T}\hat{\Omega}^{-\frac{1}{2}}\hat{\mu}(0) &\underset{d}{\to} \bar{\mu}\big(1+(\theta^*)^2\big)^{-\frac{1}{2}} - \bar{D}\big(1+(\theta^*)^2\big)^{-\frac{1}{2}}\theta^* + \psi_f^*(0),\\ \sqrt{T}\hat{\Omega}^{-\frac{1}{2}}\hat{D}(0) &\underset{d}{\to} \bar{D}\big(1+(\theta^*)^2\big)^{-\frac{1}{2}} + \bar{\mu}\big(1+(\theta^*)^2\big)^{-\frac{1}{2}}\theta^* + \psi_{\theta.f}^*(0),\end{aligned} \tag{30}$$

with $\psi_f^*(0)$ and $\psi_{\theta.f}^*(0)$ independent standard normal k_f-dimensional random vectors, $\bar{\mu} = \Omega^{-\frac{1}{2}}\tilde{\mu}_f(\theta^*)\frac{1}{\sqrt{1+(\theta^*)^2}} = \lim_{T\to\infty}\sqrt{T}V_{ff}(\theta^*)^{-\frac{1}{2}}\mu_f(\theta^*)$, $\tilde{\mu}_f(\theta^*) = \lim_{T\to\infty}\sqrt{T}\mu_f(\theta^*)$, $\mu_f(\theta^*) = \mu(0) + J(0)\theta^*$, $\bar{D} = \Omega^{-\frac{1}{2}}\tilde{D}(\theta^*)\sqrt{1+(\theta^*)^2} = \lim_{T\to\infty}\sqrt{T}V_{\theta\theta}(\theta^*)^{-\frac{1}{2}}D(\theta^*)$, $\tilde{D}(\theta^*) = \lim_{T\to\infty}\sqrt{T}D(\theta^*)$, $D(\theta^*) = J(0) - \mu_f(\theta^*)\frac{\theta^*}{1+(\theta^*)^2} = (\mu_f(0)\vdots J(0))\binom{-\theta^*}{1}\frac{1}{1+(\theta^*)^2}$, $\bar{\mu}'\bar{D} \equiv 0$, since $\tilde{\mu}_f(\theta^*)'\Omega^{-1}\tilde{D}(\theta^*) \equiv 0$.

PROOF. See the Supplemental Appendix. □

Theorem 2 implies the overall limit behavior of the sample score at the hypothesized value, as provided in the corollary below.

COROLLARY 2. *Under the conditions of Theorem 2, the limit behavior of the sample score at the hypothesized value, 0, is*

$$\begin{aligned}T\hat{s}(0) = T\hat{\mu}(0)'\hat{\Omega}^{-1}\hat{D}(0) &\underset{d}{\to} \big(1+(\theta^*)^2\big)^{-1}\theta^*\big(\bar{\mu}'\bar{\mu} - \bar{D}'\bar{D}\big)\\ &+ \big(1+(\theta^*)^2\big)^{-\frac{1}{2}}\big[\psi_f^*(0)'\big(\bar{D}+\bar{\mu}\theta^*\big) + \psi_{\theta.f}^*(0)'\big(\bar{\mu}-\bar{D}\theta^*\big)\big]\\ &+ \psi_f^*(0)'\psi_{\theta.f}^*(0).\end{aligned} \tag{31}$$

The mean of the limit behavior of the sample score in (31) equals zero when the normalized identification strength measure based on the Jacobian, $\bar{D}'\bar{D}$, and misspecification strength measure, $\bar{\mu}'\bar{\mu}$, are equal. It shows that the pseudo-true value θ^* is then not identified.

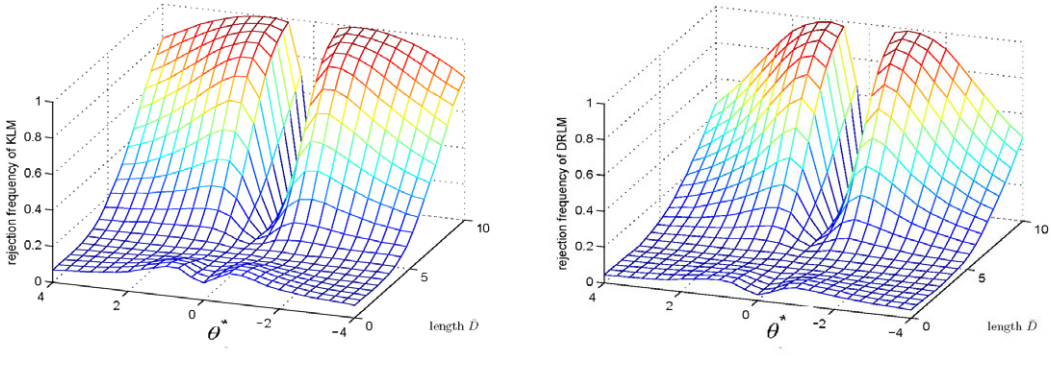
3.4 *Simulated power of DRLM*

Next, we illustrate the power of 5% significance DRLM tests of $H_0 : \theta^* = 0$ compared to other weak identification robust tests including GMM-AR, KLM, and LR tests; see, for example, Stock and Wright (2000), Moreira (2003), and Kleibergen (2005).

We consider a simulation setting with $k_f = 25$ moment equations, weak misspecification, $\bar{\mu}'\bar{\mu} = 10$, and varying identification strength. When there is no misspecification, Andrews, Moreira, and Stock (2006) show that the LR test is optimal, so we refrain from using that setting. Figures 3 and 4 show the power curves of KLM (panel 3.1), DRLM (panel 3.2), LR (panel 4.1), and GMM-AR (panel 4.2) tests of $H_0 : \theta^* = 0$ for various identification strengths and a fixed amount of misspecification.

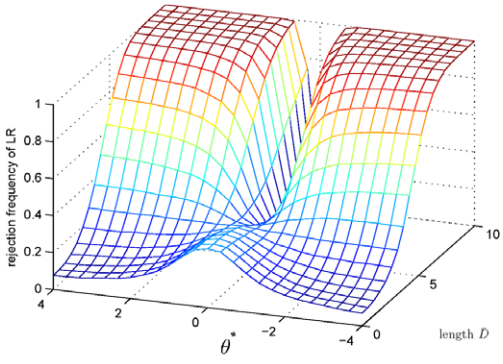
The power curves of the different tests in Figures 3 and 4 show that only the DRLM test is size correct for all settings of the identification strength. The size distortion of some of the other weak identification robust tests can be quite pronounced, which especially holds for the LR and GMM-AR tests. For the LR test, the rejection frequency at zero decreases from 30% to 8% when the identification strength increases, while for the KLM test, it decreases from 10% to 5%. The rejection frequency of the GMM-AR test at zero is equal to 36% for all identification strengths.

Corollary 2 shows that the pseudo-true value is not identified when the identification strength equals the amount of misspecification, $\bar{D}'\bar{D} = \bar{\mu}'\bar{\mu}$. This holds for a value of around three ($\approx \sqrt{10}$) on the "length $\bar{D}$" axis, which explains why the power curves for all tests are flat in the opposite direction at this value. For the DRLM test, the power curve is flat with a rejection frequency, which is proven to be at most 5%, while for the other tests it exceeds 5%, and for the LR and GMM-AR tests even by a substantial amount.

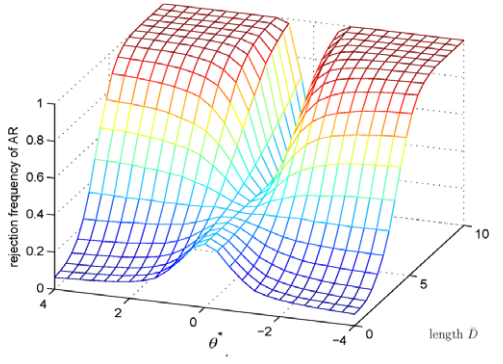


Panel 3.1: KLM Panel 3.2: DRLM

FIGURE 3. Power of 5% significance KLM and DRLM tests of $H_0 : \theta^* = 0$ with misspecification, $\bar{\mu}'\bar{\mu} = 10$, $k_f = 25$, $\Sigma = I_2$.



Panel 4.1: LR



Panel 4.2: GMM-AR

FIGURE 4. Power of 5% significance LR and GMM-AR tests of $H_0 : \theta^* = 0$ with misspecification, $\bar{\mu}'\bar{\mu} = 10$, $k_f = 25$, $\Sigma = I_2$.

When $\bar{D}'\bar{D}$ is less than $\bar{\mu}'\bar{\mu}$, the minimal value of the population continuous updating objective function is no longer at θ^* but at $\theta^1 = -\frac{1}{\theta^*}$.[9] The GMM-AR statistic is the sample analog of the population continuous updating objective function, which explains why its power is minimal at zero and maximal at infinity when "length $\bar{D}$" exceeds three, so $\bar{D}'\bar{D} > \bar{\mu}'\bar{\mu}$, and vice versa when "length $\bar{D}$" is less than three, so $\bar{D}'\bar{D} < \bar{\mu}'\bar{\mu}$. In the latter case, the pseudo-true value is not identified because it is at infinity. Kleibergen and Zhan (2024a) prove that the pseudo-true value of the CUE is not identified when the population value of the traditional rank statistic of the Jacobian, which they label *IS*, equals the minimal value of the CUE population objective function, which they label *MISS* and equals the population value of the J-statistic. They also show that $IS - MISS$ is always nonnegative and provides a measure of the identification strength in misspecified linear GMM whose sample analog is a (quasi-) likelihood ratio no-identification statistic. When $\bar{D}'\bar{D}$ is less than $\bar{\mu}'\bar{\mu}$, $MISS = IS$, which further indicates that the pseudo-true value is then not identified.

Figure 5 shows the distribution function of the misspecification J-statistic, which equals the minimal value of the GMM-AR statistic when the null hypothesis holds, so for values of θ^* equal to zero. It shows the distribution function for three different values of the identification strength $\bar{D}'\bar{D}$: 0, 10, and 100. Recognizing that the 95% critical value of the $\chi^2(24)$ distribution, since $k_f - 1 = 24$, is about 36.42, Figure 5 shows that we never reject no misspecification at the 5% significance level when $\bar{D}'\bar{D}$ equals 0, 7% of the times when $\bar{D}'\bar{D} = 10$, and 33% when $\bar{D}'\bar{D}$ equals 100. This indicates the difficulty of detecting misspecification; see also Gospodinov, Kan, and Robotti (2017).

The power curves of the LR and GMM-AR tests in Figure 4 increase when moving away from the hypothesized value and when the identification strength exceeds

[9]The first element of the limit behavior of the sample score in (31) equals zero at the two stationary points of the CUE population objective function: θ^* and $-\frac{1}{\theta^*}$. The sign of the Hessian also results from it and is indicated by $\bar{\mu}'\bar{\mu} - \bar{D}'\bar{D}$. Hence, when $\bar{\mu}'\bar{\mu} - \bar{D}'\bar{D} < 0$, the stationary points at θ^* and $-\frac{1}{\theta^*}$ are the minimum and maximum, respectively, and vice versa when $\bar{\mu}'\bar{\mu} - \bar{D}'\bar{D} > 0$; see Lemma 2(i) in the Supplemental Appendix.

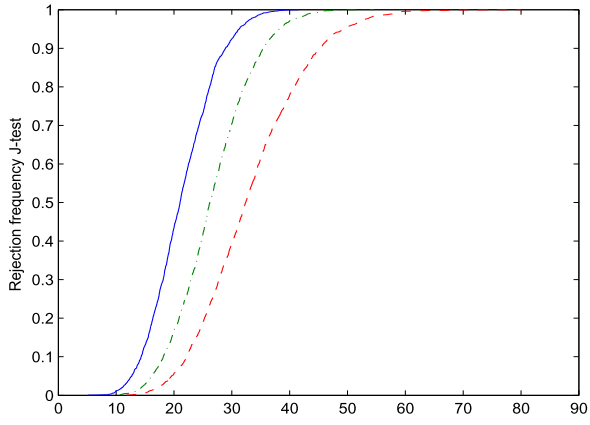


FIGURE 5. Distribution function of the J-statistic for misspecification when $\theta^* = 0$ holds, solid line: $\bar{D}'\bar{D} = 0$, dash-dot: $\bar{D}'\bar{D} = 10 =$ strength of misspecification, dashed: $\bar{D}'\bar{D} = 100$.

the measure of misspecification, $\bar{D}'\bar{D} > \bar{\mu}'\bar{\mu}$. This however does not hold for the power curves of the two LM tests in Figure 3, which are eventually decreasing. To improve the power of the DRLM test, we propose to use a property of its derivative as discussed next.

3.5 *Power improvement with $m = 1$*

The score is equal to zero at all stationary points of the CUE sample objective function, so the same holds for tests based on a quadratic form of it like, for example, the DRLM and KLM tests, as well. It explains the power decreasing away from the null hypothesis for the KLM and DRLM tests in Figure 3. Tests with better power properties therefore exist in correctly specified GMM that implicitly or explicitly combine the KLM test with an asymptotically independent J-test in either a conditional or unconditional manner; see Moreira (2003), Kleibergen (2005), Andrews, Moreira, and Stock (2006), Andrews (2016), and Andrews and Mikusheva (2016a, 2016b). In our misspecified GMM setting, this is however not possible since the limiting distribution of the J-statistic is a noncentral χ^2 distribution with an unknown noncentrality parameter. Hence, we cannot combine this limiting distribution with that of the DRLM statistic to obtain the (conditional) critical values for a combination test.

To improve the power of the DRLM test, we aim to reject hypothesized pseudo-true values of θ, which are close to a stationary point of the CUE sample objective function other than the CUE. This would be similar to the conditional or unconditional identification robust combination tests in regular GMM, which use that while the KLM test does not reject at such values of θ, J, and/or GMM-AR tests (see Anderson and Rubin (1949) and Stock and Wright (2000)) likely do. For hypothesized values of θ close to the CUE, these combination tests put most weight on the KLM test, but shift the weight toward the J and GMM-AR tests when θ is close to other stationary points; see Andrews (2016) and Kleibergen (2007). Since the limiting distributions of the J and GMM-AR statistics depend on unknown nuisance parameters in our misspecified GMM setting, it is not clear how we can use these statistics to improve power.

Stationary points of the sample CUE objective function lead to zero values of the DRLM statistic. The CUE is one of these stationary points and leads to the smallest value of the objective function. To improve the power of the DRLM test, we further reject values of θ that lead to nonsignificant values of the DRLM statistic, which result from stationary points different from the CUE. We can do so by rejecting values of θ in a closed set of nonrejected values that does not contain the CUE, for which an illustrative application is presented later in Section 4 (see Figure 7; a closed set of nonrejected values is on the left of Figure 7 as discussed for that figure).

To prove why the above power improvement rule leads to a size correct test procedure, we use the derivative of the DRLM statistic (see Lemma 12 in the Supplemental Appendix). In line with the previous Section 3.3, we also focus on $m = 1$ when discussing power improvement in Theorem 3 below.

THEOREM 3. *When Assumptions 1 and 2 hold, $m = 1$, $f_T(\theta, X)$ is linear in θ, and testing $H_0 : \theta^* = \theta_0^*$, with θ^* the pseudo-true value, at the $100 \times \alpha\%$ significance level, the power improved DRLM testing procedure that rejects a pseudo-true value of θ both if*:

1. *the DRLM statistic is significant at the $100 \times \alpha\%$ significance level or*
2. *the DRLM statistic is not significant but the hypothesized pseudo-true value of θ lies in a closed set of nonsignificant values, which does not contain the CUE with some significant values in between the closed set and the CUE,*

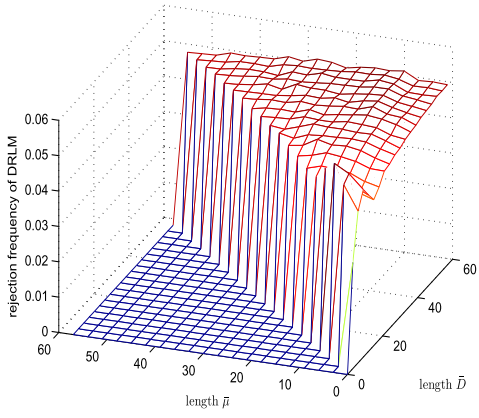
is a test procedure of $H_0 : \theta^* = \theta_0^*$ with size $100 \times \alpha\%$.

PROOF. See the Supplemental Appendix. □

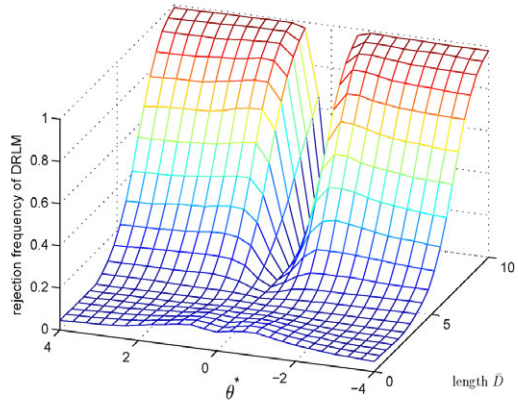
While the generic specification of the DRLM test is for a stationary point of the population continuous updating objective function, the power improved DRLM test from Theorem 3 explicitly tests for the minimizer. Later, Section 4 contains an empirical application (see Figure 7) that illustrates the power improvement rule from Theorem 3, in particular, its Step 2. When computing the size of the test at the hypothesized value of, say, zero, we therefore have to ascertain that it is the minimizer of the population objective function. For the setup in Figures 1–2, which uses the limit expression of the DRLM statistic in (29), the population minimizer is at zero if the amount of misspecification is less than the strength of identification so the length of $\bar{\mu}$ is less than that of $\bar{D}$. When the length of $\bar{\mu}$ exceeds that of $\bar{D}$, the minimizer of the population objective function is at $\pm\infty$. In standard GMM, there is no misspecification, so the minimal value of the population objective function is equal to zero. The amount of misspecification is then always less than or equal to the identification strength, so the hypothesized value automatically corresponds with the minimizer of the population objective function.

3.6 *Simulated power of power improved DRLM*

Figure 6 shows the rejection frequency and power of the power improved DRLM test. Panel 6.1 contains the rejection frequency when the minimizer of the population continuous updating objective function equals the hypothesized value, which is zero. It therefore does not show the rejection frequency for values where the length of $\bar{\mu}$ exceeds that



Panel 6.1: Rejection frequency



Panel 6.2: Power

FIGURE 6. Power improved 5% significance DRLM tests of $H_0 : \theta^* = 0$ using a conditional 95% critical value as a function of the lengths of $\bar{\mu}$ and $\bar{D}$ (panel 6.1), $\bar{\mu}'\bar{\mu} = 10$ (panel 6.2), $k_f = 25$, $\Sigma = I_2$.

of $\bar{D}$, since the hypothesized value does then not correspond with the minimizer of the population objective function, which is at $\pm\infty$. The rejection frequencies in panel 6.1 are computed using the calibrated conditional critical values. In line with Theorem 3, panel 6.1 shows that the power improvement does not affect the size of the DRLM test when the hypothesized value equals the minimizer of the continuous updating population objective function.

Panel 6.2 of Figure 6 shows power curves for the power improved DRLM test. It uses the same setup as for the power curves in Figures 3 and 4. At first glance, the power improved DRLM test in panel 6.2 seems minorly-size distorted because its rejection frequency can reach 8% when the length of $\bar{D}$ is below 3. At these values, the minimizer of the population CUE objective function is however at $\pm\infty$ since the misspecification exceeds the identification strength, so the 8% is indicative of power and not of size distortion. When the minimizer is at zero, so the identification strength exceeds the amount of misspecification, that is, the length of $\bar{D}$ exceeds 3, the rejection frequency of the DRLM test is at most 5% and reveals no size distortion. For the latter setting, panel 6.2 also shows that the power curves are (almost) monotonic and the decrease in power that we observed for the DRLM test in panel 3.2 is no longer present.

4. APPLICATIONS

To show the importance and relevance of the DRLM test for applied research where linear moment equations are common, we briefly revisit the linear models considered in Adrian, Etula, and Muir (2014) and He, Kelly, and Manela (2017) using our DRLM test and the identification robust GMM-AR, KLM, and LR tests; see also Kleibergen (2009) and Kleibergen and Zhan (2020). For the linear IV regression model, the Supplemental Appendix contains an empirical study using the data of Card (1995).

Adrian, Etula, and Muir (2014) propose a leverage risk factor ("*LevFac*") for asset pricing. The leverage level is the ratio of total assets over the difference between total assets and liabilities, and the leverage risk factor equals its log change. The empirical study of Adrian, Etula, and Muir (2014) uses quarterly data between 1968Q1 and 2009Q4. Following Lettau, Ludvigson, and Ma (2019), we extend the time period to 1963Q3–2013Q4 and use $N = 25$ size and book-to-market sorted portfolios as test assets. Adrian, Etula, and Muir (2014) show that the leverage factor prices the cross-section of many test portfolios, as reflected by the significant Fama–MacBeth (FM) (1973) and Kan–Robotti–Shanken (KRS) t-statistics on the risk premium reported in Table 1. The KRS t-statistic is robust to misspecification but not to weak identification; see Kan, Robotti, and Shanken (2013).

He, Kelly, and Manela (2017) propose the banking equity-capital ratio factor ("*EqFac*") for asset pricing. We consider one of their specifications with "*EqFac*" and the market return "R_m" as the two factors. The significant FM and KRS t-statistics for the risk premium on "*EqFac*" in Table 1 show that this factor is considered to be priced by the test assets.

DRLM: Adrian, Etula, and Muir (2014)

Using the same data as for Table 1, Figure 7 shows the p-values for testing the risk premium on the leverage factor (horizontal line) using the DRLM, GMM-AR, KLM, and LR tests. We also apply the LR no-identification test of Kleibergen and Zhan (2024a), whose test statistic 3.55 is considerably below its 95% conditional critical value of 83.3 (for conditioning statistic equal to 119), so we cannot reject the hypothesis that the pseudo-true value is not identified at the 5% (and much larger) significance level. Most of the p-values in Figure 7 are therefore above the 5% level, which implies that none of the DRLM, GMM-AR, KLM, and LR tests leads to tight 95% confidence intervals for the risk premium on the leverage factor as shown in Table 1. Given the smallish p-value of the J-test, 0.20, and the likely weak identification of the risk premium on the leverage factor reflected by the unbounded 95% confidence sets, misspecification could be present, so it would be appropriate to use the DRLM test.

The p-values of the DRLM test in Figure 7 are equal to one at two different points. The p-values of the GMM-AR test show that one of these two points relates to the minimal value of the GMM-AR test and the other one to the maximal value of the GMM-AR test. Using the power enhancement rule for DRLM stated in Theorem 3, we can reject nonsignificant values resulting from the DRLM test that lie within the closed interval indicated by the significant maximizer of the GMM-AR statistic (around the left peak in Figure 7, which does not contain the CUE), so the nonsignificant p-values of the DRLM test, which occur around the maximizer of the GMM-AR test can all be categorized as significant ones according to the power enhancement rule. The resulting 95% confidence set for the DRLM test rejects a zero value of the risk premium of the leverage factor and is reported in Table 1 alongside the one which results from just applying the DRLM test. The FM and KRS t-statistics reported in Table 1 also reject a zero value of the risk premium, but these tests are not reliable because of the potentially weak identification of the risk premium of the leverage factor and the likely misspecification reflected

TABLE 1. Inference on risk premia λ_F in Adrian, Etula, and Muir (2014) and He, Kelly, and Manela (2017). The test assets are the $N = 25$ size and book-to-market portfolios from 1963Q3 to 2013Q4 taken from Lettau, Ludvigson, and Ma (2019). "*LevFac*" is the leverage factor of Adrian, Etula, and Muir (2014). "*EqFac*" is the banking equity-capital ratio factor of He, Kelly, and Manela (2017). "R_m" is the market return. The estimate of λ_F and the FM t-statistic result from the Fama–MacBeth (1973) two-pass procedure. The KRS t-statistic is based on the KRS t-test of Kan, Robotti, and Shanken (2013). The point estimates of λ_F are identical to those reported in Lettau, Ludvigson, and Ma (2019). The F-statistic rank test is from Kleibergen and Zhan (2020) for testing $H_0 : \text{rank}(\beta) = m - 1$, with m the number of risk factors. The LR no-identification test and 95% conditional critical value are from Kleibergen and Zhan (2024a).

	Adrian, Etula, and Muir (2014)	He, Kelly, and Manela (2017)	
	LevFac	R_m	*EqFac*
Estimate of λ_F	13.91	1.19	6.88
FM t	3.58	0.81	2.14
KRS t	2.55	0.77	2.10
CUE of λ_F	51.77	23.22	94.02
95% confidence set			
FM t	(6.29, 21.54)	$(-1.67, 4.05)$	(0.57, 13.19)
KRS t	(3.22, 24.60)	$(-1.84, 4.22)$	(0.46, 13.30)
DRLM	$(-\infty, -91.4) \cup (-9.2, 1.7) \cup (17.8, +\infty)$	$(-\infty, +\infty)$	$(-\infty, +\infty)$
DRLM (power enh.)	$(-\infty, -91.4) \cup (17.8, +\infty)$	$(-\infty, +\infty)$	$(-\infty, +\infty)$
GMM-AR	$(-\infty, -101.3) \cup (18.3, +\infty)$	$(-\infty, -64.6) \cup (8.1, +\infty)$	$(-\infty, -244.1) \cup (37.7, +\infty)$
KLM	$(-\infty, -185.5) \cup (-5.7, -0.9) \cup (20.8, +\infty)$	$(-\infty, -7.2) \cup (-4.7, -0.3) \cup (1.0, +\infty)$	$(-\infty, -23.8) \cup (-8.1, 1.7) \cup (11.8, +\infty)$
LR	$(-\infty, -274.2) \cup (21.8, +\infty)$	$(-\infty, -9.7) \cup (2.2, +\infty)$	$(-\infty, -33.8) \cup (16.2, +\infty)$
IS χ^2-statistic (p-value)	31.97 (0.13)	35.88 (0.04)	
Rank F-statistic (p-value)	1.17 (0.28)	1.33 (0.16)	
LR no-identification (95%)	3.55 (83.3)	0.57 (56.3)	

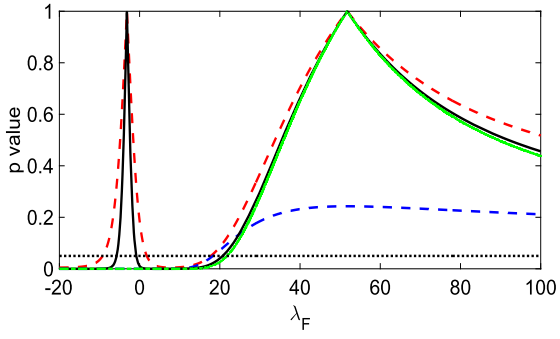


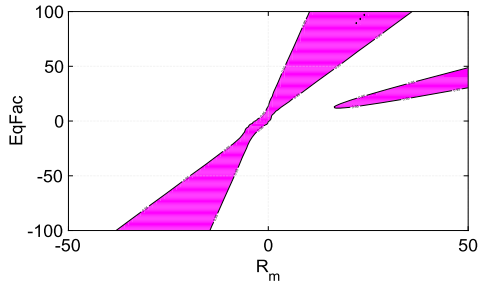
FIGURE 7. Adrian, Etula, and Muir (2014). p-value from the DRLM (dashed red), GMM-AR (dashed blue), KLM (solid black), LR (dash-dotted green), and the 5% level (dotted black). J-statistic (=minimum GMM-AR) equals 28.42, with p-value of 0.20 resulting from $\chi^2(N-2)$, $IS = 31.97$.

by the smallish p-value of the J-test. Since the zero risk premium is rejected by DRLM, the leverage factor does appear to have its pricing ability; on the other hand, since the confidence set of the risk premium is wide, we cannot precisely infer the leverage risk premium by using the adopted data.

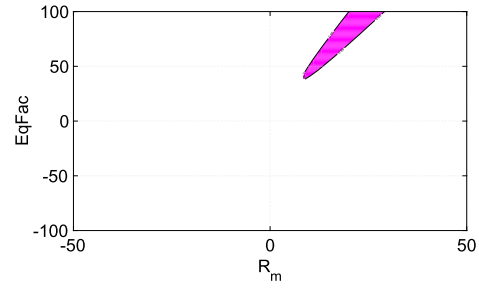
DRLM: He, Kelly, and Manela (2017)

Figure 8 shows the joint 95% confidence sets (shaded areas) of the risk premia on the banking equity-capital ratio factor "*EqFac*" and the market return "R_m," from using the DRLM, GMM-AR, KLM, and LR tests. The p-value of the J-test shows that misspecification is likely present, so it is appropriate to use the DRLM test for the confidence set of the minimizer of the population continuous updating objective function. The LR no-identification statistic of Kleibergen and Zhan (2024a), 0.57 is far below its 95% conditional critical value, 56.3 (for conditioning statistic equal to 114), so we cannot reject the hypothesis that the pseudo-true value is not identified. The 95% confidence sets of the DRLM and KLM tests consist of two or three rather disjoint sets. The power enhancement rule for the DRLM test shows that the smaller disjoint closed set can be discarded for the joint 95% confidence set that results from the DRLM test. The resulting 95% confidence set from the DRLM test includes a zero value for the risk premium on "*EqFac*" and is also unbounded, which indicates that the pricing ability of "*EqFac*" is under doubt.

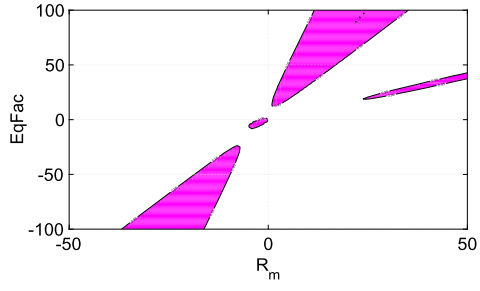
Gospodinov and Robotti (2021) also criticize the two-factor model of He, Kelly, and Manela (2017). They warn that "*EqFac*" and "R_m" are closely related, so jointly using them could lead to a reduced rank of the beta matrix. To compare with Figure 8, we replace the "*EqFac*" risk factor with the "SMB" (small minus big) factor from Fama and French (1993) and similarly construct Figure 9. The GMM-AR test now indicates misspecification, since it rejects every hypothesized risk premia as shown in panel 9.2, so the 95% confidence set that results from the GMM-AR test is empty. The LR no-identification statistic is now 69 with a 5% conditional critical value of 56.4 (for conditioning statistic equal to 214, see Kleibergen and Zhan (2024a)), so we reject the hypothesis that the pseudo-true value is not identified at the 5% significance level. Our



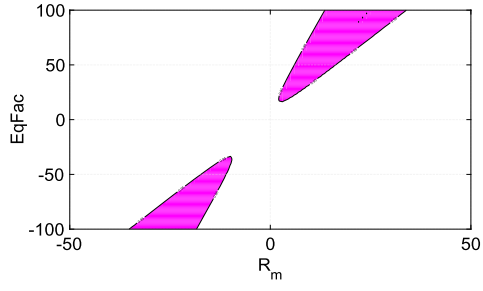
Panel 8.1: DRLM



Panel 8.2: GMM-AR



Panel 8.3: KLM



Panel 8.4: LR

FIGURE 8. He, Kelly, and Manela (2017). 95% confidence sets from DRLM, GMM-AR, KLM, and LR. J-statistic (minimum of GMM-AR) equals 35.32, with p-value of 0.036 resulting from $\chi^2(N-3)$, $IS = 35.88$.

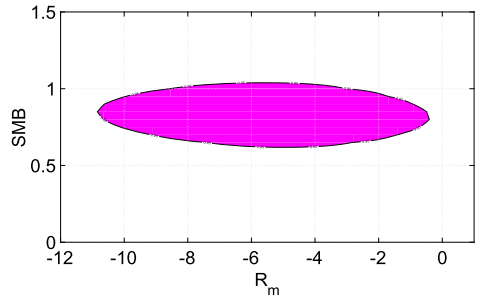
DRLM test, which allows for misspecification therefore yields a tight confidence set in panel 9.1. This tight confidence set, in contrast with the wide one in panel 8.1, indicates that the pricing ability of "*EqFac*" differs substantially from "SMB," that is, "SMB" leads to stronger identification of risk premia than "*EqFac*." Because of the misspecification, the 95% confidence sets resulting from the KLM and LR tests are not representative for the minimizer of the population objective function.

5. EXTENSIONS

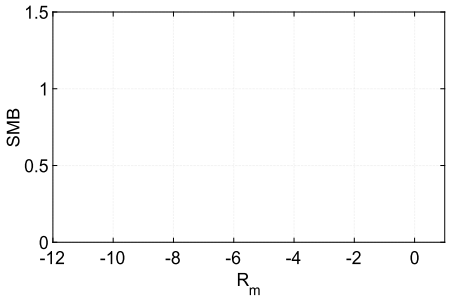
We next briefly discuss how the DRLM test can be used for subvector inference. Following up, we state the DRLM test that also allows for strong misspecification. Thereafter, we illustrate the size properties of the DRLM test for a setting with nonlinear moment equations and discuss an application to it.

5.1 *Subvector inference*

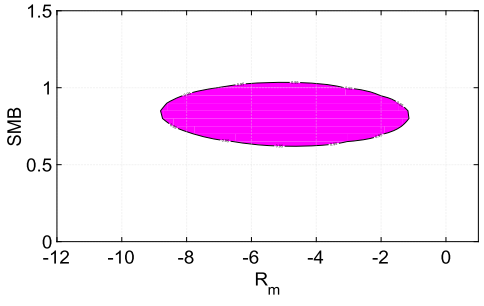
The expression of the DRLM statistic applies to settings where the structural parameter vector potentially has multiple elements, so the pseudo-true value $\theta^* = (\theta_1^* \dots \theta_m^*)'$ with $m \geq 1$. Many times, practitioners are interested in constructing confidence sets on a sub-



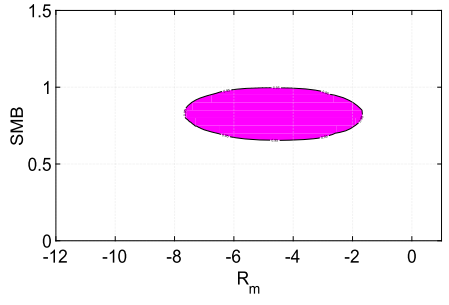
Panel 9.1: DRLM



Panel 9.2: GMM-AR



Panel 9.3: KLM



Panel 9.4: LR

FIGURE 9. R_m and SMB. 95% confidence sets from DRLM, GMM-AR, KLM, and LR. J-statistic (minimum of GMM-AR) equals 59.34, with p-value of 0.00 resulting from $\chi^2(N-3)$, $IS = 128.35$.

vector or individual element of the structural parameter vector. We discuss how to use the DRLM test for constructing such a confidence set in the Supplemental Appendix.

5.2 *DRLM test for strong misspecification*

For the strong misspecification case, that is, $\mu_f(\theta^*)$ is considered as a fixed nonzero vector instead of a drifting sequence to zero, the weight matrix involved in the DRLM statistic needs to be modified. For the general robustness of the DRLM test, we extend Assumption 1 to Assumption 1*, which concerns the joint limit behavior of the sample moment, its derivative, and the covariance matrix estimator.

ASSUMPTION 1*. *For a value of θ equal to the minimizer of the continuous updating population objective function, θ^*, we assume that the joint limit behavior of the sample moment, its derivative, and the covariance matrix estimator accords with the central limit theorem*:

$$\sqrt{T}\begin{pmatrix} f_T(\theta^*, X) - \mu_f(\theta^*) \\ \mathrm{vec}(q_T(\theta^*, X) - J(\theta^*)) \\ \mathrm{vech}(\hat{V}_{ff}(\theta^*) - V_{ff}(\theta^*)) \end{pmatrix} \underset{d}{\to} \begin{pmatrix} \psi_f(\theta^*) \\ \psi_q(\theta^*) \\ \psi_{ff}(\theta^*) \end{pmatrix} \sim N(0, \mathcal{V}(\theta^*)), \tag{32}$$

where the $\frac{1}{2}k_f(k_f+1)$ *dimensional vector* $\operatorname{vech}(V_{ff}(\theta^*))$ *contains the unique* $\frac{1}{2}k_f(k_f+1)$ *elements of the symmetric* $k_f \times k_f$ *dimensional matrix* $V_{ff}(\theta^*)$.

Assumption 1* is identical to the result implied by Theorem 2 of Hall and Inoue (2003) when using $V_{ff}(\theta^*)$ as the weight matrix. Assumption 1* extends Assumption 1 because it also defines the limit behavior of the covariance matrix estimator $\hat{V}_{ff}(\theta^*)$, which affects the limit behavior of the sample score in case of strong misspecification. It applies under the conditions listed in Hall and Inoue (2003), foremost Assumptions A.6–9, and Hansen and Lee (2021), Assumptions 2–4. Because the covariance between the limit behaviors of $f_T(\theta^*, X)$ and $\hat{D}(\theta^*)$ equals zero, and the estimator of this covariance also equals zero, its limit behavior does not affect the sample score in case of strong misspecification. Theorem 4 provides the appropriate weight matrix for the extended DRLM statistic that allows for both weak and strong misspecification and states that the same bounding result (25) holds for more general settings, which cover strong misspecification. To avoid confusion, we label the test statistic in Theorem 4 as $\widetilde{\mathrm{DRLM}}(\theta_0^*)$, in contrast with the previous $\mathrm{DRLM}(\theta_0^*)$ in Theorem 1.

THEOREM 4. *The extended DRLM statistic for testing* $H_0 : \theta^* = \theta_0^*$ *reads*

$$\widetilde{\mathrm{DRLM}}(\theta_0^*) = T \times \hat{s}(\theta_0^*)\widehat{\mathcal{W}}(\theta_0^*)^{-1}\hat{s}(\theta_0^*)', \tag{33}$$

where the weight matrix is given by $\widehat{\mathcal{W}}(\theta_0^*) = \hat{W}(\theta_0^*) + \hat{W}_s(\theta_0^*)$, *with* $\hat{W}(\theta_0^*)$ *resulting from* (23) *and* $\hat{W}_s(\theta_0^*) =$

$$\begin{pmatrix} \hat{V}_{ff}(\theta_0^*)^{-1}\hat{D}(\theta_0^*) \\ (I_m \otimes \hat{V}_{ff}(\theta_0^*)^{-1} f_T(\theta_0^*, X)) \\ -D'_{k_f}(\hat{V}_{ff}(\theta_0^*)^{-1} f_T(\theta_0^*, X) \otimes \hat{V}_{ff}(\theta_0^*)^{-1}\hat{D}(\theta_0^*)) \end{pmatrix}' \begin{pmatrix} 0 & 0 & \hat{V}_{f,ff}(\theta_0^*) \\ 0 & 0 & \hat{V}_{\theta,ff}(\theta_0^*) \\ \hat{V}_{ff,f}(\theta_0^*) & \hat{V}_{ff,\theta}(\theta_0^*) & \hat{V}_{ff,ff}(\theta_0^*) \end{pmatrix}$$
$$\begin{pmatrix} \hat{V}_{ff}(\theta_0^*)^{-1}\hat{D}(\theta_0^*) \\ (I_m \otimes \hat{V}_{ff}(\theta_0^*)^{-1} f_T(\theta_0^*, X)) \\ -D'_{k_f}(\hat{V}_{ff}(\theta_0^*)^{-1} f_T(\theta_0^*, X) \otimes \hat{V}_{ff}(\theta_0^*)^{-1}\hat{D}(\theta_0^*)) \end{pmatrix}, \tag{34}$$

where $\hat{V}_{f,ff}(\theta_0^*) = \hat{V}_{ff,f}(\theta_0^*)'$, $\hat{V}_{\theta,ff}(\theta_0^*) = \hat{V}_{ff,\theta}(\theta_0^*)'$, *and* $\hat{V}_{ff,ff}(\theta_0^*)$ *are the consistent estimators of the* $k_f \times \frac{1}{2}k_f(k_f+1)$, $mk_f \times \frac{1}{2}k_f(k_f+1)$, *and* $\frac{1}{2}k_f(k_f+1) \times \frac{1}{2}k_f(k_f+1)$ *dimensional covariances matrices between* $f_T(\theta_0^*, X)$ *and* $\operatorname{vech}(\hat{V}_{ff}(\theta_0^*))$, $\operatorname{vec}(\hat{D}(\theta_0^*))$ *and* $\operatorname{vech}(\hat{V}_{ff}(\theta_0^*))$, *and of* $\operatorname{vech}(\hat{V}_{ff}(\theta_0^*))$ *itself;* D_{k_f} *is the* $k_f^2 \times \frac{1}{2}k_f(k_f+1)$ *dimensional duplication matrix, so* $\operatorname{vec}(V_{ff}(\theta_0^*)) = D_{k_f}\operatorname{vech}(V_{ff}(\theta_0^*))$.

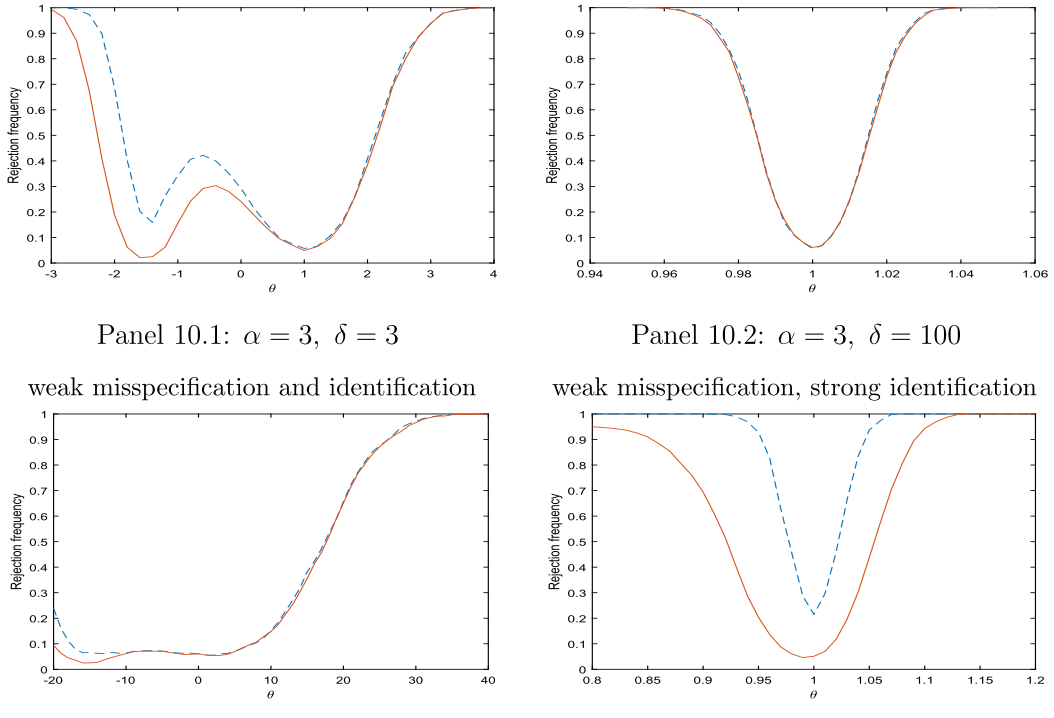
Under Assumptions 1* and 2, the extended DRLM test using the test statistic in (33) is size correct as in (25) for all settings of $\mu_f(\theta^*)$ and $D(\theta^*)$, that is, local to zero or fixed.

PROOF. See the Supplemental Appendix. □

Theorem 4 shows that the additional component $\hat{W}_s(\theta_0^*)$ in (34) has to be added to the weak misspecification/identification weight matrix $\hat{W}(\theta_0^*)$ to make the test also robust to strong misspecification. All components in $\hat{W}(\theta_0^*)$ consist of quadratic forms in $f_T(\theta_0^*, X)$ and $\hat{D}(\theta_0^*)$, while those in $\hat{W}_s(\theta_0^*)$ are at least of third order. It explains why $\hat{W}_s(\theta_0^*)$ vanishes if $f_T(\theta_0^*, X)$ or $\hat{D}(\theta_0^*)$ is of a lower order, compared to $\hat{W}(\theta_0^*)$. When $f_T(\theta_0^*, X)$ and $\hat{D}(\theta_0^*)$ converge in probability to a finite population defined constant, $\hat{W}(\theta_0^*)$ and $\hat{W}_s(\theta_0^*)$ are of the same order of magnitude, so ignoring $\hat{W}_s(\theta_0^*)$ would make the DRLM test size distorted.

To illustrate the effect of strong misspecification on the size of the DRLM test, we conduct a small simulation experiment where the values of α and δ drive the magnitude of misspecification and strength of identification, respectively, as in Figure 10.[10]

Figure 10 presents the rejection frequencies of the 5% significance DRLM test from Theorem 1, which uses a weight matrix suitable for weak misspecification, and the 5% significance extended DRLM test resulting from Theorem 4, whose weight matrix also allows for strong misspecification.



Panel 10.1: $\alpha = 3,\ \delta = 3$ weak misspecification and identification

Panel 10.2: $\alpha = 3,\ \delta = 100$ weak misspecification, strong identification

Panel 10.3: $\alpha = 50,\ \delta = 3$ strong misspecification, weak identification

Panel 10.4: $\alpha = 50,\ \delta = 100$ strong misspecification and identification

FIGURE 10. Rejection frequencies of 5% significance DRLM tests of $H_0 : \theta^* = 1$. DRLM test from Theorem 1 (dashed blue), DRLM test from Theorem 4 (solid red).

[10]The data generating process uses a single linear factor model with $k_f = 10$, $T = 1000$, $\beta_i = \frac{\delta}{\sqrt{T}}(\frac{i-1}{k_f-1})$, $\mu_{R,i} = \frac{\alpha}{\sqrt{T}}$, $\mu_F = 0$, $\Omega = I_{k_f}$, $Q_{FF} = 1$. β_i and $\mu_{R,i}$ are the ith element in β and μ_R, respectively.

Figure 10 shows that only in panel 10.4, which depicts strong misspecification and identification, there is a noticeable size distortion of the DRLM test from Theorem 1. Panels 10.1–10.3 show that the sizes of the two DRLM tests are basically identical while there is a small power loss when using the more robust DRLM test from Theorem 4.

5.3 *Nonlinear moment equations*

The DRLM test applies to general nonlinear GMM settings with unrestricted covariance matrices. In this subsection, we first illustrate the size of the DRLM test using nonlinear moment equations, for which we also conduct an empirical study afterwards.

5.3.1 *Size of DRLM in a nonlinear asset pricing model* We illustrate the size of the DRLM test by using the nonlinear moment equation resulting from a constant relative rate of risk aversion (CRRA) utility function (see, e.g., Hansen and Singleton (1982)):

$$E\left[\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(\iota_{k_f}+R_{t+1})-\iota_{k_f}\right]=\mu_f(\delta,\gamma), \tag{35}$$

with δ the discount factor, which is kept fixed at the value used in the simulation experiment, $\delta_0=0.95$, γ the relative rate of risk aversion, C_t consumption at time t, R_{t+1} a k_f-dimensional vector of asset returns, and ι_{k_f} a k_f-dimensional vector of ones. The setup of the simulation experiment, which is also used in Kleibergen and Zhan (2020) for illustrating the GMM-AR test, is laid out in the Supplemental Appendix. Unlike for the linear moment equations with i.i.d. homoskedastic setting, there is no analytical manner to compute the pseudo-true value, so it has to be done numerically.

Similar to Figures 1–2, Figure 11 shows the rejection frequencies of 5% significance GMM-AR and DRLM tests of the pseudo-true value of γ denoted by γ^*. The constant c in Figure 11 reflects the amount of misspecification, while $\tilde{c}$ reflects the identification

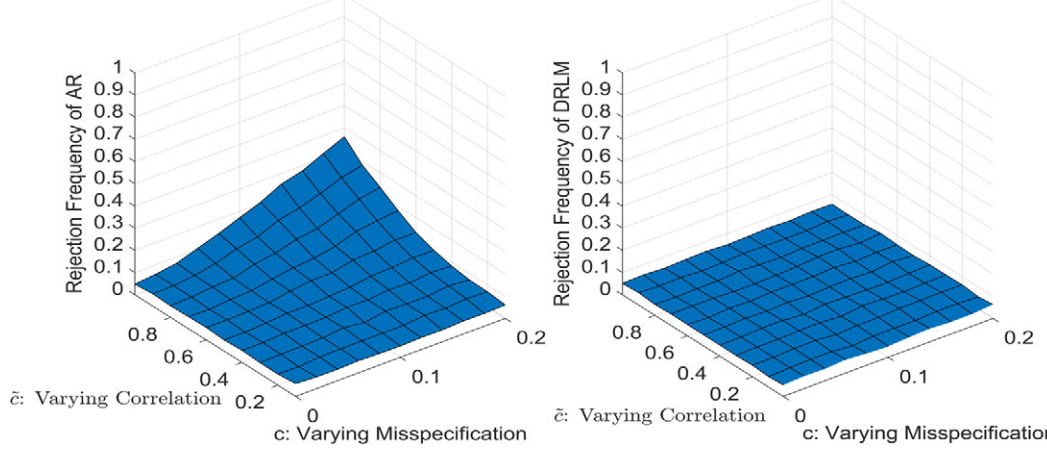


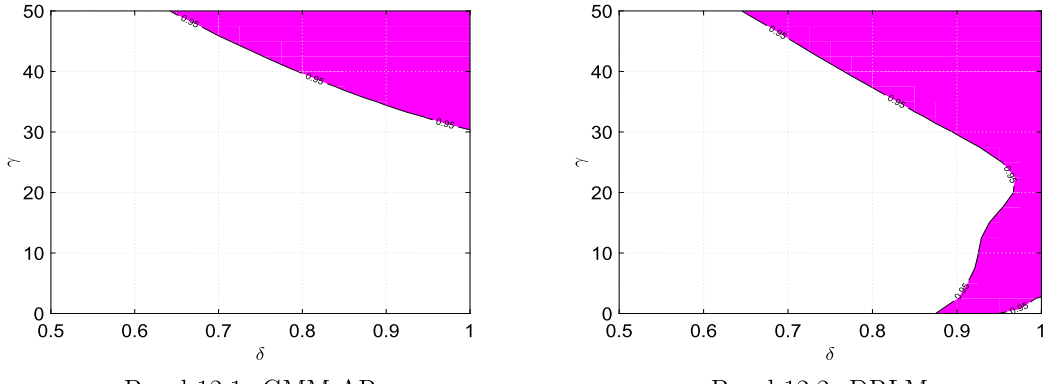
FIGURE 11. Rejection frequencies of 5% significance GMM-AR and DRLM tests of $H_0: \gamma^* = \gamma_0^*$ with $m=1$, $k_f=5$ as a function of misspecification c, and strength of identification $\tilde{c}$.

strength; see the Supplemental Appendix for further details. We note that the pseudo-true value γ^* is a function of $(c, \tilde{c})$, so the reported rejection frequencies in Figure 11 are for different hypothesized values of γ^*. The left panel of Figure 11 shows that the GMM-AR test gets size distorted when the misspecification increases. This is unlike the DRLM test in the right panel of Figure 11, which remains size correct for all values of the identification and misspecification strengths.

The Supplemental Appendix conducts additional simulation studies for illustrating the size of the DRLM test including a Monte Carlo study that employs the data generation process of Hansen and Lee (2021) for the linear IV regression model, and a simulation experiment that jointly tests the pseudo-true values of δ and γ in (35), so $m = 2$. Since the resulting findings are in line with Figures 1–2 and 11, we relegate them to the Supplemental Appendix.

5.3.2 *Application: Nonlinear asset pricing model with multiple parameters* We next use the DRLM test to construct the joint 95% confidence set for the discount rate and risk aversion for the moment function (35) resulting from a CRRA utility function. The data are yearly consumption and asset returns from 1928–2014. We use the six portfolios sorted by size and book-to-market from French's website for asset returns. The consumption series results from Kroencke (2017) and is also used in Kleibergen and Zhan (2020). For practicality, Figure 12 just reports on the region $\delta \in [0.5, 1]$ and $\gamma \in [0, 50]$ since values outside this region are unlikely to be economically meaningful. The shaded areas in panels 12.1 and 12.2 contain the values of (δ, γ) that are not rejected at the 5% level by the GMM-AR and DRLM tests, respectively.

Panel 12.1 shows that the GMM-AR test leads to large values for γ. For plausible values of δ, the 95% confidence set resulting from the GMM-AR test consists of values of γ



Panel 12.1: GMM-AR

Panel 12.2: DRLM

FIGURE 12. 95% confidence regions of (δ, γ) for CRRA utility. Notes: δ, the discount factor; γ, the relative rate of risk aversion. We focus on the region $\delta \in [0.5, 1]$, $\gamma \in [0, 50]$. The shaded areas contain the values of (δ, γ) that are not rejected at the 5% level by the GMM-AR and DRLM tests, respectively. The data are yearly consumption and asset returns from 1928–2014; the six portfolio returns sorted by size and book-to-market are from French's website, while the consumption series is from Kroencke (2017).

that exceed 30. These large values of γ are generally considered unrealistic, since they lead to the so-called, equity premium-risk-free rate puzzle; see, e.g., Savov (2011). In contrast with the joint confidence set of (δ, γ) resulting from the GMM-AR test in panel 12.1, panel 12.2 shows that the DRLM test leads to values of γ that are economically more reasonable: the 95% confidence set resulting from the DRLM test contains values of γ smaller than 30. More specifically, these plausible values of γ also occur jointly with reasonable values of δ mostly around $0.9 \sim 1$. Projecting the joint confidence set of (δ, γ) on the axes of δ and γ however leads to wide confidence sets for each of them, which therefore indicate the difficulty of precisely identifying these two parameters by using the adopted yearly data. Kleibergen and Zhan (2020) therefore call for longer time series and improved consumption measures for consumption-based asset pricing studies. We also expect to conduct further power improvements for DRLM using nonlinear moment conditions in future work, which can potentially lead to narrower confidence sets.

The difference between panel 12.1 and panel 12.2 results since the DRLM test allows for misspecification while the GMM-AR test does not, while both tests are robust to weak identification. Once we account for misspecification, the additional uncertainty leads to an expanded confidence set for (δ, γ) in panel 12.2, which contains economically meaningful values while panel 12.1 does not.

In the Supplemental Appendix, we extend our empirical analysis to a similar setting with Epstein–Zin (1989) preferences, which involve more time-consuming grid-search in the 3-dimensional space. The resulting findings are similar to those in Figure 12.

6. Conclusions

We show that it is generally feasible to conduct reliable inference on the pseudo-true value of the structural parameters resulting from the population continuous updating GMM objective function in case of misspecification and weak identification using the DRLM test. Hence, the DRLM test provides an important tool for conducting trustworthy inference for empirically relevant settings. We use the DRLM test to analyze empirical applications, which are plagued by both weak identification and misspecification issues: Adrian, Etula, and Muir (2014), He, Kelly, and Manela (2017), and a nonlinear asset pricing model with multiple parameters. They show that other inference procedures can seriously underestimate the uncertainty concerning the structural parameters when both misspecification and weak identification matter.

References

Adrian, Tobias, Erkko Etula, and Tyler Muir (2014), "Financial intermediaries and the cross-section of asset returns." *Journal of Finance*, 69, 2557–2596. [0297, 0313, 0314, 0315, 0316, 0323]

Anderson, Theodore and Herman Rubin (1949), "Estimation of the parameters of a single equation in a complete set of stochastic equations." *Annals of Mathematical Statistics*, 21, 570–582. [0297, 0311]

Andrews, Donald and Xu Cheng (2012), "Estimation and inference with weak, semi-strong and strong identification." *Econometrica*, 80, 2153–2211. [0295]

Andrews, Donald, Marcelo Moreira, and James Stock (2006), "Optimal two-sided invariant similar tests for instrumental variables regression." *Econometrica*, 74, 715–752. [0297, 0309, 0311]

Andrews, Isaiah (2016), "Conditional linear combination tests for weakly identified models." *Econometrica*, 84, 2155–2182. [0297, 0303, 0311]

Andrews, Isaiah (2019), "On the structure of IV estimands." *Journal of Econometrics*, 211 (1), 294–307. [0301]

Andrews, Isaiah and Anna Mikusheva (2016a), "A geometric approach to nonlinear econometric models." *Econometrica*, 84, 1249–1264. [0295, 0303, 0311]

Andrews, Isaiah and Anna Mikusheva (2016b), "Conditional inference with a functional nuisance parameter." *Econometrica*, 84, 1571–1612. [0295, 0303, 0311]

Andrews, Isaiah, James Stock, and Liyang Sun (2019), "Weak instruments in IV regression: Theory and practice." *Annual Review of Economics*, 11, 727–753. [0295]

Beaulieu, Marie-Claude, Jean-Marie Dufour, and Lynda Khalaf (2013), "Identification-robust estimation and testing of the zero-beta capm." *Review of Economic Studies*, 80, 892–924. [0295]

Card, David (1995), "Using geographic variation in college proximity to estimate the return to schooling." In *Aspects of Labour Market Behaviour: Essays in Honor of John Vanderkamp* (Louis Christofides, Kenneth Grant, and Robert Swidinsky, eds.), 201–222, University of Toronto Press, Toronto, Canada. NBER Working Paper 4483, 1993. [0313]

Cheng, Xu, Winston Dou, and Zhipeng Liao (2022), "Macro-finance decoupling: Robust evaluations of macro asset pricing models." *Econometrica*, 90 (2), 685–713. [0298]

Dufour, Jean-Marie (1997), "Some impossibility theorems in econometrics with applications to structural and dynamic models." *Econometrica*, 65, 1365–1388. [0295, 0296]

Epstein, Larry and Stanley Zin (1989), "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework." *Econometrica*, 57, 937–969. [0323]

Evdokimov, Kirill and Michal Kolesar "Inference in instrumental variables analysis with heterogeneous treatment effects." 2018. Working paper, Princeton University. [0297]

Fama, Eugene and Kenneth French (1993), "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics*, 33, 3–56. [0307, 0316]

Fama, Eugene and James MacBeth (1973), "Risk, return and equillibrium: Empirical tests." *Journal of Political Economy*, 81, 607–636. [0297, 0314, 0315]

Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti (2014), "Misspecification-robust inference in linear asset-pricing models with irrelevant factors." *Review of Financial Studies*, 27, 2139–2170. [0296]

Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti (2017), "Spurious inference in reduced-rank regresson models." *Econometrica*, 85, 1613–1628. [0296, 0310]

Gospodinov, Nikolay and Cesare Robotti (2021), "Common pricing across asset classes: Empirical evidence revisited." *Journal of Financial Economics*, 140 (1), 292–324. [0316]

Guggenberger, Patrik, Frank Kleibergen, and Sophocles Mavroeidis (2019), "A more powerful Anderson Rubin test in linear instrumental variables regression." *Quantitative Economics*, 10, 487–526. [0308]

Hall, Alastair and Atsushi Inoue (2003), "The large sample behaviour of the generalized method of moments esimator in misspecified models." *Journal of Econometrics*, 50, 1029–1054. [0296, 0304, 0319]

Han, Sukjin and Adam McCloskey (2019), "Estimation and inference with a (nearly) singular Jacobian." *Quantitative Economics*, 10, 1019–1068. [0295]

Hansen, Bruce and Seojeong Lee (2021), "Inference for iterated GMM under misspecification." *Econometrica*, 89 (3), 1419–1447. [0296, 0298, 0304, 0319, 0322]

Hansen, Lars Peter (1982), "Large sample properties of generalized method moments estimators." *Econometrica*, 50, 1029–1054. [0295, 0296, 0298]

Hansen, Lars Peter, John Heaton, and Amir Yaron (1996), "Finite sample properties of some alternative GMM estimators." *Journal of Business and Economic Statistics*, 14, 262–280. [0296, 0298]

Hansen, Lars Peter and Kenneth Singleton (1982), "Generalized instrumental variable estimation of nonlinear rational expectations models." *Econometrica*, 50, 1269–1286. [0321]

Hausman, Jerry (1983), "Specification and estimation of simultaneous equations systems." In *Handbook of Econometrics*, Vol. 1 (Zvi Griliches and Michael Intrilligator, eds.). Elsevier Science, Amsterdam. [0301]

He, Zhiguo, Brian Kelly, and Asaf Manela (2017), "Intermediary asset pricing: New evidence from many asset classes." *Journal of Financial Economics*, 126, 1–35. [0297, 0313, 0314, 0315, 0316, 0317, 0323]

Imbens, Guido and Joshua Angrist (1994), "Identification and estimation of local average treatment efects." *Econometrica*, 62, 467–475. [0297]

Kan, Raymond, Cesare Robotti, and Jay Shanken (2013), "Pricing model performance and the two-pass cross-sectional regression methodology." *Journal of Finance*, 68, 2617–2649. [0296, 0297, 0299, 0314, 0315]

Kan, Raymond and Chu Zhang (1999), "Two-pass tests of asset pricing models with useless factors." *Journal of Finance*, 54, 203–235. [0296]

Kleibergen, Frank (2002), "Pivotal statistics for testing structural parameters in instrumental variables regression." *Econometrica*, 70, 1781–1803. [0295, 0297]

Kleibergen, Frank (2005), "Testing parameters in GMM without assuming that they are identified." *Econometrica*, 73, 1103–1124. [0295, 0297, 0298, 0299, 0301, 0302, 0303, 0307, 0309, 0311]

Kleibergen, Frank (2007), "Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics." *Journal of Econometrics*, 139, 181–216. [0311]

Kleibergen, Frank (2009), "Tests of risk premia in linear factor models." *Journal of Econometrics*, 149, 149–173. [0295, 0297, 0313]

Kleibergen, Frank and Sophocles Mavroeidis (2009), "Weak instrument robust tests in GMM and the new Keynesian Phillips curve." *Journal of Business and Economic Statistics*, 27, 293–311. [0295]

Kleibergen, Frank and Zhaoguo Zhan (2020), "Robust inference for consumption-based asset pricing." *Journal of Finance*, 75, 507–550. [0295, 0297, 0313, 0315, 0321, 0322, 0323]

Kleibergen, Frank and Zhaoguo Zhan (2024a), "Testing for identification in potentially misspecified linear GMM." Unpublished Working paper, University of Amsterdam. [0310, 0314, 0315, 0316]

Kleibergen, Frank and Zhaoguo Zhan (2024b), "Online appendix to "Double robust inference for continuous updating GMM"." *Quantitative Economics Supplemental Material*, 15, https://doi.org/10.3982/QE2347. [0298]

Kroencke, Tim (2017), "Asset pricing without garbage." *Journal of Finance*, 72, 47–98. [0322]

Lee, Seojeong (2018), "A consistent variance estimator for 2SLS when instruments identify different LATEs." *Journal of Business and Economic Statistics*, 36, 400–410. [0296, 0297]

Lettau, Martin, Sydney Ludvigson, and Sai Ma (2019), "Capital share risk in U.S. asset pricing." *Journal of Finance*, 74, 1753–1792. [0314, 0315]

Maasoumi, Esfandiar (1990), "How to live with misspecification if you must." *Journal of Econometrics*, 44, 67–86. [0296]

Maasoumi, Esfandiar and Peter Phillips (1982), "On the behavior of inconsistent instrumental variable estimators." *Journal of Econometrics*, 19, 183–201. [0296]

Mavroeidis, Sophocles, Mikkel Plagborg-Moller, and James Stock (2014), "Empirical evidence on inflation expectations in the new-Keynesian Phillips curve." *Journal of Economic Literature*, 52, 124–188. [0295]

Moreira, Marcelo (2003), "A conditional likelihood ratio test for structural models." *Econometrica*, 71, 1027–1048. [0295, 0297, 0303, 0309, 0311]

Newey, Whitney and Kenneth West (1987), "A simple, positive semi-definite, heteroskedasticity and autocorrelation." *Econometrica*, 55 (3), 703–708. [0302]

Peñaranda, Francisco and Enrique Sentana (2015), "A unifying approach to the empirical evaluation of asset pricing models." *Review of Economics and Statistics*, 97 (2), 412–435. [0298]

Sargan, John Denis (1958), "The estimation of economic relationships using instrumental variables." *Econometrica*, 26, 393–415. [0296]

Savov, Alexi (2011), "Asset pricing with garbage." *Journal of Finance*, 72, 47–98. [0323]

Staiger, Douglas and James Stock (1997), "Instrumental variables regression with weak instruments." *Econometrica*, 65, 557–586. [0295, 0296, 0300, 0303]

Stock, James and Jonathan Wright (2000), "GMM with weak identification." *Econometrica*, 68, 1055–1096. [0295, 0298, 0303, 0309, 0311]

White, Halbert (1980), "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity." *Econometrica*, 48, 817–838. [0302]

White, Halbert (1994), *Estimation, Inference and Specification Analysis*. Econometric Society Monographs. Cambridge University Press. [0296]

Co-editor Stéphane Bonhomme handled this manuscript.

Manuscript received 27 February, 2023; final version accepted 5 November, 2024; available online 5 November, 2024.

The replication package for this paper is available at https://doi.org/10.5281/zenodo.14025673. The Journal checked the data and codes included in the package for their ability to reproduce the results in the paper and approved online appendices.