Double robust continuous updating GMM

Frank Kleibergen Zhaoguo Zhan*

November 16, 2020

Abstract

We propose the double robust Lagrange multiplier (DRLM) statistic for testing hypotheses specified on the pseudo-true value of the structural parameters in the generalized method of moments. The pseudo-true value is defined as the minimizer of the population continuous updating objective function of Hansen et al. (1996) and equals the true value of the structural parameter in the absence of misspecification. The (bounding) χ^2 limiting distribution of the DRLM test is robust to both misspecification and weak identification of the structural parameters, hence its name. Weak identification robust tests are size distorted in case of misspecification while misspecification tests are virtually powerless under weak identification, see Gospodinov et al. (2017), so the DRLM test removes an important obstacle for conducting reliable inference in these empirically relevant settings. To emphasize its importance for applied work, we use the DRLM test to analyze data from Card (1995), Adrian et al. (2014), and He et al. (2017).

Keywords: weak identification, misspecification, robust inference, Lagrange multiplier.

1 Introduction

Little more than twenty 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 e.g. Staiger and Stock (1997), Stock and Wright (2000), Kleibergen (2002, 2005, 2009), Moreira (2003), Andrews and Cheng (2012), and Andrews and Mikusheva (2016a,b). 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 amongst practitioners, see e.g. Kleibergen and Mavroeidis (2009), Mavroeidis et al. (2014), Andrews et al. (2019) and Kleibergen and Zhan (2020).

*Frank Kleibergen (corresponding author, f.r.kleibergen@uva.nl) is at the University of Amsterdam. Zhaoguo Zhan (zzhan@kennesaw.edu) is at Kennesaw State University.

The weak identification robust inference procedures lead to inference that is centered around the continuous updating estimator (CUE) of Hansen et al. (1996). GMM requests the moment condition to hold at a (unknown) true value of the parameter which is then also the minimizer of the population continuous updating objective function. The inference resulting from weak identification robust inference procedures concerning hypotheses specified on the true value of the structural parameters remains reliable under varying degrees of identification. When there is no value of the structural parameters where the GMM moment conditions exactly hold, the structural model is rendered misspecified and we refer to the minimizer of the (population continuous updating) GMM objective function as the pseudo-true value. The pseudo-true value depends on the (population) objective function at hand and different objective functions lead to distinct pseudo-true values. We use the minimizer of the population continuous updating objective function as the pseudo-true value because of its invariance properties and since weak identification robust tests lead to inference that is centered around it. In case of misspecification, these inference procedures for testing hypotheses specified on the pseudo-true value 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, however, 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 et al. (2017). Weak identification robust inference procedures thus came about to overcome the general critique of non-robustness of traditional inference procedures to varying identification strengths, see Staiger and Stock (1997), but are similarly non-robust to misspecification.

Arguably, the first to emphasize the importance 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, see Hall and Inoue (2003). In asset pricing models, for example, it is now generally accepted that misspecification, alongside weak identification, is an important empirical issue, see e.g. Kan et al. (2013) and Kleibergen and Zhan (2020). Kan et al. (2013) therefore developed misspecification robust t-statistics for the Fama-MacBeth (FM) (1973) two-pass estimator, *i.e.* the typical estimator employed to estimate risk premia in linear asset pricing models. These misspecification robust t-statistics are, however, not robust to weak identification so identical to the weak identification robust inference procedures, they cannot deal with the empirically relevant settings of both misspecification and weak identification. We therefore extend the weak identification robust score or Lagrange multiplier (KLM) test from Kleibergen

(2002, 2005, 2009) to a double robust Lagrange multiplier (DRLM) test. This DRLM test is size correct and robust to both misspecification and weak identification, hence its name. The DRLM test is a quadratic form of the score function which equals zero at all stationary points of the CUE sample objective function. This is also so for the KLM test and explains its power problems, see e.g. Andrews et al. (2006). To overcome the power problems of the KLM test, it can be combined in a conditional or unconditional manner with the Anderson-Rubin (AR) (1949) test, see e.g. Andrews (2016). Andrews et al. (2006) show that the conditional likelihood ratio 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. We use the maximal invariant to show that in case of misspecification, it is not obvious how to improve the power of the DRLM test by such combination arguments since the tests with which the DRLM test is to be combined have non-central limiting distributions under misspecification. We therefore improve the power of the DRLM test by exploiting the specification of its derivative with respect to the structural parameters.

The rest of the paper is organized as follows. In the second section, we discuss continuous updating GMM with misspecification, and show how a structural interpretation can be obtained from the pseudo-true value. In the third section, we introduce the DRLM test and prove that it is size correct. We illustrate the latter in a simulation experiment and also propose data-dependent critical values to reduce the conservativeness of the DRLM test for settings of both weak misspecification and weak identification. The fourth section conducts a power study of the DRLM test and other weak identification robust tests. 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 the power improvement rule and shows that the resulting test procedure has generally good power. The fifth section shows how to deal with multiple structural parameters. The sixth section conducts a simulation experiment using nonlinear GMM with an asset pricing Euler moment equation that results from a constant relative rate of risk aversion utility function. The seventh section applies the DRLM test to risk premia using asset pricing data from Adrian et al. (2014) and He et al. (2017), and to analyze the return on education using data from Card (1995) for which local average treatment effects that differ over the instruments can lead to misspecification, see Imbens and Angrist (1994). Especially for the risk premium parameters, we show that usage of other inference procedures understates the uncertainty of the risk measures because of the misspecification and weak identification present. The eighth section concludes.[1]

[1]Technical details and additional material are relegated to the Online Appendix.

2 GMM with potential misspecification

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 $f(.,.)$ 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 over identified and there are more moment equations than structural parameters so $k_f > m$. The population moment function of $f(\theta, X_t)$ equals $\mu_f(\theta)$:

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

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 $\mu_f(\theta_0) = 0$. We analyze θ using the continuous updating setting of Hansen et al. (1996). We use it because of its invariance properties and since it leads to inference using identification robust statistics in standard GMM, see e.g. Stock and Wright (2000) and Kleibergen (2005). 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 $f_T(\theta, X) = \frac{1}{T}\sum_{t=1}^T f_t(\theta)$, $f_t(\theta) = f(\theta, X_t)$:

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

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:

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

The minimizer of the population objective function satisfies the first order condition (FOC) stated in Theorem 1.

Theorem 1: The FOC (divided by two) for a stationary point θ^s of the population objective function reads:

$$\frac{1}{2}\frac{\partial}{\partial\theta'} Q_p(\theta^s) = 0 \quad \Leftrightarrow \quad \mu_f(\theta^s)' V_{ff}(\theta^s)^{-1} D(\theta^s) = 0, \tag{5}$$

with

$$D(\theta) = \quad J(\theta) - \left[V_{\theta_1 f}(\theta)V_{ff}(\theta)^{-1}\mu_f(\theta)\dots V_{\theta_m f}(\theta)V_{ff}(\theta)^{-1}\mu_f(\theta)\right] \tag{6}$$

and $J(\theta) = \frac{\partial}{\partial\theta'}\mu_f(\theta)$,

$$V_{\theta_i f}(\theta) = \quad E\left[\lim_{T\to\infty} T(\tfrac{\partial}{\partial\theta_i}(f_T(\theta,X) - \mu_f(\theta)))\,(f_T(\theta,X) - \mu_f(\theta))'\right], \qquad i = 1,\dots,m. \tag{7}$$

Proof. See the Online Appendix and Kleibergen (2005). ■

Theorem 1 shows that if there is a unique value of θ, θ_0, for which $\mu_f(\theta_0) = 0$, then also $\theta^* = \theta_0$ and $D(\theta_0) = J(\theta_0)$. The misspecification thus implies that the recentered Jacobian $D(\theta^*)$ differs from the population value $J(\theta^*)$ in other instances.

Running example 1: Linear asset pricing model The linear asset pricing model shows the extent to which the mean of an $(N+1)$-dimensional vector of asset returns $\mathcal{R}_t$ is spanned by the betas of m risk factors contained in the m-dimensional vector F_t. It is reflected by the moment function:

$$\mu_f(\lambda_0, \lambda_F) = E(\mathcal{R}_t) - \iota_{N+1}\lambda_0 - \mathcal{B}\lambda_F, \tag{8}$$

with ι_{N+1} an $(N+1)$-dimensional vector of ones, $\mathcal{B}$ an $(N+1)\times m$ dimensional matrix:

$$\mathcal{B} = \text{cov}(\mathcal{R}_t, F_t)\text{var}(F_t)^{-1}, \tag{9}$$

and λ_0 is the zero-beta return, λ_F is the m-dimensional vector of risk premia.

The asset pricing moment equation in (8) can be more compactly written by removing the zero-beta return which we accomplish by taking the asset returns in deviation of the $(N+1)$-th asset return:[2]

$$R_t = \begin{pmatrix} \mathcal{R}_{1t} \\ \vdots \\ \mathcal{R}_{Nt} \end{pmatrix} - \iota_N \mathcal{R}_{(N+1)t}, \qquad \beta = \begin{pmatrix} \mathcal{B}_1 \\ \vdots \\ \mathcal{B}_N \end{pmatrix} - \iota_N \mathcal{B}_{N+1}, \tag{10}$$

for $\mathcal{R}_t = (\mathcal{R}_{1t}\dots\mathcal{R}_{(N+1)t})'$, $\mathcal{B} = (\mathcal{B}_1'\dots\mathcal{B}_{N+1}')'$. The removal of the zero-beta return leads to the moment function:

$$\mu_f(\lambda_F) = \mu_R - \beta\lambda_F, \tag{11}$$

[2] This is without loss of generality since our results are invariant with respect to the asset return which is subtracted, see Kleibergen and Zhan (2020).

with $\mu_R = E(R_t)$ and $\beta = cov(R_t, F_t)var(F_t)^{-1}$.

The mean asset returns are not necessarily fully spanned by the β's and we therefore analyze the pseudo-true value of the risk premia λ_F^* which is the minimizer of the population continuous updating objective function:

$$Q_p(\lambda_F) = (\mu_R - \beta\lambda_F)' \left[\text{Var}\left(\sqrt{T}\left(\bar{R} - \hat{\beta}\lambda_F\right)\right)\right]^{-1} (\mu_R - \beta\lambda_F), \tag{12}$$

since $f_T(\lambda_F, X) = \bar{R} - \hat{\beta}\lambda_F$, with $\bar{R} = \frac{1}{T}\sum_{t=1}^T R_t$ and $\hat{\beta} = \frac{1}{T}\sum_{t=1}^T R_t\bar{F}_t' \left(\frac{1}{T}\sum_{j=1}^T \bar{F}_j\bar{F}_j'\right)^{-1}$, $\bar{F}_t = F_t - \bar{F}$, $\bar{F} = \frac{1}{T}\sum_{t=1}^T F_t$. The population continuous updating objective function results from a generalized reduced rank problem, see also Kleibergen (2007):

$$Q_p(\lambda_F) = \min_{D\in\mathbb{R}^{N\times m}} Q_p(\lambda_F, D) \tag{13}$$

with $D(\lambda_F) = \arg\min_{D\in\mathbb{R}^{N\times m}} Q_p(\lambda_F, D)$ and

$$\begin{aligned} Q_p(\lambda_F, D) = \quad & \left[\text{vec}\left(\left(\mu_R \vdots \beta\right) + D\left(\lambda_F \vdots I_m\right)\right)\right]' \left[\text{Var}\left(\sqrt{T}\left(\bar{R}' \vdots \text{vec}(\hat{\beta})'\right)'\right)\right]^{-1} \\ & \left[\text{vec}\left(\left(\mu_R \vdots \beta\right) + D\left(\lambda_F \vdots I_m\right)\right)\right]. \end{aligned} \tag{14}$$

The minimal value over (λ_F, D) of the objective function (14) is invariant with respect to the reduced rank specification implied by $D(\lambda_F \vdots I_m)$. When using another reduced rank specification, say, $E(I_m \vdots \phi)$, with E an $N \times m$ matrix and ϕ an m-dimensional vector, it leads to the same value of the optimized objective function over (ϕ, E). Hence, restrictions imposed on this specification, like, for example, $\phi_1 = 0$, with ϕ_1 the top element of ϕ, which imposes a reduced rank value on just β, lead to a larger (or equal) value of the minimized objective function. This setting is such that the objective function reflects the identification strength of λ_F as reflected by the distance of β from a reduced rank value. For the pseudo-true value λ_F^* to reflect risk premia and thus to have a structural interpretation, the identification strength has therefore to be larger than or equal to the misspecification. In standard GMM without misspecification, the minimal value of the continuous updating objective function equals zero so reduced rank values of β cannot lower the minimal value of the objective function and its minimizer always has a structural interpretation. We next further illustrate this for a simplified setting of the asset pricing model.

When $\mu_F = E(F_t) = 0$ and $\hat{\beta}$ results from a regression of $\bar{R}_t$ on $\bar{F}_t$ in which the error term is assumed to be i.i.d. with $N \times N$ dimensional covariance matrix Ω, Lemma 1 in the Online Appendix shows that $\bar{R}$ and $\hat{\beta}$ are independently normally distributed in large samples, see also Shanken (1992) and Kleibergen (2009). The population continuous updating objective function (12) then simplifies to:

$$Q_p(\lambda_F) = \tfrac{1}{1+\lambda_F' Q_{\bar{F}\bar{F}}^{-1}\lambda_F}(\mu_R - \beta\lambda_F)'\Omega^{-1}(\mu_R - \beta\lambda_F), \tag{15}$$

with $Q_{\bar{F}\bar{F}}$ =var(F_t), so its minimal value equals the smallest root of the characteristic polynomial:

$$\left| \tau \begin{pmatrix} 1 & 0 \\ 0 & Q_{\bar{F}\bar{F}}^{-1} \end{pmatrix} - \left(\mu_R \vdots \beta\right)' \Omega^{-1} \left(\mu_R \vdots \beta\right) \right| = 0. \tag{16}$$

Proposition 1. Using a value of λ_F, λ_F^s, that satisfies the FOC in Theorem 1, the smallest root of the characteristic polynomial in (16) equals either

$$\tfrac{1}{1+\lambda_F^{s\prime} Q_{\bar{F}\bar{F}}^{-1}\lambda_F^s} \left(\mu_R - \beta\lambda_F^s\right)' \Omega^{-1} \left(\mu_R - \beta\lambda_F^s\right) \tag{17}$$

or the smallest root of the characteristic polynomial:

$$\left|\tau(Q_{\bar{F}\bar{F}} + \lambda_F^s\lambda_F^{s\prime})^{-1} - D(\lambda_F^s)'\Omega^{-1}D\left(\lambda_F^s\right)\right| = 0, \tag{18}$$

with $D(\lambda_F) = -\beta - \left(\mu_R - \beta\lambda_F\right)\lambda_F' Q_{\bar{F}\bar{F}}^{-1}(1+\lambda_F' Q_{\bar{F}\bar{F}}^{-1}\lambda_F)^{-1} = -(\beta Q_{\bar{F}\bar{F}} + \mu_R\lambda_F')(Q_{\bar{F}\bar{F}} + \lambda_F\lambda_F')^{-1}$.

Proof. The rewriting of (16) to obtain the above is conducted in the Online Appendix. ■

Without misspecification, there is a value of λ_F^s for which (17) is equal to zero so it is the smallest root of the characteristic polynomial. Proposition 1 therefore shows that in models with misspecification, the minimizer of the population objective function is not necessarily associated with misspecification. For example, when $m = 1$, $\beta = 0$ and $\mu_R \neq 0$, the roots of the characteristic polynomial in (16) equal zero, attained for $\lambda_F = \pm\infty$, and $\mu_R'\Omega^{-1}\mu_R$, attained at $\lambda_F = 0$, so the smallest root is then associated with the identification strength reflected by β and the largest one with misspecification. The pseudo-true value does thus only represent the risk premia and has a structural interpretation when the identification strength is at least as large as the misspecification. This setting is used in Kan and Zhang (1999) to point at the misbehavior of traditional inference methods; see also Gospodinov et al. (2017).

Running example 2: Linear instrumental variables regression model For the linear instrumental variables (IV) regression model:

$$
\begin{aligned}
y = & \quad X\beta + \varepsilon \\
X = & \quad Z\Pi + V,
\end{aligned} \tag{19}
$$

with β and Π $m \times 1$ and $k \times m$ matrices containing unknown parameters, $y = (y_1 \ldots y_T)'$ and $X = (X_1 \ldots X_T)'$ $T \times 1$ and $T \times m$ dimensional matrices containing the endogenous variables, $Z = (Z_1 \ldots Z_T)'$ a $T \times k$ matrix containing the instrumental variables, $\varepsilon = (\varepsilon_1 \ldots \varepsilon_T)'$ and $V = (V_1 \ldots V_T)'$ are $T \times 1$ and $T \times m$ matrices of errors. The moment function is:

$$
\mu_f(\beta) = \sigma_{Zy} - \Sigma_{ZX}\beta, \tag{20}
$$

with $\sigma_{Zy} = E((Z_t - \mu_Z)(y_t - \mu_y))$, $\Sigma_{ZX} = E((Z_t - \mu_Z)(X_t - \mu_X)') = Q_{\bar{Z}\bar{Z}}\Pi$, $Q_{\bar{Z}\bar{Z}} = E((Z_t - \mu_Z)(Z_t - \mu_Z)')$, $\mu_y = E(y_t)$, $\mu_X = E(X_t)$, $\mu_Z = E(Z_t)$. When $u_t = \varepsilon_t + V_t'\beta$ and V_t are i.i.d. distributed with mean zero and covariance matrix $\Omega = \left(\begin{smallmatrix} \omega_{uu} & \omega_{uV} \\ \omega_{Vu} & \Omega_{VV} \end{smallmatrix}\right)$, the population continuous updating objective function of the linear IV regression model is:

$$
Q_p(\beta) = \tfrac{1}{\omega_{uu} - 2\omega_{uV}\beta + \beta'\Omega_{VV}\beta}(\sigma_{Zy} - \Sigma_{ZX}\beta)' Q_{\bar{Z}\bar{Z}}^{-1}(\sigma_{Zy} - \Sigma_{ZX}\beta). \tag{21}
$$

Along the same lines as for the linear asset pricing model, the minimal value of this population continuous updating objective function equals the smallest root of a characteristic polynomial:

$$
\left| \tau\Omega - \left(\sigma_{Zy} \vdots \Sigma_{ZX}\right)' Q_{\bar{Z}\bar{Z}}^{-1} \left(\sigma_{Zy} \vdots \Sigma_{ZX}\right) \right| = 0. \tag{22}
$$

If there is no value of β for which $\mu_f(\beta) = 0$, identical to the characteristic polynomial of the linear asset pricing model, the smallest root of the characteristic polynomial is only associated with misspecification when the misspecification is less than the identification strength.

Misspecified linear IV regression models are of interest in several settings, for example, when analyzing treatment effects. In case of multiple discrete instruments and heterogeneous treatment effects, the local average treatment effects of Imbens and Angrist (1994) differ over the instruments so the linear IV regression model using all these instruments is misspecified. The pseudo-true value is then a function of these local average treatment effects. We lateron provide an empirical illustration of this using data from Card (1995) in Section 7. Kolesár et al. (2015) provide another example of how a misspecified linear IV regression model can render a structural interpretation. Similarly, Kan et al. (2013) give a structural interpretation to the misspecified linear factor model as minimizing the pricing errors. In the Online Appendix, we provide further discussions on how a structural interpretation can be given to these models in case of misspecification. It is also important to realize that the identification of the structural parameters is often rather weak in applied settings in which case misspecification tests have very little power, see Gospodinov et al. (2017). The identification robust tests needed because of weak identification then become size distorted for testing the pseudo-true value in the presence of misspecification, so it is important to have tests which remain size correct for these empirically relevant settings.

3 Double robust score test

The sample analog of the population continuous updating objective function is the sample objective function for the continuous updating estimator (CUE) of Hansen et al. (1996):

$$\hat{Q}_s(\theta) = f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X), \tag{23}$$

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

$$\hat{\theta} = \arg\min_{\theta \in \mathbb{R}^m} \hat{Q}_s(\theta). \tag{24}$$

To construct the large sample behavior of test statistics centered around the CUE, we 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. *For a value of θ equal to the minimizer of the continuous updating population objective function, θ^*, the $k_f \times 1$ dimensional derivative of $f_t(\theta)$ with respect to θ_i,*

$$q_{it}(\theta) = \tfrac{\partial f_t(\theta)}{\partial \theta_i} : k_f \times 1, \qquad\qquad i = 1, \ldots, m, \tag{25}$$

is such that the joint limiting 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 theorem:

$$\tfrac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} \bar{f}_t(\theta) \\ \bar{q}_t(\theta) \end{pmatrix} \underset{d}{\rightarrow} \begin{pmatrix} \psi_f(\theta) \\ \psi_\theta(\theta) \end{pmatrix} \sim N(0, V(\theta)), \tag{26}$$

where $\psi_f : k_f \times 1$, $\psi_\theta : k_\theta \times 1$, $k_\theta = mk_f$, *and* $V(\theta)$ *is a positive semi-definite symmetric* $(k_f + k_\theta) \times (k_f + k_\theta)$ *matrix,*

$$V(\theta) = \begin{pmatrix} V_{ff}(\theta) & V_{f\theta}(\theta) \\ V_{\theta f}(\theta) & V_{\theta\theta}(\theta) \end{pmatrix}, \tag{27}$$

with $V_{\theta f}(\theta) = V_{f\theta}(\theta)' = (V_{\theta_1 f}(\theta)' \dots V_{\theta_m f}(\theta)')'$, $V_{\theta\theta}(\theta) = (V_{\theta_i\theta_j}(\theta)) : i, j = 1, \dots, m$; *and* $V_{ff}(\theta)$, $V_{\theta_i f}(\theta)$, $V_{\theta_i\theta_j}(\theta)$ *are* $k_f \times k_f$ *dimensional matrices for* $i, j = 1, \dots, m$, *and*

$$V(\theta) = \quad \mathrm{var}\left(\lim_{T\to\infty} \sqrt{T} \begin{pmatrix} f_T(\theta, X) \\ vec(q_T(\theta, X)) \end{pmatrix}\right), \tag{28}$$

with $q_T(\theta, X) = \frac{\partial f_T(\theta,X)}{\partial \theta'}|_\theta = \frac{1}{T}\sum_{t=1}^T (q_{1t}(\theta) \dots q_{mt}(\theta))$.

Assumption 1 requests a joint central limit theorem to hold for the sample moments and their derivative with respect to θ. It is satisfied under mild conditions which are listed in Kleibergen (2005), like, for example, finite r-th moments for $r > 2$, mixing conditions for the sample moments in case of time-series data. Allowing for a positive semi-definite covariance matrix $V(\theta)$ is important for applications, like, for example, linear dynamic panel data models. We next also use Assumption 2 from Kleibergen (2005) which concerns the convergence of the covariance matrix estimator $\hat{V}(\theta)$.

Assumption 2. *The convergence behavior of the covariance matrix estimator* $\hat{V}(\theta)$ *towards* $V(\theta)$ *is such that*

$$\hat{V}(\theta) \underset{p}{\to} V(\theta) \;\; \textit{and} \;\; \frac{\partial \mathrm{vec}(\hat{V}_{ff}(\theta))}{\partial \theta'} \underset{p}{\to} \frac{\partial \mathrm{vec}(V_{ff}(\theta))}{\partial \theta'}. \tag{29}$$

The CUE satisfies the FOC for a minimum of the CUE sample objective function.

Theorem 2: The FOC (divided by two) for a stationary point $\hat{\theta}^s$ of the CUE sample objective function reads:

$$\frac{1}{2}\frac{\partial}{\partial \theta'}\hat{Q}_s(\hat{\theta}^s) = 0 \quad \Leftrightarrow \quad f_T(\hat{\theta}^s, X)'\hat{V}_{ff}(\hat{\theta}^s)^{-1}\hat{D}(\hat{\theta}^s) = 0, \tag{30}$$

with

$$\hat{D}(\theta) = \quad q_T(\theta, X) - \left[\hat{V}_{\theta_1 f}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X) \dots \hat{V}_{\theta_m f}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X)\right] \tag{31}$$

and

$$\hat{V}(\theta) = \begin{pmatrix} \hat{V}_{ff}(\theta) & \hat{V}_{f\theta}(\theta) \\ \hat{V}_{\theta f}(\theta) & \hat{V}_{\theta\theta}(\theta) \end{pmatrix}, \tag{32}$$

with $\hat{V}_{\theta f}(\theta) = \hat{V}_{f\theta}(\theta)' = (\hat{V}_{\theta_1 f}(\theta)' \dots \hat{V}_{\theta_m f}(\theta)')'$, $\hat{V}_{\theta\theta}(\theta) = (\hat{V}_{\theta_i\theta_j}(\theta)) : i,j = 1,\dots,m$; and $\hat{V}_{ff}(\theta)$, $\hat{V}_{\theta_i f}(\theta)$, $\hat{V}_{\theta_i\theta_j}(\theta)$ are $k_f \times k_f$ dimensional matrices for $i,j = 1,\dots,m$.

Proof. It follows along the lines of the proof of Theorem 1; see also Kleibergen (2005). ■

Theorem 2 shows that the FOC of the sample CUE objective function can in an identical manner be factorized as the FOC of the population continuous updating objective function. Theorem 3 shows that the two components in which the FOC of the sample objective function factorizes are independently distributed in large samples.

Theorem 3: When Assumptions 1 and 2 hold and for θ^* the pseudo-true value minimizing the population continuous updating objective function:

$$\begin{aligned} \sqrt{T}\left(f_T(\theta^*, X) - \mu_f(\theta^*)\right) &\underset{d}{\rightarrow} \psi_f(\theta^*), \\ \sqrt{T}\mathrm{vec}\left(\hat{D}(\theta^*) - D(\theta^*)\right) &\underset{d}{\rightarrow} \psi_{\theta.f}(\theta^*), \end{aligned} \tag{33}$$

where $\psi_{\theta.f}(\theta^*) = \psi_\theta(\theta^*) - V_{\theta f}(\theta^*)V_{ff}(\theta^*)^{-1}\psi_f(\theta^*)$ and

$$\begin{aligned} \psi_f(\theta^*) &\sim N(0, V_{ff}(\theta^*)), \\ \psi_{\theta.f}(\theta^*) &\sim N(0, V_{\theta\theta.f}(\theta^*)), \end{aligned} \tag{34}$$

with $V_{\theta\theta.f}(\theta) = V_{\theta\theta}(\theta) - V_{\theta f}(\theta)V_{ff}(\theta)^{-1}V_{f\theta}(\theta)$, and $\psi_{\theta.f}(\theta^*)$ is independent of $\psi_f(\theta^*)$.

Proof. See the Online Appendix and Lemma 1 in Kleibergen (2005). ■

In standard GMM using the CUE objective function, the sample moment $f_T(\theta, X)$ is centered at zero at the true value so we can use different identification robust statistics, 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, b). In our misspecified GMM setting, the sample moment is not centered at zero so we can not use any of these statistics. We therefore propose a misspecification robust score statistic which uses that the expected value of the limit of the derivative of the sample objective function:

$$s(\theta) = \tfrac{1}{2}\tfrac{\partial}{\partial\theta'}\hat{Q}_s(\theta) = \quad f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta), \tag{35}$$

is equal to zero at the pseudo-true value θ^*.

Theorem 4: When Assumptions 1 and 2 hold, θ^* is the minimizer of the population continuous updating objective function and

$$\begin{aligned}\bar{\mu}_f(\theta^*) = & \quad E\left[\lim_{T\to\infty}\sqrt{T}f_T(\theta^*,X)\right] \\ \bar{D}(\theta^*) = & \quad E\left[\lim_{T\to\infty}\sqrt{T}\left(q_T(\theta^*,X)-\right.\right. \\ & \left.\left.\left[V_{\theta_1 f}(\theta^*)V_{ff}(\theta^*)^{-1}f_T(\theta^*,X)\ldots V_{\theta_m f}(\theta^*)V_{ff}(\theta^*)^{-1}f_T(\theta^*,X)\right]\right)\right],\end{aligned} \tag{36}$$

with $\bar{\mu}_f(\theta^*)$ and $\bar{D}(\theta^*)$ finite valued k_f and $k_f \times m$ dimensional continuously differentiable functions of θ^*, so $\bar{\mu}_f(\theta^*)'V_{ff}(\theta^*)^{-1}\bar{D}(\theta^*) \equiv 0$, the limit behavior of (half) the derivative of the CUE sample objective function at θ^* is characterized by:

$$\begin{aligned}Ts(\theta^*) \underset{d}{\to} & \quad \bar{\mu}_f(\theta^*)'V_{ff}(\theta^*)^{-1}\Psi_{\theta.f}(\theta^*) + \psi_f(\theta^*)'V_{ff}(\theta^*)^{-1}\left[\bar{D}(\theta^*)+\Psi_{\theta.f}(\theta^*)\right] \\ = & \quad \left(\bar{\mu}_f(\theta^*)+\psi_f(\theta^*)\right)'V_{ff}(\theta^*)^{-1}\Psi_{\theta.f}(\theta^*) + \psi_f(\theta^*)'V_{ff}(\theta^*)^{-1}\bar{D}(\theta^*),\end{aligned} \tag{37}$$

with $\mathrm{vec}(\Psi_{\theta.f}(\theta^*)) = \psi_{\theta.f}(\theta^*)$, so the expected value of the limit of the derivative of the sample CUE objective function is equal to zero at the pseudo-true value θ^* :

$$E\left[\lim_{T\to\infty} T \times s(\theta^*)\right] = 0. \tag{38}$$

Proof. See the Online Appendix. ■

Theorem 4 states in (37) two equivalent expressions of the limit behavior of the score of the CUE sample objective function. Each of these two expressions consists of two components which are products of independently distributed random variables. Since (at least) one of the random variables in these products has mean zero, the mean of the limit behavior of the score is equal to zero as well. Theorem 4 uses local to zero sequences for $\mu_f(\theta)$ and $D(\theta)$ which are orthogonal at the pseudo-true value θ^*. This is without loss of generality (wlog). We just use it to save on notation since it avoids that certain bounded random variables get multiplied by diverging objects which would imply that the expectation becomes ill defined.

We use the two limit expressions of the score in (37) to construct a weighting matrix which results in a size correct test based on a quadratic form of the score. To start out, we note that the second component of the first limit expression in (37) is the limit of the score used in the KLM test from Kleibergen (2005). We can therefore use, as in Kleibergen (2005), the conditional expectation of its outer product given $\bar{D}(\theta^*) + \Psi_{\theta.f}(\theta^*)$:

$$T\hat{D}(\theta^*)'\hat{V}_{ff}(\theta^*)^{-1}\hat{D}(\theta^*) \underset{d}{\rightarrow} E_{\psi_f(\theta^*)|\bar{D}(\theta^*)+\Psi_{\theta.f}(\theta^*)}(\left[\bar{D}(\theta^*)+\Psi_{\theta.f}(\theta^*)\right]' V_{ff}(\theta^*)^{-1}\psi_f(\theta^*)\psi_f(\theta^*)'$$
$$V_{ff}(\theta^*)^{-1}\left[\bar{D}(\theta^*)+\Psi_{\theta.f}(\theta^*)\right]|\bar{D}(\theta^*)+\Psi_{\theta.f}(\theta^*)=\sqrt{T}\hat{D}(\theta^*)). \quad (39)$$

since $\hat{V}_{ff}(\theta^*) \underset{p}{\rightarrow} V_{ff}(\theta^*)$, $\sqrt{T}\hat{D}(\theta^*) \underset{d}{\rightarrow} \bar{D}(\theta^*)+\Psi_{\theta.f}(\theta^*)$ so it provides an estimator of $\bar{D}(\theta^*)+\Psi_{\theta.f}(\theta^*)$, for this component in the weighting matrix.

Since $\sqrt{T}\hat{\mu}_f(\theta^*)=\sqrt{T}f_T(\theta^*,X) \underset{d}{\rightarrow} \bar{\mu}_f(\theta^*)+\psi_f(\theta^*)$, $\sqrt{T}\hat{\mu}_f(\theta^*)$ provides an estimator of $\bar{\mu}_f(\theta^*)+\psi_f(\theta^*)$, we can in a similar manner construct the conditional expectation of the first component of the second limit expression of the score in (37):

$$\begin{array}{r}T\left(I_m\otimes\hat{V}_{ff}(\theta^*)^{-1}\hat{\mu}_f(\theta^*)\right)'\hat{V}_{\theta\theta.f}(\theta^*)\left(I_m\otimes\hat{V}_{ff}(\theta^*)^{-1}\hat{\mu}_f(\theta^*)\right) \underset{d}{\rightarrow} \\ E_{\psi_{\theta.f}(\theta^*)|\bar{\mu}_f(\theta^*)+\psi_f(\theta^*)}\left((\bar{\mu}_f(\theta^*)+\psi_f(\theta^*))'V_{ff}(\theta^*)^{-1}\Psi_{\theta.f}(\theta^*)\Psi_{\theta.f}(\theta^*)'\right. \\ \left.V_{ff}(\theta^*)^{-1}(\bar{\mu}_f(\theta^*)+\psi_f(\theta^*))|\bar{\mu}_f(\theta^*)+\psi_f(\theta^*)=\sqrt{T}\hat{\mu}_f(\theta^*)\right) = \\ E_{\psi_{\theta.f}(\theta^*)|\bar{\mu}_f(\theta^*)+\psi_f(\theta^*)}\left(\left(I_m\otimes V_{ff}(\theta^*)^{-1}(\bar{\mu}_f(\theta^*)+\psi_f(\theta^*))\right)'\psi_{\theta.f}(\theta^*)\psi_{\theta.f}(\theta^*)'\right. \\ \left.\left(I_m\otimes V_{ff}(\theta^*)^{-1}(\bar{\mu}_f(\theta^*)+\psi_f(\theta^*))\right)|\bar{\mu}_f(\theta^*)+\psi_f(\theta^*)=\sqrt{T}\hat{\mu}_f(\theta^*)\right).\end{array} \quad (40)$$

Theorem 5 shows that we can use the sum of the components in (39) and (40) as a weighting matrix for a double robust score or Lagrange multiplier test.

Definition 1. The double robust score or Lagrange multiplier (DRLM) statistic for testing $\mathrm{H}_0:\theta=\theta^*$, with θ^* the pseudo-true value, is:

$$\begin{array}{l}DRLM(\theta^*)= \\ \quad T^2\times f_T(\theta^*,X)'\hat{V}_{ff}(\theta^*)^{-1}\hat{D}(\theta^*) \\ \quad \left[T\times\left(I_m\otimes\hat{V}_{ff}(\theta^*)^{-1}f_T(\theta^*,X)\right)'\hat{V}_{\theta\theta.f}(\theta^*)\left(I_m\otimes\hat{V}_{ff}(\theta^*)^{-1}f_T(\theta^*,X)\right)+\right. \\ \quad \left.T\times\hat{D}(\theta^*)'\hat{V}_{ff}(\theta^*)^{-1}\hat{D}(\theta^*)\right]^{-1}\hat{D}(\theta^*)'\hat{V}_{ff}(\theta^*)^{-1}f_T(\theta^*,X).\end{array} \quad (41)$$

Theorem 5: When Assumptions 1 and 2 hold and given the specifications in (36), the limit behavior of DRLM(θ^*) under $\mathrm{H}_0:\theta=\theta^*$, with θ^* the minimizer of the population continuous updating objective function, is bounded according to:

$$\lim_{T\to\infty} DRLM(\theta^*)\preceq\chi^2(m). \quad (42)$$

Proof. See the Online Appendix, which also provides an extension to Assumptions 1 and 2 by stating the parameter space of the distributions which render the DRLM test size correct; see also Andrews and Guggenberger (2017). ■

We next use the DRLM statistic to test the risk premia in the linear asset pricing model with i.i.d. errors.

Running example 1: Linear asset pricing model For a DRLM test of the risk premia, we need the specification of its different components for the linear asset pricing model with i.i.d. errors:

$$\begin{aligned} f_T(\lambda_F, X) = & \quad \bar{R} - \hat{\beta}\lambda_F \\ \hat{D}(\lambda_F) = & \quad -\hat{\beta} - (\bar{R} - \hat{\beta}\lambda_F)(1 + \lambda_F' \hat{Q}_{\bar{F}\bar{F}}^{-1} \lambda_F)^{-1} \lambda_F' \hat{Q}_{\bar{F}\bar{F}}^{-1} \\ = & \quad -\tfrac{1}{T} \textstyle\sum_{t=1}^{T} R_t (\bar{F}_t + \lambda_F)' \left[\tfrac{1}{T} \textstyle\sum_{t=1}^{T} (\bar{F}_t + \lambda_F)(\bar{F}_t + \lambda_F)' \right]^{-1} \\ \hat{V}_{ff}(\lambda_F) = & \quad (1 + \lambda_F' Q_{\bar{F}\bar{F}}^{-1} \lambda_F) \hat{\Omega} \\ \hat{V}_{\theta\theta.f}(\lambda_F) = & \quad (\hat{Q}_{\bar{F}\bar{F}} + \lambda_F \lambda_F')^{-1} \otimes \hat{\Omega}, \end{aligned} \tag{43}$$

so the specification of the DRLM test reads:

$$\begin{aligned} DRLM(\lambda_F^*) = & \quad T(1 + \lambda_F^{*\prime} \hat{Q}_{\bar{F}\bar{F}}^{-1} \lambda_F^*)^{-1} (\bar{R} - \hat{\beta}\lambda_F^*)' \hat{\Omega}^{-1} \hat{D}(\lambda_F^*) \\ & \Big[(1 + \lambda_F^{*\prime} \hat{Q}_{\bar{F}\bar{F}}^{-1} \lambda_F^*)^{-1} (\bar{R} - \hat{\beta}\lambda_F^*)' \hat{\Omega}^{-1} (\bar{R} - \hat{\beta}\lambda_F^*) (Q_{\bar{F}\bar{F}} + \lambda_F^* \lambda_F^{*\prime})^{-1} + \\ & \hat{D}(\lambda_F^*)' \hat{\Omega}^{-1} \hat{D}(\lambda_F^*) \Big]^{-1} \hat{D}(\lambda_F^*)' \hat{\Omega}^{-1} (\bar{R} - \hat{\beta}\lambda_F^*) \\ = & \quad \hat{\mu}(\lambda_F)^{*\prime} \hat{D}(\lambda_F^*)^* \left[\hat{\mu}(\lambda_F)^{*\prime} \hat{\mu}(\lambda_F)^* I_m + \hat{D}(\lambda_F^*)^{*\prime} \hat{D}(\lambda_F^*)^* \right]^{-1} \hat{D}(\lambda_F^*)^{*\prime} \hat{\mu}(\lambda_F)^*, \end{aligned} \tag{44}$$

with $\hat{\mu}(\lambda_F)^* = \sqrt{T} \hat{\Omega}^{-\frac{1}{2}} (\bar{R} - \hat{\beta}\lambda_F)(1 + \lambda_F' \hat{Q}_{\bar{F}\bar{F}}^{-1} \lambda_F)^{-\frac{1}{2}} = \sqrt{T} \hat{V}_{ff}(\lambda_F)^{-\frac{1}{2}} f_T(\lambda_F, X)$, and $\hat{D}(\lambda_F)^* = \sqrt{T} \hat{\Omega}^{-\frac{1}{2}} \hat{D}(\lambda_F)(\hat{Q}_{\bar{F}\bar{F}} + \lambda_F \lambda_F')^{\frac{1}{2}}$.

Corollary 1. When Assumptions 1 and 2 hold and under i.i.d. errors, the limit behavior of the DRLM statistic under $\mathrm{H}_0 : \lambda_F = \lambda_F^*$ is characterized by:

$$\begin{aligned} DRLM(\lambda_F^*) \underset{d}{\to} & \quad \left[\psi_f' (\bar{D} + \Psi_{\theta.f}) + \bar{\mu}' \Psi_{\theta.f} \right] [(\bar{\mu} + \psi_f)'(\bar{\mu} + \psi_f) I_m + \\ & \left(\bar{D} + \Psi_{\theta.f} \right)' \left(\bar{D} + \Psi_{\theta.f} \right) \Big]^{-1} \left[(\bar{D} + \Psi_{\theta.f})' \psi_f + \Psi_{\theta.f}' \bar{\mu} \right] \\ \preceq & \quad \chi^2(m), \end{aligned} \tag{45}$$

with $\bar{\mu} = \Omega^{-\frac{1}{2}} \bar{\mu}(\lambda_F^*)(1 + \lambda_F^{*\prime} Q_{\bar{F}\bar{F}}^{-1} \lambda_F^*)^{-\frac{1}{2}}$, $\bar{D} = \Omega^{-\frac{1}{2}} \bar{D}(\lambda_F^*)(Q_{\bar{F}\bar{F}} + \lambda_F^* \lambda_F^{*\prime})^{\frac{1}{2}}$, $\bar{\mu}' \bar{D} \equiv 0$ and ψ_f and $\Psi_{\theta.f}$ $N \times 1$ and $N \times m$ dimensional random matrices that consist of independent standard normal random variables.

The limit behavior of the DRLM statistic in Corollary 1 shows that it under H_0 only depends on two parameters, the lengths of $\bar{\mu}$ and $\bar{D}$ and is dominated by a $\chi^2(m)$ distribution. Figure 1 shows the rejection frequencies of 5% significance DRLM tests with a 95% $\chi^2(1)$ critical value as a function of the lengths of $\bar{\mu}$ and $\bar{D}$ for a one factor setting, so $m = 1$, and $N = 25$. The latter number corresponds with the twenty-five Fama-French size and book-to-market sorted portfolios which are the default in the asset pricing literature, see Fama and French (1993).

Figure 1: Rejection frequency of 5% significance DRLM tests of $\mathrm{H}_0 : \lambda_F = \lambda_F^*$ using a 95% $\chi^2(1)$ critical value as a function of the lengths of $\bar{\mu}$ and $\bar{D}$, $m = 1$, $N = 25$.

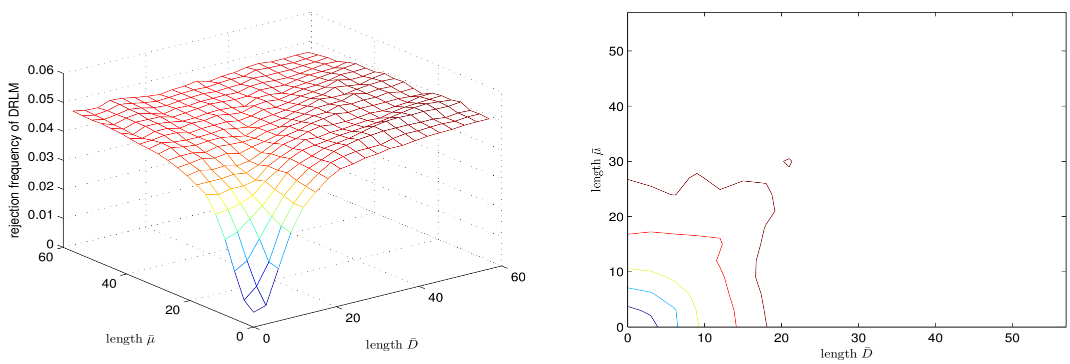


Figure 2: Rejection frequency of 5% significance KLM tests of $\mathrm{H}_0 : \lambda_F = \lambda_F^*$ using a 95% $\chi^2(1)$ critical value as a function of the lengths of $\bar{\mu}$ and $\bar{D}$, $m = 1$, $N = 25$.

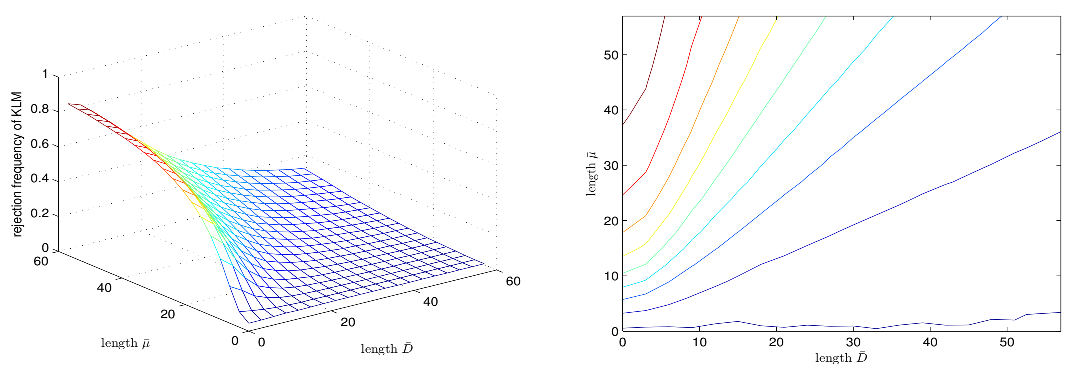
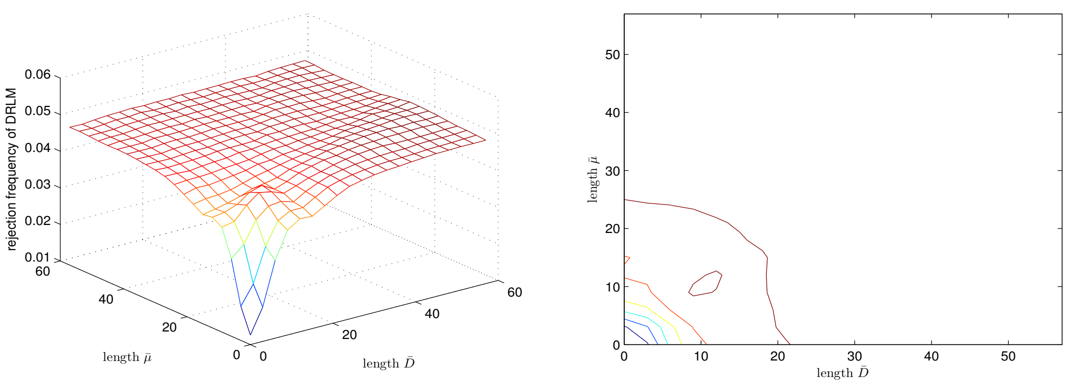


Figure 1 shows that the DRLM test is size correct since its rejection frequency does not exceed 5% for any length of $\bar{\mu}$ and $\bar{D}$. For comparison, Figure 2 presents the rejection frequencies of the

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. To reduce the conservativeness of the DRLM test at these low values, we calibrate a feasible conditional critical value function based on the maximum of $\hat{\mu}(\lambda_F)^{*\prime}\hat{\mu}(\lambda_F)^*$ and $\hat{D}(\lambda_F)^{*\prime}\hat{D}(\lambda_F)^*$. When the maximum of these is less than two-hundred and fifty, we calibrated a 95% conditional critical value function based on $\max(\hat{\mu}(\lambda_F)^{*\prime}\hat{\mu}(\lambda_F)^*,\ \hat{D}(\lambda_F)^{*\prime}\hat{D}(\lambda_F)^*)$.[3] The contour lines in Figure 3 show that the conservativeness of a 5% significance DRLM test has been reduced substantially from an area where the maximal length of $\bar{\mu}$ and $\bar{D}$ is less than twenty to an area where their sum is less than ten.

Figure 3: Rejection frequency of 5% significance DRLM tests of $\mathrm{H}_0 : \lambda_F = \lambda_F^*$ using a conditional 95% critical value as a function of the lengths of $\bar{\mu}$ and $\bar{D}$, $m = 1$, $N = 25$.



4 Power

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. This leads to the somewhat oddly behaved power of the KLM test in regular GMM. Tests with better power properties therefore exist in 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 et al. (2006), Andrews (2016) and Andrews and

[3] The conditional critical value function we calibrated for Figure 3 is $f(r) = 2.4 + (\lfloor r \rfloor^{0.35}) \times (3.84 - 2.4)/(250^{0.35})$ for $r \leq 250$ and $f(r) = 3.84$ for $r > 250$, with r the conditioning variable and $\lfloor . \rfloor$ the entier function.

Mikusheva (2016a, b). In our misspecified GMM setting, this is, however, not possible since the limiting distribution of the J-test is a non-central χ^2 distribution with an unknown non-centrality parameter. Hence, we can not combine this limiting distribution with that of the DRLM test to obtain the (conditional) critical values for a combination test.

To improve the power of a $100 \times \alpha\%$ significance DRLM test, we can reject hypothesized values of θ for which a $100 \times \alpha\%$ significance DRLM test is not significant but 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 is non-significant at such values of θ, J and/or GMM Anderson-Rubin (AR) tests, see Anderson and Rubin (1949) and Stock and Wright (2000), can be significant. For hypothesized values of θ close to the CUE, these combination tests put most weight on the KLM test but shift the weight towards 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 tests depend on unknown nuisance parameters in our misspecified GMM setting, it is not clear how we can use these tests to improve power. To improve the power of a $100 \times \alpha\%$ significance DRLM test, we can reject values of θ for which the DRLM test is not significant at the $100 \times \alpha\%$ level but which are on a line from the hypothesized value to the CUE where in between the hypothesized value and the CUE there are significant values of the DRLM test. We next lay out the different steps needed to turn this into a size correct test for stylized linear GMM settings.

Theorem 6: **a.** For a given data set of realized values and a linear moment equation, the sum of $f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)$ and $\text{vec}(\hat{D}(\theta))'\hat{V}_{\theta\theta.f}(\theta)^{-1}\text{vec}(\hat{D}(\theta))$ does not vary over θ.

b. When $m = 1$ and $f_T(\theta,X)$ is linear in θ, the derivative of DRLM(θ) with respect to θ reads:

$$\begin{aligned}
&\tfrac{1}{2}\tfrac{\partial}{\partial\theta}DRLM(\theta) = T\left(\tfrac{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)}{\left[f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta.f}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)+\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)\right]}\right)\times \\
&\quad\left\{\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) - 2f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta f}(\theta)\hat{V}_{ff}(\theta)^{-1}D_T(\theta,X)-\right. \\
&\quad f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta.f}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X) + 2f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)\times \\
&\quad\left.\tfrac{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta f}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta.f}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)+\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta f}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)}{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta.f}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)+\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)}\right\}.
\end{aligned} \tag{46}$$

c. When the data is i.i.d., $m = 1$, and $f_T(\theta,X)$ is linear in $\theta : \hat{V}(\theta)$ has a Kronecker product structure so we can specify $\hat{V}_{ff}(\theta) = \hat{v}_{ff}(\theta)\hat{V}$, $\hat{V}_{\theta f}(\theta) = \hat{v}_{\theta f}(\theta)\hat{V}$ and $\hat{V}_{\theta\theta.f}(\theta) = \hat{v}_{\theta\theta.f}(\theta)\hat{V}$, with $\hat{v}_{ff}(\theta)$, $\hat{v}_{\theta f}(\theta)$, $\hat{v}_{\theta\theta.f}(\theta)$ scalar functions of θ and $\hat{V}$ a $k_f \times k_f$ dimensional covariance matrix estimator, and the derivative of DRLM(θ) is:

$$\frac{1}{2}\frac{\partial}{\partial\theta}DRLM(\theta)=\left(\frac{\left(\hat{V}_{ff}(\theta)^{-\frac{1}{2}}f_T(\theta,X)\right)'\left(\hat{V}_{\theta\theta.f}(\theta)^{-\frac{1}{2}}\hat{D}(\theta)\right)}{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)+\hat{D}(\theta)'\hat{V}_{\theta\theta.f}(\theta)^{-1}\hat{D}(\theta)}\right)\times$$
$$\left(T\times\hat{D}(\theta)'\hat{V}_{\theta\theta.f}(\theta)^{-1}\hat{D}(\theta)-T\times f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)\right)\left(\frac{\hat{v}_{\theta\theta.f}(\theta)}{\hat{v}_{ff}(\theta)}\right)^{\frac{1}{2}}.$$

Proof. See the Online Appendix. ■

Running example 1: Linear asset pricing model Theorem 6c shows that for the one factor linear asset pricing model with i.i.d. errors, the derivative of the DRLM statistic is proportional to the difference between the GMM-AR statistic, $T\times f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)$, and an independently distributed statistic reflecting the strength of identification, $T\times\hat{D}(\theta)'\hat{V}_{\theta\theta.f}(\theta)^{-1}\hat{D}(\theta)$. Theorem 6a further shows that, for a given data set of realized values, the sum of these two statistics does not depend on θ. Given a realized data set, the DRLM statistic considered as a function of θ thus attains its maximum when both statistics are identical so they equal half their sum.

Corollary 2. For the one factor linear asset pricing model with i.i.d. errors, the maximal value of the DRLM statistic as a function of λ_F is attained at the value of λ_F where the GMM-AR statistic, $T\times f_T(\lambda_F,X)'\hat{V}_{ff}(\lambda_F)^{-1}f_T(\lambda_F,X)$, equals half the sum of $T\times f_T(\lambda_F,X)'\hat{V}_{ff}(\lambda_F)^{-1}f_T(\lambda_F,X)$ and $T\times\hat{D}(\lambda_F)'\hat{V}_{\theta\theta.f}(\lambda_F)^{-1}\hat{D}(\lambda_F)$.

Using Corollary 2 and the sample equivalent of the characteristic polynomial in (16), we can solve for the value of λ_F that maximizes the DRLM statistic for a given data set of realized values. We do so by not equating the characteristic polynomial to zero but to half the sum of $T\times f_T(\lambda_F,X)'\hat{V}_{ff}(\lambda_F)^{-1}f_T(\lambda_F,X)$ and $T\times\hat{D}(\lambda_F)'\hat{V}_{\theta\theta.f}(\lambda_F)^{-1}\hat{D}(\lambda_F)$, which, as stated in Theorem 6a, is constant over λ_F. We can then straightforwardly solve for the value of λ_F that maximizes the DRLM statistic in a data set of realized values. We can next use this maximizer to improve the power of $100\times\alpha\%$ significance DRLM tests.

The power of a $100\times\alpha\%$ significance DRLM test of $\mathrm{H}_0:\lambda_F=\lambda_F^1$ can be improved by rejecting H_0 alongside for significant values of DRLM(λ_F^1) also when both:

1. The maximal value of the DRLM statistic for the analyzed data set is significant at the $100\times\alpha\%$ level.

2. The DRLM statistic evaluated at λ_F^1 is insignificant at the $100\times\alpha\%$ level but λ_F^1 lies within the closed interval indicated by the maximizers of the DRLM statistic that does not contain the CUE.

The above algorithm rejects H_0 alongside for significant values of DRLM(λ_F^1) also when there is a significant value of the DRLM statistic on the line between λ_F^1 and the CUE. To show that the above algorithm leads to a size correct test, we compute its rejection frequency when testing $\mathrm{H}_0 : \lambda_F = 0$ using the setup from Figures 1-3. While the generic specification of the DRLM statistic tests for a stationary point of the population continuous updating objective function, the above algorithm explicitly tests the minimizer. 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 1-3, which uses the limit expression of the DRLM test in (45), the population minimizer is at zero if the 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 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.

Figure 4: Rejection frequency of 5% significance tests of $\mathrm{H}_0 : \lambda_F = 0$ using power improved DRLM and a conditional 95% critical value as a function of the lengths of $\bar{\mu}$ and $\bar{D}$, $m = 1$, $N = 25$.

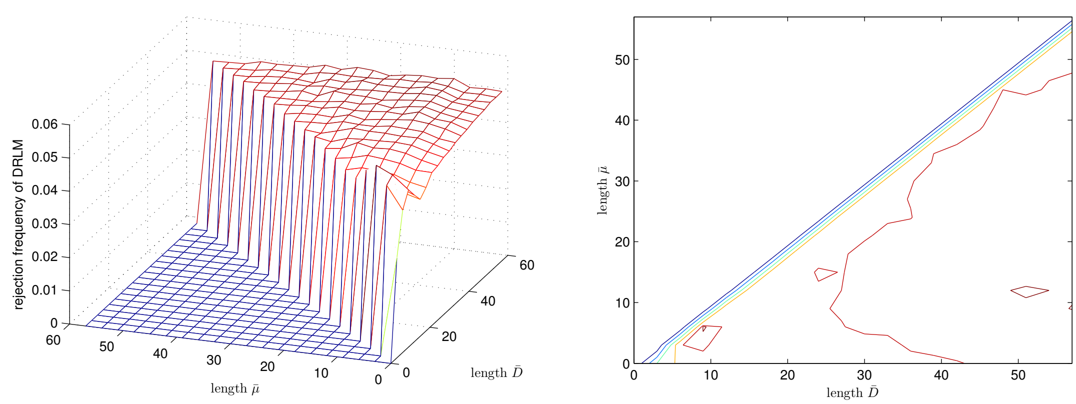


Figure 4 shows the rejection frequency of the power improved DRLM test when the minimizer of the population continuous updating objective function equals the hypothesized value which is zero. Figure 4 does therefore not show the rejection frequency for values where the length of $\bar{\mu}$ exceeds that 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 Figure 4 are computed using the conditional critical values explained previously. Figure 4 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.

Power analysis We use the one factor linear asset pricing model to compare the power and size of different identification robust test procedures with that of the DRLM test. For the power analysis, the minimizer of the population continuous updating objective function is the pseudo-true value λ_F^* while we test for a zero value under the null hypothesis. We then map out the power curve by changing the pseudo-true value and keeping the hypothesized value, zero, fixed. Theorem 7 states the limiting distributions of the different components of the DRLM statistic for testing the hypothesis of interest used for the power analysis.

Theorem 7: For testing $\mathrm{H}_0 : \lambda_F = \lambda_F^1 = 0$, the limit behavior of the components of the DRLM statistic in the one factor linear asset pricing model with i.i.d. errors, $m = 1$ and $Q_{\bar{F}\bar{F}} = 1$, while the pseudo-true value equals λ_F^*, are characterized by:

$$\begin{array}{rcl} \sqrt{T}\hat{\Omega}^{-\frac{1}{2}}\bar{R} & \underset{d}{\rightarrow} & \bar{\mu}(1+(\lambda_F^*)^2)^{-\frac{1}{2}} - \bar{D}(1+(\lambda_F^*)^2)^{-\frac{1}{2}}\lambda_F^* + \psi_f^*(\lambda_F^1 = 0) \\ \sqrt{T}\hat{\Omega}^{-\frac{1}{2}}\hat{D}(\lambda_F^1 = 0) & \underset{d}{\rightarrow} & \bar{D}(1+(\lambda_F^*)^2)^{-\frac{1}{2}} + \bar{\mu}(1+(\lambda_F^*)^2)^{-\frac{1}{2}}\lambda_F^* + \psi_{\theta.f}^*(\lambda_F^1 = 0), \end{array} \tag{47}$$

with $\psi_f^*(\lambda_F^1 = 0)$, $\psi_{\theta.f}^*(\lambda_F^1 = 0)$ independent standard normal N dimensional random vectors, $\mu^* = \lim_{T\to\infty}\sqrt{T}\mu_f(\lambda_F^*)$, $\mu_f(\lambda_F^*) = \mu_R - \beta\lambda_F^*$, $D^* = \lim_{T\to\infty}\sqrt{T}D(\lambda_F^*)$, $D(\lambda_F^*) = -\beta - \mu_f(\lambda_F^*)\lambda_F^{*\prime}(Q_{\bar{F}\bar{F}} + \lambda_F^*\lambda_F^{*\prime})^{-1}$, $\bar{\mu} = \Omega^{-\frac{1}{2}}\mu^*(1+\lambda_F^{*\prime}Q_{\bar{F}\bar{F}}^{-1}\lambda_F^*)^{-\frac{1}{2}}$, $\bar{D} = \Omega^{-\frac{1}{2}}D^*(Q_{\bar{F}\bar{F}} + \lambda_F^*\lambda_F^{*\prime})^{\frac{1}{2}}$, so $\bar{\mu}'\bar{D} \equiv 0$.

Proof. See the Online Appendix. ■

The specification in Theorem 7 is such that, since $\bar{\mu}'\bar{D} \equiv 0$, λ_F^* is the minimizer of the population continuous updating objective function when the length of $\bar{D}$, which reflects the strength of identification, is larger than or equal to the length of $\bar{\mu}$, which reflects misspecification. The product of the limit behavior of both components in (47):

$$\begin{array}{l} T\bar{R}'\hat{\Omega}^{-1}\hat{D}(\lambda_F^1 = 0) \underset{d}{\rightarrow} (1+(\lambda_F^*)^2)^{-1}\lambda_F^*\left(\bar{\mu}'\bar{\mu} - \bar{D}'\bar{D}\right) + \\ \quad (1+(\lambda_F^*)^2)^{-\frac{1}{2}}\left[\psi_f^*(\lambda_F^1 = 0)'\left(\bar{D} + \bar{\mu}\lambda_F^*\right) + \psi_{\theta.f}^*(\lambda_F^1 = 0)'\left(\bar{\mu} - \bar{D}\lambda_F^*\right)\right], \end{array} \tag{48}$$

further shows that identification is problematic when the lengths of $\bar{\mu}$ and $\bar{D}$ are equal so the misspecification and identification strengths are identical.

We next analyze the power of identification robust test statistics and the DRLM test for a number of settings of misspecification: no misspecification; weak misspecification; and mild misspecification.

No misspecification We first compare the power of the DRLM test with existing identification robust tests when no misspecification is present so all of these tests are size correct. The figures in Panels 5-6 and Figure 7 show the different power curves. Panel 5 shows the power curves of the KLM test of Kleibergen (2002, 2005, 2009) and the DRLM test for various identification strengths and no misspecification. The power of the KLM test is known to be non-monotonic which is in line with Figure 5.1. Figure 5.2 shows that power curves of the DRLM test are non-monotonic as well.

Panel 5: Power of 5% significance KLM and DRLM tests of

$\mathrm{H}_0 : \lambda_F = 0$ with no misspecification, $N = 25,\ Q_{\bar{F}\bar{F}} = 1$

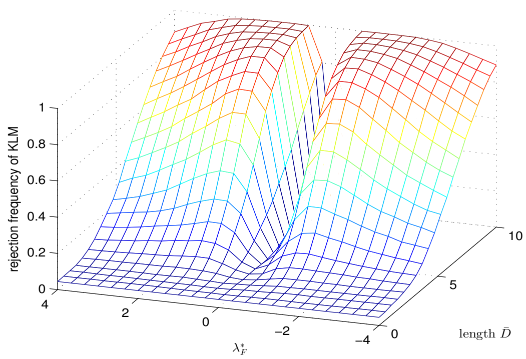


Figure 5.1: KLM

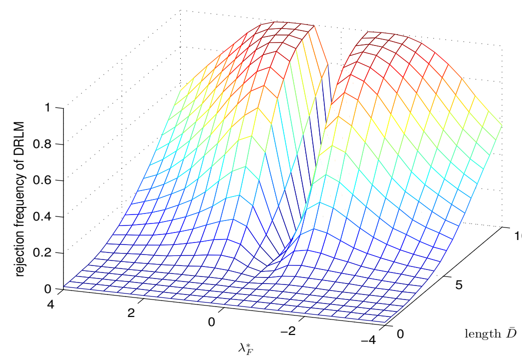


Figure 5.2: DRLM

Panel 6: Power of 5% significance LR and size and power improved

DRLM tests of $\mathrm{H}_0 : \lambda_F = 0$ with no misspecification, $N = 25,\ Q_{\bar{F}\bar{F}} = 1$

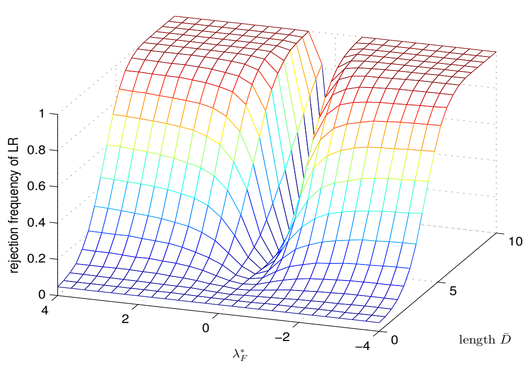


Figure 6.1: LR

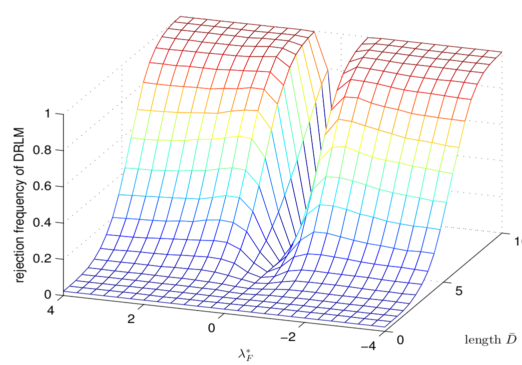
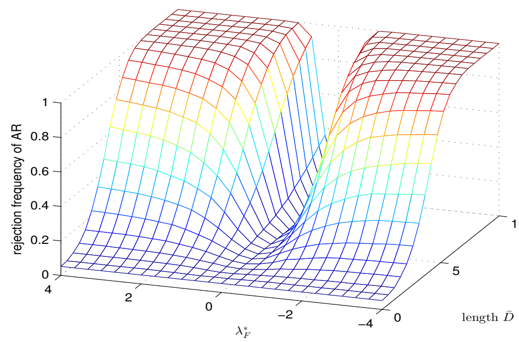


Figure 6.2: DRLM with size and power improvements

Figure 6.2 in Panel 6 shows that the size and power improved DRLM test, which uses the size and power improvement procedures discussed previously, has a nearly monotonic power curve. Figure 6.1 in Panel 6 shows power curves of the conditional likelihood ratio (LR) test of Moreira (2003) which is known to be optimal for this setting, see Andrews et al. (2006). Figure 7 shows power curves of the factor Anderson-Rubin (AR) test, see Anderson and Rubin (1949) and Kleibergen (2009).

Figure 7: Power of 5% signifance factor AR test

of $\mathrm{H}_0 : \lambda_F = 0$ with no misspecification, $N = 25,\ Q_{\bar{F}\bar{F}} = 1$



Weak misspecification We next compare the power of the different test procedures in a setting of weak misspecification where $\bar{\mu}'\bar{\mu} = 4.4$. Figures 8.1 and 8.2 in Panel 8 therefore show power curves of the KLM and DRLM tests for various identification strengths while Figures 9.1 and 9.2 in Panel 9 show power curves of the LR and size and power corrected DRLM test. Figure 10 shows power curves of the AR test. The power curves of the different test procedures are comparable to the ones in the previous figures except that we observe size distortion of the identification robust AR, KLM and LR tests. Except for the AR test, these size distortions become less when the identification strength increases. For the conditional LR test the rejection frequency at zero decreases from 15% to 9% when the identification strength increases. It equals 13% when the misspecification and identification strengths are identical. For the KLM test, it decreases from 7% to 5%. For the AR test, the rejection frequency at zero equals 15% for all settings of the identification strength since no estimator of the identification strength is involved in the AR test. For the DRLM and size and power improved DRLM tests, we observe no size distortion.

What is striking is that, for small values of the identification strength, the power of the identification robust AR and LR tests decreases when λ_F^* moves away from zero. This results since when the misspecification strength exceeds the identification strength, the population continuous updating objective function is maximized at zero instead of minimized. The population continuous updating objective function is then minimized when λ_F equals $\pm\infty$. When the strength of identification equals zero, so the length of $\bar{D}=0$, the moment equation (11) is, however, still not satisfied at these values of λ_F so the LR, KLM and AR tests remain size distorted even at these values. Moving away from zero at these settings of the identification strength, however, in general reduces the sample continuous updating objective function which then leads to a lower rejection frequency of these tests. For values of the identification strength exceeding the misspecification, the population continuous updating objective function is minimized at zero so we then no longer observe a reduction of the rejection frequency when λ_F^* moves away from zero.

Panel 8: Power of 5% significance KLM and DRLM tests of
$\text{H}_0 : \lambda_F = 0$ with misspecification, $\bar{\mu}'\bar{\mu} = 4.4,\ N = 25,\ Q_{\bar{F}\bar{F}} = 1$

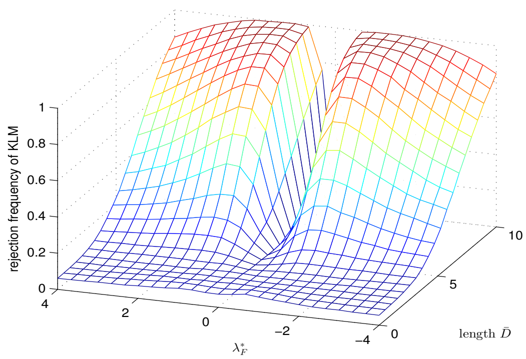


Figure 8.1: KLM

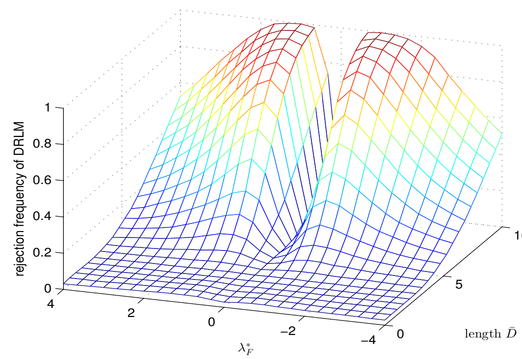


Figure 8.2: DRLM

Panel 9: Power of 5% significance LR and size and power improved
DRLM tests of $\text{H}_0 : \lambda_F = 0$ with misspecification, $\bar{\mu}'\bar{\mu} = 4.4,\ N = 25,\ Q_{\bar{F}\bar{F}} = 1$

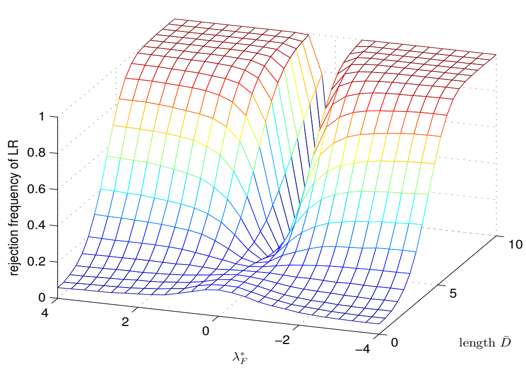


Figure 9.1: LR

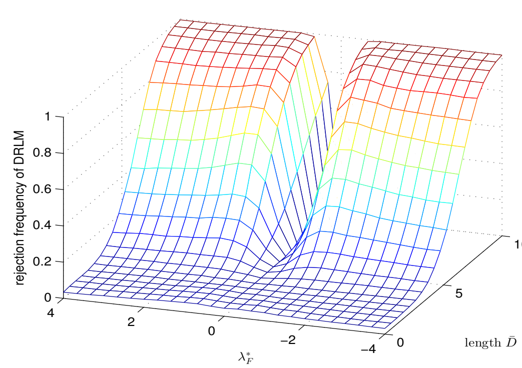
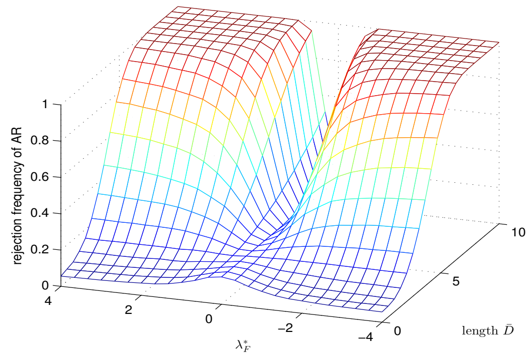


Figure 9.2: DRLM with size and power improvements

Figure 10: Power of 5% significance AR tests of $\mathrm{H}_0 : \lambda_F = 0$

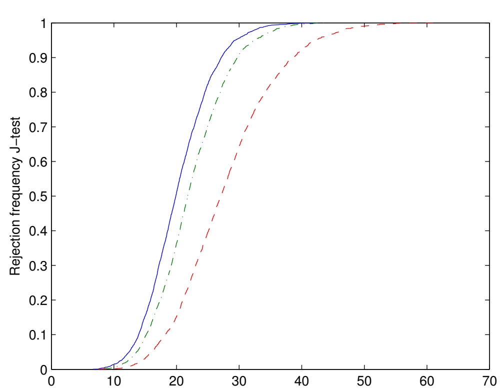
with misspecification, $\bar{\mu}'\bar{\mu} = 4.4,\ N = 25,\ Q_{\bar{F}\bar{F}} = 1$



To show the difficulty of detecting the weak misspecification used in Panels 8-9 and Figure 10, Figure 11 shows the simulated distribution function of the misspecification J-test, which equals the minimal value of the AR test for the simulated data, when the null hypothesis holds, so for values of λ_F^* equal to zero. Figure 11 shows the distribution function of the misspecification J-test for three different values of the identification strength $\bar{D}'\bar{D}$: 0, 4.4 and 100. In Guggenberger et al. (2012), it is shown that the distribution function of the J-test is a non-increasing function of the identification strength. Recognizing that the 95% critical value of the $\chi^2(24)$ distribution, since $N-1 = 24$, equals 36.42, Figure 11 shows that we never reject no misspecification at the 5% significance level when $\bar{D}'\bar{D}$ equals 0 or 4.4 and we only do so in 15% of the cases when $\bar{D}'\bar{D}$ equals 100. This illustrates the difficulty of detecting weak misspecification.

Figure 11: Distribution function of J-test for misspecification when $\mathrm{H}_0 : \lambda_F = 0$ holds,

solid line: $\bar{D}'\bar{D} = 0,$ dash-dot: $\bar{D}'\bar{D} = 4.4 =$ strength of misspecification, dashed: $\bar{D}'\bar{D} = 100.$



Mild misspecification We next increase the amount of misspecification to $\bar{\mu}'\bar{\mu} = 10$, which is still quite small since there are twenty-five moment equations. Panels 12-13 and Figure 14 show that the increased misspecification exacerbates the size distortion of the AR, KLM and LR tests compared to the previous setting of weak misspecification. For the conditional LR test, the rejection frequency at zero decreases from 30% to 8% when the identification strength increases. When the misspecification and identification strengths coincide, the rejection frequency of the LR test is 27% when $\lambda_F^* = 0$. For the KLM test, the rejection frequency decreases from 10% to 5%. For the DRLM and size and power improved DRLM test, we observe either no size distortion and a rejection frequency of 8% which decreases to 5% when the identification strength increases. The minor size distortion of the size and power improved DRLM test only occurs when the misspecification exceeds the strength of identification, so the hypothesized value is not the minimizer of the population objective function, and is not present when the identification strength is larger than or equal to the misspecification. The rejection frequency of the AR test is equal to 36% for all identification strengths. When the misspecification strength exceeds the identification strength, the maximum of the population continuous updating objective function is situated at $\lambda_F^* = 0$, which explains why the rejection frequency of the AR and LR tests decreases away from $\lambda_F^* = 0$ for low values of the identification strength. For values of the identification strength which exceed the misspecification strength, we see no decrease of the rejection frequency when λ_F^* moves away from zero.

Panel 12: Power of 5% significance KLM and DRLM tests of

$H_0 : \lambda_F = 0$ with misspecification, $\bar{\mu}'\bar{\mu} = 10,\ N = 25,\ Q_{\bar{F}\bar{F}} = 1$

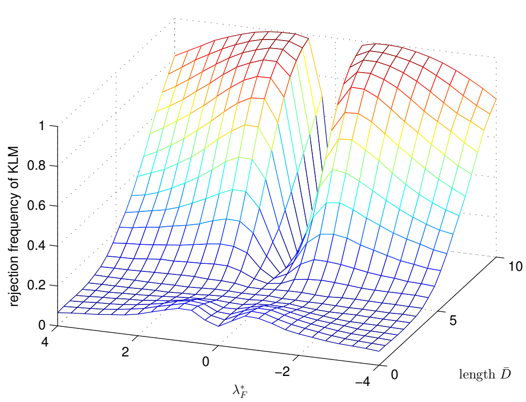


Figure 12.1: KLM

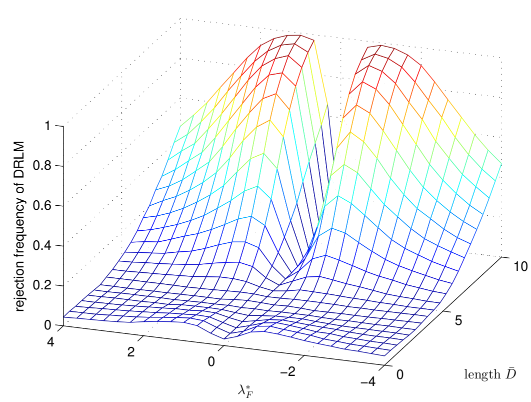


Figure 12.2: DRLM

Panel 13: Power of 5% significance LR and size and power improved

DRLM tests of $\mathrm{H}_0 : \lambda_F = 0$ with misspecification, $\bar{\mu}'\bar{\mu} = 10,\ N = 25,\ Q_{\bar{F}\bar{F}} = 1$

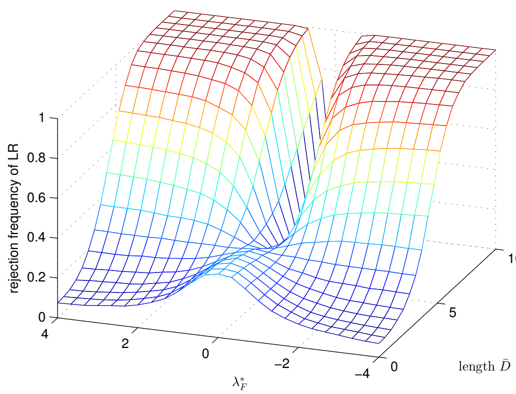


Figure 13.1: LR

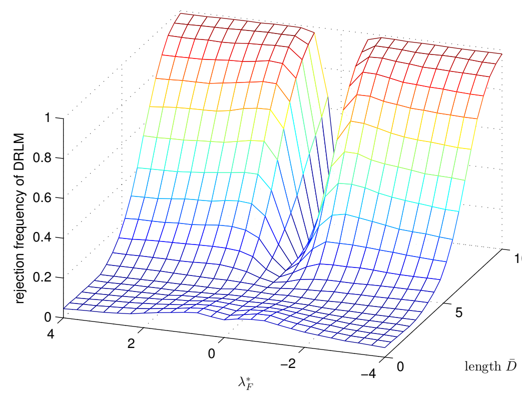


Figure 13.2: DRLM with size and power improvements

Panel 14: Power of 5% significance AR tests of $\mathrm{H}_0 : \lambda_F = 0$

with misspecification, $\bar{\mu}'\bar{\mu} = 10,\ N = 25,\ Q_{\bar{F}\bar{F}} = 1$

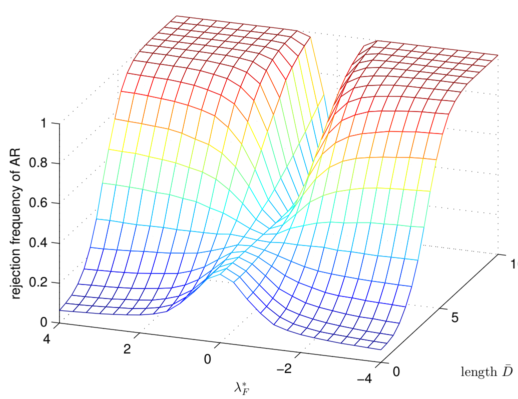
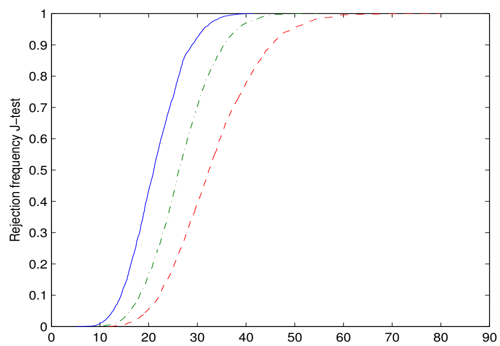


Figure 15 shows the distribution function of the misspecification J-test, which equals the minimal value of the AR test, when the null hypothesis holds, so for values of λ_F^* equal to zero. It shows the distribution function for three different values of the identification strength $\bar{D}'\bar{D}$: 0, 10 and 100. Recognizing that the 95% critical value of the $\chi^2(24)$ distribution, since $N - 1 = 24$, equals 36.42, Figure 15 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 the mild misspecification present in the simulated data. To show that the power issues discussed previously for both the identification robust tests and the misspecification J-test do not result from the somewhat large number of moment equations, 25, we discuss a somewhat smaller simulation experiment with fewer moment conditions in the Online Appendix.

Figure 15: Distribution function of J-test for misspecification when $\mathrm{H}_0 : \lambda_F = 0$ holds, solid line: $\bar{D}'\bar{D} = 0$, dash-dot: $\bar{D}'\bar{D} = 10 =$ strength of misspecification, dashed: $\bar{D}'\bar{D} = 100$.



More power improvements? We further analyze the power of invariant tests for which we use that they are a function of the maximal invariant. We therefore construct the maximal invariant for a stylized setting of the linear asset pricing model with independent normal errors and a fixed number of observations. In order to do so, we first conduct a singular value decomposition of $\Omega^{-\frac{1}{2}} \begin{pmatrix} \ddot{\mu}_R & \ddot{\beta} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & Q_{\bar{F}\bar{F}}^{\frac{1}{2}} \end{pmatrix}$, with $\ddot{\mu}_R = \sqrt{T}\mu_R$, $\ddot{\beta} = \sqrt{T}\beta$, which is invariant to transformations and whose sample estimator has an identity covariance matrix.

Theorem 8: A singular value decomposition of $\Omega^{-\frac{1}{2}} \begin{pmatrix} \ddot{\mu}_R & \ddot{\beta} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & Q_{\bar{F}\bar{F}}^{\frac{1}{2}} \end{pmatrix}$ results in:

$$\begin{array}{l} \Omega^{-\frac{1}{2}} \begin{pmatrix} \ddot{\mu}_R & \ddot{\beta} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & Q_{\bar{F}\bar{F}}^{\frac{1}{2}} \end{pmatrix} = \mathcal{USV}' = \\ -\Omega^{-\frac{1}{2}} D(\lambda_F^*) \begin{pmatrix} \lambda_F^* & I_m \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & Q_{\bar{F}\bar{F}}^{\frac{1}{2}} \end{pmatrix} + \Omega^{\frac{1}{2}} D(\lambda_F^*)_{\perp} \delta \begin{pmatrix} \lambda_F^* & I_m \end{pmatrix}_{\perp} \begin{pmatrix} 1 & 0 \\ 0 & Q_{\bar{F}\bar{F}}^{-\frac{1}{2}} \end{pmatrix}, \end{array} \tag{49}$$

which makes $D(\lambda_F^*)$ proportional to the expression provided below (18) when the latter is evaluated at λ_F^*, $\mathcal{U}$ an $N \times N$ dimensional orthonormal matrix, $\mathcal{V}$ an $(m+1) \times (m+1)$ dimensional orthonormal matrix, and S an $N \times (m+1)$ dimensional diagonal matrix with the singular values in decreasing order on the main diagonal:

$$\mathcal{U}=\begin{pmatrix}\mathcal{U}_{11} & \mathcal{U}_{12}\\ \mathcal{U}_{21} & \mathcal{U}_{22}\end{pmatrix},\ S=\begin{pmatrix}\mathcal{S}_{1} & 0\\ 0 & \mathcal{S}_{2}\end{pmatrix}\text{ and }\mathcal{V}=\begin{pmatrix}\mathcal{V}_{11} & \mathcal{V}_{12}\\ \mathcal{V}_{21} & \mathcal{V}_{22}\end{pmatrix}, \tag{50}$$

where $\mathcal{U}_{11}$, $\mathcal{S}_1$, $\mathcal{V}_{21}$ are $m\times m$ dimensional matrices; $\mathcal{S}_2$ is an $(N-m)\times 1$ dimensional matrix, $\mathcal{V}_{11}'$, $\mathcal{V}_{22}$ are $m\times 1$ dimensional vectors, $\mathcal{U}_{12}, \mathcal{U}_{21}$, and $\mathcal{U}_{22}$ are $m\times(N-m)$, $(N-m)\times m$ and $(N-m)\times(N-m)$ dimensional matrices and $\mathcal{V}_{12}$ is a scalar. The $N\times(N-m)$ dimensional matrix $D(\lambda_F^*)_\perp$ is the orthogonal complement of $D(\lambda_F^*)$, $D(\lambda_F^*)_\perp' D(\lambda_F^*)\equiv 0$, $D(\lambda_F^*)_\perp'\Omega D(\lambda_F^*)_\perp\equiv I_{N-m}$; $\begin{pmatrix}\lambda_F^* & I_m\end{pmatrix}_\perp$ the $1\times(m+1)$ dimensional orthogonal complement of $\begin{pmatrix}\lambda_F^* & I_m\end{pmatrix}$, $\begin{pmatrix}\lambda_F^* & I_m\end{pmatrix}\begin{pmatrix}\lambda_F^* & I_m\end{pmatrix}_\perp'\equiv 0$ and $\begin{pmatrix}\lambda_F^* & I_m\end{pmatrix}_\perp\begin{pmatrix}1 & 0\\ 0 & Q_{\bar F\bar F}^{-1}\end{pmatrix}\begin{pmatrix}\lambda_F^* & I_m\end{pmatrix}_\perp'\equiv 1$, $\begin{pmatrix}\lambda_F^* & I_m\end{pmatrix}_\perp=\begin{pmatrix}1 & -\lambda_F^{*\prime}\end{pmatrix}\left(1+\lambda_F^{*\prime}Q_{\bar F\bar F}^{-1}\lambda_F^*\right)^{-\frac{1}{2}}$:

$$D(\lambda_F^*)=-\Omega^{\frac{1}{2}}\mathcal{U}_1 S_1\mathcal{V}_{21}'Q_{\bar F\bar F}^{-\frac{1}{2}},\ \lambda_F^*=Q_{\bar F\bar F}^{\frac{1}{2}}\mathcal{V}_{21}'^{-1}\mathcal{V}_{11}',\ \delta=(\mathcal{U}_{22}\mathcal{U}_{22}')^{-\frac{1}{2}}\mathcal{U}_{22}S_2\mathcal{V}_{12}'(\mathcal{V}_{12}\mathcal{V}_{12}')^{-\frac{1}{2}}. \tag{51}$$

Proof. See the Online Appendix and also Kleibergen and Paap (2006). ■

The squared singular values are the roots of the characteristic polynomial in (16) so λ_F^* in Theorem 8 is the pseudo-true value of the risk premia. The population moment $\mu_f(\lambda_F)$ results from post-multiplying $\Omega^{-\frac{1}{2}}\begin{pmatrix}\mu_R & \beta\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & Q_{\bar F\bar F}^{\frac{1}{2}}\end{pmatrix}$ by $\begin{pmatrix}1\\ -Q_{\bar F\bar F}^{-\frac{1}{2}}\lambda_F\end{pmatrix}$, which is spanned by $\begin{pmatrix}1 & 0\\ 0 & Q_{\bar F\bar F}^{-\frac{1}{2}}\end{pmatrix}\begin{pmatrix}\lambda_F^* & I_m\end{pmatrix}_\perp'$, and pre-multiplying by $\Omega^{\frac{1}{2}}$. The derivative of the population objective function at λ_F then results as:

$$\begin{aligned}\sqrt{T}\mu_f(\lambda_F)'\Omega^{-1}D(\lambda_F)=&-\begin{pmatrix}1\\ -\lambda_F\end{pmatrix}'\begin{pmatrix}\lambda_F^* & I_m\end{pmatrix}'D(\lambda_F^*)'\Omega^{-1}D(\lambda_F)+\\ &\begin{pmatrix}1\\ -\lambda_F\end{pmatrix}'\begin{pmatrix}1 & 0\\ 0 & Q_{\bar F\bar F}^{-1}\end{pmatrix}\begin{pmatrix}\lambda_F^* & I_m\end{pmatrix}_\perp'\delta' D(\lambda_F^*)_\perp' D(\lambda_F),\end{aligned} \tag{52}$$

which equals zero when λ_F is the pseudo-true value but also at the other stationary points. When there is no misspecification, $\delta=0$ and $D(\lambda_F^*)=-\ddot\beta$ so

$$\sqrt{T}\mu_f(\lambda_F)'\Omega^{-1}D(\lambda_F)=\begin{pmatrix}1\\ -\lambda_F\end{pmatrix}'\begin{pmatrix}\lambda_F^* & I_m\end{pmatrix}'\ddot\beta'\Omega^{-1}D(\lambda_F)=(\lambda_F^*-\lambda_F)'\,\ddot\beta'\Omega^{-1}D(\lambda_F), \tag{53}$$

and $\ddot{\beta}$ is the only nuisance parameter.

Andrews et al. (2006) construct the two sided power envelope for testing the single structural parameter in a linear instrumental variables regression model with independent normal errors and a known value of the covariance matrix. This power envelope directly extends to the linear one factor asset pricing model with independent normal errors and no misspecification. It is then of interest to analyze if such a power envelope can be constructed in case of misspecification. Andrews et al. (2006) construct the power envelope using the maximal invariant which is stated in Theorem 9 alongside its distribution for the one factor linear asset pricing model with independent normal errors and known covariance matrices of the errors and factors.

Theorem 9: The maximal invariant, $S = \begin{pmatrix} S_{\perp\perp} & S'_{\lambda_F^1\perp} \\ S_{\lambda_F^1\perp} & S_{\lambda_F^1\lambda_F^1} \end{pmatrix}$, for testing $\mathrm{H}_0 : \lambda_F = \lambda_F^1$ in the one factor linear asset pricing model with independent normal errors and known values of the covariance matrices of the errors, Ω, and factors, $Q_{\bar{F}\bar{F}}$, is the quadratic form of:

$$\sqrt{T}\Omega^{-\frac{1}{2}} \begin{pmatrix} \bar{R} & \hat{\beta} \end{pmatrix} \left(\begin{pmatrix} 1 \\ -\lambda_F^1 \end{pmatrix} (1 + \lambda_F^{1\prime} Q_{\bar{F}\bar{F}}^{-1} \lambda_F^1)^{-\frac{1}{2}} \ \vdots \ \begin{pmatrix} 1 & 0 \\ 0 & Q_{\bar{F}\bar{F}} \end{pmatrix} \begin{pmatrix} \lambda_F^1 & I_m \end{pmatrix}' (Q_{\bar{F}\bar{F}} + \lambda_F^1 \lambda_F^{1\prime})^{-\frac{1}{2}} \right). \tag{54}$$

When $m = 1$, it has a non-central Wishart distribution with T degrees of freedom, identity scale matrix and non-centrality parameter:

Correct specification:

$$\begin{pmatrix} (\lambda_F^* - \lambda_F^1)(1 + (\lambda_F^1)^2 Q_{\bar{F}\bar{F}}^{-1})^{-\frac{1}{2}} \\ (Q_{\bar{F}\bar{F}} + (\lambda_F^1)^2)^{-\frac{1}{2}} \left(Q_{\bar{F}\bar{F}} + \lambda_F^* \lambda_F^{1\prime}\right) \end{pmatrix} \ddot{\beta}'\Omega^{-1}\ddot{\beta} \begin{pmatrix} (\lambda_F^* - \lambda_F^1)(1 + (\lambda_F^1)^2 Q_{\bar{F}\bar{F}}^{-1})^{-\frac{1}{2}} \\ (Q_{\bar{F}\bar{F}} + (\lambda_F^1)^2)^{-\frac{1}{2}} \left(Q_{\bar{F}\bar{F}} + \lambda_F^* \lambda_F^{1\prime}\right) \end{pmatrix}' \tag{55}$$

Misspecification:

$$\begin{pmatrix} (\lambda_F^* - \lambda_F^1)(1 + (\lambda_F^1)^2 Q_{\bar{F}\bar{F}}^{-1})^{-\frac{1}{2}} \\ (Q_{\bar{F}\bar{F}} + (\lambda_F^1)^2)^{-\frac{1}{2}} \left(Q_{\bar{F}\bar{F}} + \lambda_F^* \lambda_F^{1\prime}\right) \end{pmatrix} D(\lambda_F^*)'\Omega^{-1}D(\lambda_F^*) \begin{pmatrix} (\lambda_F^* - \lambda_F^1)(1 + (\lambda_F^1)^2 Q_{\bar{F}\bar{F}}^{-1})^{-\frac{1}{2}} \\ (Q_{\bar{F}\bar{F}} + (\lambda_F^1)^2)^{-\frac{1}{2}} \left(Q_{\bar{F}\bar{F}} + \lambda_F^* \lambda_F^{1\prime}\right) \end{pmatrix}' +$$
$$\begin{pmatrix} (1 + (\lambda_F^1)^2 Q_{\bar{F}\bar{F}}^{-1})^{-\frac{1}{2}} \left(1 + \lambda_F^* Q_{\bar{F}\bar{F}}^{-1} \lambda_F^1\right) \\ -(Q_{\bar{F}\bar{F}} + (\lambda_F^1)^2)^{-\frac{1}{2}} \left(\lambda_F^* - \lambda_F^1\right) \end{pmatrix} (1 + (\lambda_F^*)^2 Q_{\bar{F}\bar{F}}^{-1})^{-1} \delta'\delta \begin{pmatrix} (1 + (\lambda_F^1)^2 Q_{\bar{F}\bar{F}}^{-1})^{-\frac{1}{2}} \left(1 + \lambda_F^* Q_{\bar{F}\bar{F}}^{-1} \lambda_F^1\right) \\ -(Q_{\bar{F}\bar{F}} + (\lambda_F^1)^2)^{-\frac{1}{2}} \left(\lambda_F^* - \lambda_F^1\right) \end{pmatrix}', \tag{56}$$

where the specifications of $D(\lambda_F^*)$ and δ are stated in Theorem 8.

Proof. See the Online Appendix. ■

The elements of the maximal invariant in Theorem 9 are such that:

$$\begin{aligned} S_{\lambda_F^1\lambda_F^1} = &\ T\hat{D}(\lambda_F^1)'\hat{V}_{\theta\theta.f}(\lambda_F^1)^{-1}\hat{D}(\lambda_F^1) \\ S_{\perp\perp} = &\ Tf_T(\lambda_F^1, X)'\hat{V}_{ff}(\lambda_F^1)^{-1}f_T(\lambda_F, X) \\ S_{\lambda_F^1\perp} = &\ T\left(\hat{V}_{ff}(\lambda_F^1)^{-\frac{1}{2}}f_T(\lambda_F, X)\right)'\left(\hat{V}_{\theta\theta.f}(\lambda_F^1)^{-\frac{1}{2}}\hat{D}(\lambda_F^1)\right). \end{aligned} \tag{57}$$

Since $1+(\lambda_F^1)^2Q_{\bar{F}\bar{F}}^{-1}$ is known, the distribution of the maximal invariant in Theorem 9 is a function of three unknown parameters: $D(\lambda_F^*)'\Omega^{-1}D(\lambda_F^*)$, $\delta'\delta$ and $\left(\lambda_F^*-\lambda_F^1\right)$. Under $\mathrm{H}_0 : \lambda_F = \lambda_F^1 = \lambda_F^*$, $\lambda_F^*-\lambda_F^1 = 0$ so one of these three parameters is pinned down.

Corollary 3. Under $\mathrm{H}_0 : \lambda_F = \lambda_F^*$, the non-centrality parameter of the non-central Wishart distribution of the maximal invariant equals:

$$\begin{aligned} &\text{Correct specification: } \begin{pmatrix} 0 \\ 1 \end{pmatrix}(Q_{\bar{F}\bar{F}}+(\lambda_F^*)^2)\ddot{\beta}'\Omega^{-1}\ddot{\beta}\begin{pmatrix} 0 \\ 1 \end{pmatrix}' \\ &\text{Misspecification: } \begin{pmatrix} 0 \\ 1 \end{pmatrix}(Q_{\bar{F}\bar{F}}+(\lambda_F^*)^2)D(\lambda_F^*)'\Omega^{-1}D(\lambda_F^*)\begin{pmatrix} 0 \\ 1 \end{pmatrix}' + \begin{pmatrix} 1 \\ 0 \end{pmatrix}\delta'\delta\begin{pmatrix} 1 \\ 0 \end{pmatrix}'. \end{aligned} \tag{58}$$

Corollary 3 shows that under H_0 and correct specification, the three different elements of the maximal invariant depend on only one unknown parameter, $(Q_{\bar{F}\bar{F}}+(\lambda_F^*)^2)\ddot{\beta}'\Omega^{-1}\ddot{\beta}$. Since the $S_{\lambda_F^1\lambda_F^1}$-element of the maximal invariant is a sufficient statistic for it and independently distributed of the other elements of the maximal invariant, we can condition on $S_{\lambda_F^1\lambda_F^1}$ to construct the power envelope and for optimally combining the two other elements of the maximal invariant, $S_{\lambda_F^1\perp}$ and $S_{\perp\perp}$, to improve the power of testing H_0, see Andrews et al. (2006).

Under misspecification, the three elements of the maximal invariant depend on two parameters, $(Q_{\bar{F}\bar{F}}+(\lambda_F^*)^2)D(\lambda_F^*)'\Omega^{-1}D(\lambda_F^*)$ and $\delta'\delta$. These are estimated using $S_{\lambda_F^1\lambda_F^1}$ and $S_{\perp\perp}$ so we can no longer use $S_{\perp\perp}$ to improve the power of tests of H_0 like in case of correct specification. The $S_{\lambda_F^1\perp}$-element of the maximal invariant, which represents the score, is then the only element which can be used to test H_0 under misspecification. It is thus not obvious how to improve the power of tests of $\mathrm{H}_0 : \lambda_F = \lambda_F^*$ compared to the score test in case of misspecification.

The non-centrality parameter of the score element of the distribution of the maximal invariant,

$S_{\lambda_F^1 \perp}$:

$$\begin{array}{c}\left(\lambda_F^* - \lambda_F^1\right)(Q_{\bar{F}\bar{F}} + (\lambda_F^1)^2)^{-\frac{1}{2}}(1 + (\lambda_F^1)^2 Q_{\bar{F}\bar{F}}^{-1})^{-\frac{1}{2}} \\ \left[\left(Q_{\bar{F}\bar{F}} + \lambda_F^* \lambda_F^{1\prime}\right) D(\lambda_F^*)'\Omega^{-1} D(\lambda_F^*) - (1 + (\lambda_F^*)^2 Q_{\bar{F}\bar{F}}^{-1})^{-1} \delta'\delta \left(1 + \lambda_F^* Q_{\bar{F}\bar{F}}^{-1} \lambda_F^1\right)\right]\end{array} \tag{59}$$

shows, similar to Theorem 7, that the power of the DRLM test positively depends on the strength of identification, $D(\lambda_F^*)'\Omega^{-1}D(\lambda_F^*)$, and negatively on the misspecification, $\delta'\delta$. It further shows that under the null $\mathrm{H}_0 : \lambda_F = \lambda_F^1 = \lambda_F^*$, the non-centrality parameter equals zero when $\delta'\delta = \left(Q_{\bar{F}\bar{F}} + \lambda_F^{*2}\right) D(\lambda_F^*)'\Omega^{-1}D(\lambda_F^*)$ so λ_F is not identified when the identification strength equals the misspecification, see also (48).

5 Testing multiple and subsets of the structural parameter vector

The expressions of the DRLM test apply as well to settings where the structural parameter vector has multiple elements. The power enhancement procedure directly extends as well. Hence, we can improve the power of testing a hypothesis on the structural parameter vector at the $\alpha \times 100\%$ significance level by also rejecting it when there are significant values of the statistic on every line going from the hypothesized parameter value to the CUE.

Many times, we are interested in constructing confidence sets on the individual elements of the structural parameter vector. Subset DRLM tests of hypotheses specified on a selection of the elements of the structural parameter vector which result from substituting the CUE for the parameters left unspecified under the hypothesis of interest, are not necessarily size correct, see also Guggenberger et al. (2012). Confidence sets with the correct coverage therefore result by projecting the joint confidence set that applies to all structural parameters on the different axes.

6 Nonlinear GMM

The DRLM test applies to general non-linear GMM settings with unrestricted covariance matrices. We therefore conduct a small simulation study using the non-linear moment equation resulting from a constant relative rate of risk aversion (CRRA) utility function, see e.g. Hansen and Singleton (1982), to illustrate the size and power properties of the DRLM test in a non-linear GMM setting.

Running example 3: Constant relative risk aversion (CRRA) The moment function resulting from the CRRA utility function is, see e.g. Hansen and Singleton (1982):

$$E\left[\delta\left(\tfrac{C_{t+1}}{C_t}\right)^{-\gamma}(\iota_N+R_{t+1})-\iota_N\right]=\mu_f(\delta,\gamma), \tag{60}$$

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 and R_{t+1} an N-dimensional vector of asset returns. The sample moment function and its derivative therefore only depend on $\gamma:$

$$\begin{array}{llll} f_T(\gamma,X)= & \frac{1}{T}\sum_{t=1}^T f_t(\gamma), & f_t(\gamma)= & \delta_0\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(\iota_N+R_{t+1})-\iota_N, \\ q_T(\gamma,X)= & \frac{1}{T}\sum_{t=1}^T q_t(\gamma), & q_t(\gamma)= & -\delta_0\ln\left(\frac{C_{t+1}}{C_t}\right)\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(\iota_N+R_{t+1}). \end{array} \tag{61}$$

The covariance matrix estimators are the Eicker-White ones, see White (1980):

$$\begin{array}{rl} \hat{V}_{ff}(\gamma)= & \frac{1}{T}\sum_{t=1}^T(f_t(\gamma)-f_T(\gamma,X))(f_t(\gamma)-f_T(\gamma,X))', \\ \hat{V}_{\theta f}(\gamma)= & \frac{1}{T}\sum_{t=1}^T(q_t(\gamma)-q_T(\gamma,X))(f_t(\gamma)-f_T(\gamma,X))', \\ \hat{V}_{\theta\theta}(\gamma)= & \frac{1}{T}\sum_{t=1}^T(q_t(\gamma)-q_T(\gamma,X))(q_t(\gamma)-q_T(\gamma,X))', \\ \hat{V}_{\theta\theta.f}(\gamma)= & \hat{V}_{\theta\theta}(\gamma)-\hat{V}_{\theta f}(\gamma)\hat{V}_{ff}(\gamma)^{-1}\hat{V}_{\theta f}(\gamma)'. \end{array} \tag{62}$$

We use a log-normal data generating process to jointly simulate consumption growth and asset returns in accordance with the moment equation. Since the discount factor is fixed at its true value, γ is the single structural parameter of interest; see, for example, Savov (2011) and Kroencke (2017). The population moment function then reads:[4]

$$\mu_f(\gamma)=\begin{pmatrix} \exp\left(\ln(\delta_0)+\mu_{2,1,0}+\frac{1}{2}\left(V_{rr,11,0}+\gamma^2V_{cc,0}-2\gamma V_{rc,1,0}\right)\right) \\ \vdots \\ \exp\left(\ln(\delta_0)+\mu_{2,N,0}+\frac{1}{2}\left(V_{rr,NN,0}+\gamma^2V_{cc,0}-2\gamma V_{rc,N,0}\right)\right) \end{pmatrix}-\iota_N, \tag{63}$$

with $\mu_{2,0}=(\mu_{2,1,0}\ldots\mu_{2,N,0})'$ the mean of $r_{t+1}=\ln(1+R_{t+1})$, $V_{cc,0}$ the (scalar) variance of $\triangle c_{t+1}=\ln\left(\frac{C_{t+1}}{C_t}\right)$, $V_{rc,0}=V'_{cr,0}=(V_{rc,1,0}\ldots V_{rc,N,0})'$ the $N\times 1$ dimensional covariance between r_{t+1} and $\triangle c_{t+1}$ and $V_{rr,0}=(V_{rr,ij,0}):i,j=1,\ldots,N$, the $N\times N$ dimensional covariance matrix of r_{t+1}. The Online Appendix states the expression of the population covariance matrix $V_{ff}(\gamma)$ needed to compute the pseudo-true value γ^*:

[4]See the Online Appendix for its construction and for further details on the simulation setup.

$$\gamma^* = arg\min_\gamma \mu_f(\gamma)' V_{ff}(\gamma)^{-1} \mu_f(\gamma). \tag{64}$$

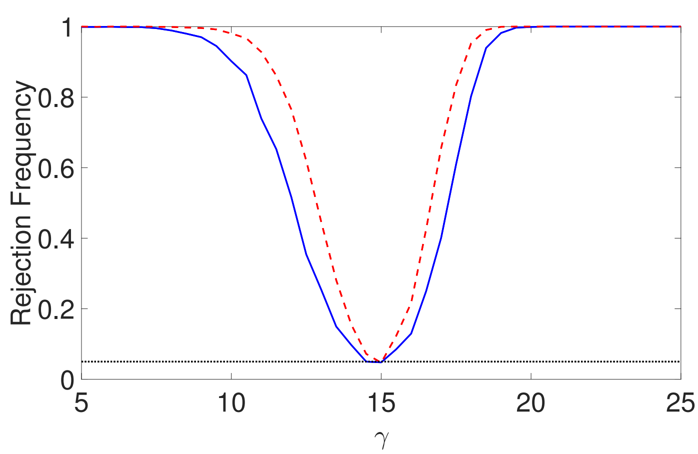
Unlike for the linear factor asset pricing model, we need to compute the pseudo-true value numerically since no closed form expression is available when there is misspecification. This also explains why we use the log-normal setting so we have an analytical expression of the population moment function and only use one structural parameter since numerical optimizing in higher dimensions is both computationally demanding and can be imprecise. We analyze GMM-AR and DRLM tests for correctly specified and misspecified settings.

Correct Specification and $N = 5$ Standard GMM operates under correct specification so (63) equals zero, which implies that:

$$\mu_{2,0} = -\iota_N \ln(\delta_0) - \tfrac{1}{2} \left[\begin{pmatrix} V_{rr,11,0} \\ \vdots \\ V_{rr,NN,0} \end{pmatrix} + \iota_N \gamma^2 V_{cc,0} - 2\gamma V_{rc,0} \right]. \tag{65}$$

We revisit the simulation study in Kleibergen and Zhan (2020), who examine the GMM-AR test on γ. We augment their simulation study by the DRLM test. Figure 16 shows the resulting power curves of the GMM-AR and DRLM tests. It indicates that GMM-AR and DRLM are both size-correct with good power in the correctly specified setting.

Figure 16: Simulated power curves of GMM-AR (solid blue) and DRLM (dashed red) tests with 5% significance under correct specification. The CRRA moment condition is imposed in the DGP with $\delta = 0.95$ and $N = 5$. The null hypothesis is $\mathrm{H}_0 : \gamma = 15$.



In addition, since we consider $N = 5$ in the DGP, there is over-identification, which helps explain the difference in power between the GMM-AR and DRLM tests.

Misspecification and $N = 5$ For misspecification, we no longer impose (65) in the DGP. Instead, we just test for the pseudo-true value of γ, denoted by γ^*. Specifically, we start with an auxiliary $\tilde{\mu}_2$ that satisfies (65), and then subtract a vector of constants to introduce misspecification in the DGP:

$$\begin{aligned}\tilde{\mu}_2 = &\quad -\iota_N \ln(\delta_0) - \tfrac{1}{2}\left[\begin{pmatrix} V_{rr,11,0} \\ \vdots \\ V_{rr,NN,0} \end{pmatrix} + \iota_N \gamma^2 V_{cc,0} - 2\gamma V_{rc,0}\right] \\ \mu_{2,0} = &\quad \tilde{\mu}_2 - c\iota_N. \end{aligned} \tag{66}$$

Figure 17.1 in Panel 17 illustrates our simulation design. When $c = 0$, $\gamma^* = 15$, and min $\mu_f' V_{ff}^{-1} \mu_f$ $= 0$, as in the previous correct specification case. When c deviates from zero, the pseudo-true value γ^* starts to differ from 15, and the objective function $\mu_f' V_{ff}^{-1} \mu_f$ in Figure 17.2 is no longer equal to zero at the pseudo-true value γ^*.

Panel 17: Pseudo-true value and population objective function as functions of the misspecification

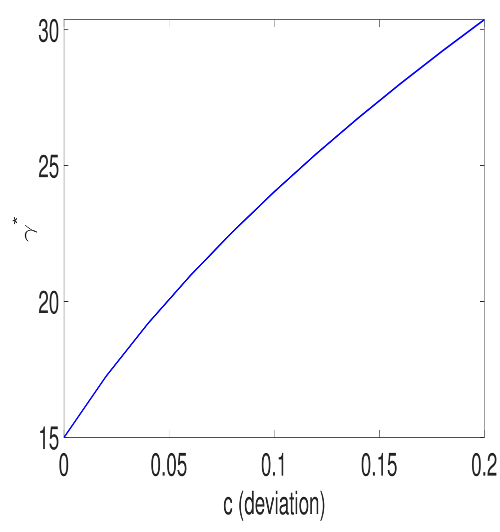


Figure 17.1: Pseudo-true value function

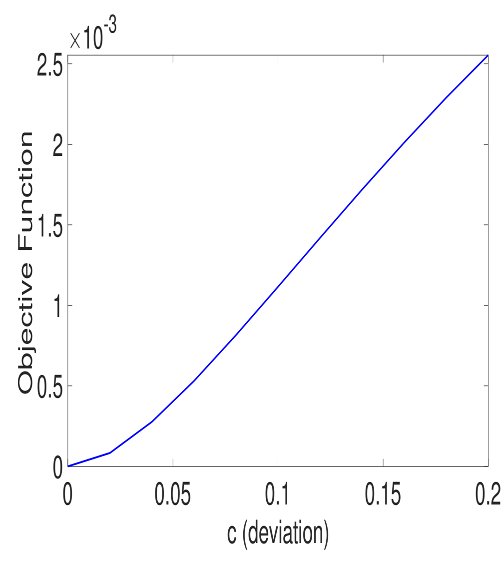
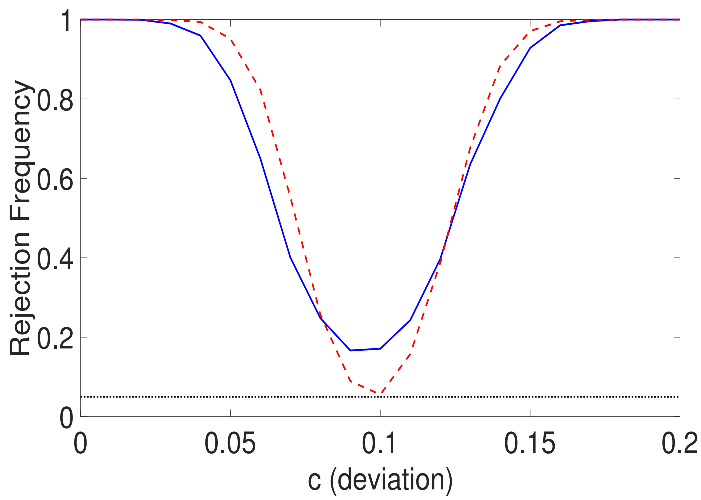


Figure 17.2: Population objective function at γ^*

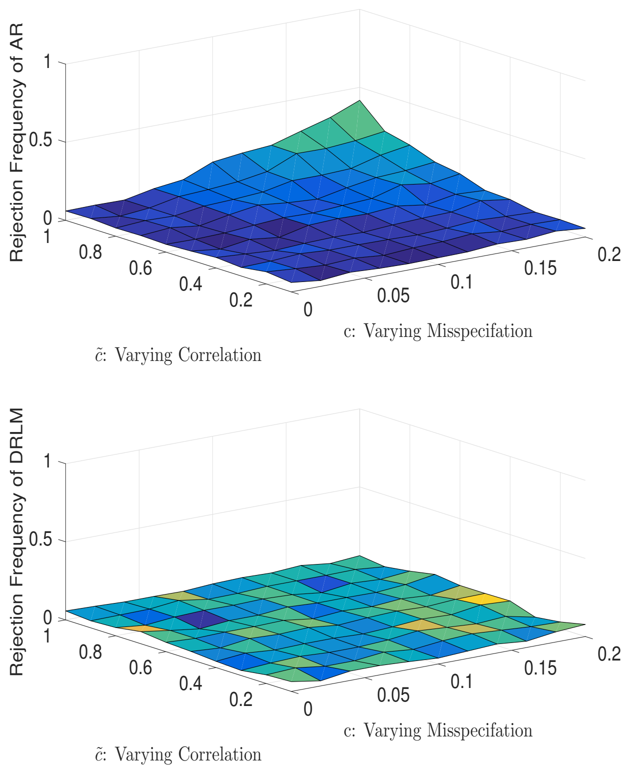
Figure 18 shows the rejection frequencies of GMM-AR and DRLM tests of $\mathrm{H}_0 : \gamma^* = 24$ which corresponds, according to Figure 17.1, with a degree of misspecification of 0.1. We consider a range of values of c from 0 to 0.2 in the DGP while we test for $\text{H}_0 : \gamma^* = 24$, or put differently, $\text{H}_0 : c = 0.1$. Figure 18 shows that the GMM-AR test rejects the null more often than the nominal significance level of 5% to reflect that the moment condition is misspecified. In contrast, since the DRLM test allows for misspecification, it has the correct rejection frequency at the hypothesized value.

Figure 18: Simulated power curves of GMM-AR (solid blue) and DRLM (dashed red) tests at the 5% significance level under misspecification. The null hypothesis $\text{H}_0 : \gamma = \gamma^* = 24$ corresponds with misspecification equal to $c = 0.1$ where c reflects the deviation for misspecification.



Size of AR and DRLM tests with $N = 5$ Furthermore, Figure 19 shows the trade-off between the identification strength and the misspecification for the rejection frequencies of GMM-AR and DRLM tests. The DGP is such that the correlation coefficient between the log-consumption growth and the log asset returns, $\rho_i = \frac{V_{rc,i,0}}{\sqrt{V_{cc,0} V_{rr,ii,0}}}$, is scaled by a constant $\tilde{c}$ to vary identification. Figure 19 shows the rejection frequencies of tests of $\text{H}_0 : \gamma = \gamma^*$ as a function of the misspecification c and strength of identification which is (partly) reflected by $\tilde{c}$. We note that the pseudo-true value γ^* is a function of $(c, \tilde{c})$ so the reported rejection frequencies in Figure 19 are for different hypothesized values of γ^*. Figure 19 shows that the GMM-AR test gets size distorted when the misspecification increases. This is unlike the DRLM test which remains size correct for all values of the identification and misspecification strengths.

Figure 19: Rejection frequencies of GMM-AR and DRLM tests of $\mathrm{H}_0 : \gamma = \gamma^*$ at the 5% significance level with $N = 5$ as a function of the strengths of identification, $\tilde{c}$, and misspecification c.



7 Applications

We apply the DRLM test and the identification robust AR, KLM and LR tests to data for two different models discussed previously: the linear asset pricing model and the linear instrumental variables regression model.

Running example 1: Linear asset pricing model We briefly revisit the linear factor models considered in Adrian et al. (2014) and He et al. (2017) using our DRLM test and the identification robust AR, KLM and LR tests, see Kleibergen (2009) and Kleibergen and Zhan (2020).

Adrian et al. (2014) propose a leverage risk factor ("$LevFac$"), where the leverage level is the ratio of total assets over the difference between total assets and total liabilities. The resulting log change of the leverage level is their leverage factor. The empirical study of Adrian et al. (2014)

uses quarterly data between 1968Q1 and 2009Q4. Following Lettau et al. (2019), we extend the time period to 1963Q3 - 2013Q4 and use $N = 25$ size and book-to-market portfolios as test assets. Adrian et al. (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 et al. (2013).

He et al. (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 two factors. As presented in Table 1, the significant FM and KRS t-statistics for the risk premium on "$EqFac$" appear to favor this factor for asset pricing.

DRLM: Adrian, Etula, and Muir (2014) Using the same data as for Table 1, Figure 20 shows the p-values for testing the risk premium on the leverage factor (horizontal line) using the DRLM, AR, KLM, and LR tests. Most of the p-values in Figure 20 are above the 5% level, which implies that none of the DRLM, 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 weak identification of the risk premium on the leverage factor reflected by the unbounded 95% confidence sets, it is likely that there is misspecification so it would be appropriate to use the DRLM test.

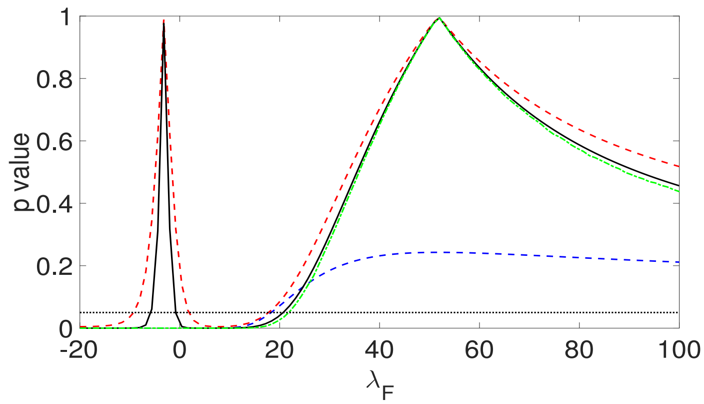
The p-values of the DRLM test in Figure 20 are equal to one at two different points. The p-values of the AR test show that one of these two points relates to the minimal value of the AR test and the other one to the maximal value of the AR test. Using the power enhancement rule for the DRLM test, we can reject non-significant values that lie within the closed interval indicated by the maximizers of the DRLM statistic that does not contain the CUE so the non-significant p-values of the DRLM test which occur around the maximizer of the 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 weak identification of the risk premium of the leverage factor and the likely misspecification reflected by the smallish p-value of the J-test.

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).

	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	2.67	0.78	2.16
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	(3.71, 24.11)	(-1.80, 4.18)	(0.65, 13.11)
DRLM	$(-\infty, -91.4) \cup (-9.2, 1.6) \cup (17.8, +\infty)$	$(-\infty, +\infty)$	$(-\infty, +\infty)$
DRLM (power enh.)	$(-\infty, -91.4) \cup (17.8, +\infty)$	$(-\infty, +\infty)$	$(-\infty, +\infty)$
AR	$(-\infty, -101.4) \cup (18.4, +\infty)$	$(-\infty, -64.6) \cup (8.1, +\infty)$	$(-\infty, -244.1) \cup (37.7, +\infty)$
KLM	$(-\infty, -185.8) \cup (-5.6, -1.0) \cup (20.8, +\infty)$	$(-\infty, -7.2) \cup (-4.7, -0.3) \cup (1.0, +\infty)$	$(-\infty, -23.8) \cup (-8.1, 1.7) \cup (11.8, +\infty)$
LR	$(-\infty, -276.2) \cup (22.0, +\infty)$	$(-\infty, -9.7) \cup (2.2, +\infty)$	$(-\infty, -33.7) \cup (16.2, +\infty)$

Figure 20: Adrian, Etula and Muir (2014). p-value from the DRLM (dashed red), AR (dashed blue), KLM (solid black), LR (dash-dotted green) and the 5% level (dotted black). J-statistic (=minimum AR) equals 28.42, with p-value of 0.20 resulting from $\chi^2(N-2)$.



DRLM: He, Kelly, and Manela (2017) Panel 21 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, AR, KLM, and LR tests. The p-value of the J-test shows that misspecification is 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 95% confidence sets of the DRLM and KLM tests have two rather disjoint areas. The power enhancement rule for the DRLM test shows that the smaller disjoint area 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 "$EqFac$" which indicates that the pricing ability of "$EqFac$" is under doubt.

To compare with Panel 21, we replace the "$EqFac$" risk factor with the "SMB" (small minus big) factor from Fama and French (1993) and similarly construct Panel 22. The AR test now signals model misspecification, since it rejects every hypothesized risk premia as shown in Figure 22.2 so the 95% confidence set that results from the AR test is empty. Our DRLM test, which allows for misspecification, yields a tight confidence set in Figure 22.1. This tight confidence set, in contrast with the wide one in Figure 21.1, indicates that the pricing ability of "$EqFac$" differs substantially from "SMB". 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.

Panel 21: He, Kelly and Manela (2017). 95% confidence sets from DRLM, AR, KLM and LR.

J-statistic (minimum of AR) equals 35.32, with p-value of 0.036 resulting from $\chi^2(N-3)$.

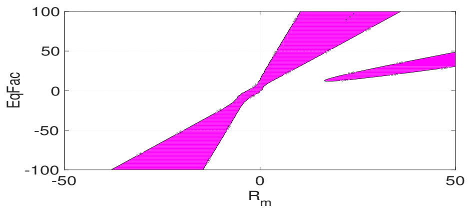


Figure 21.1: DRLM

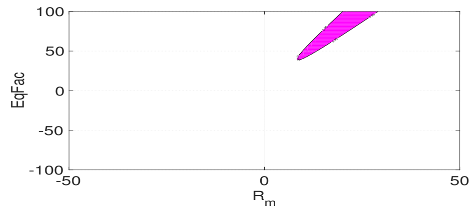


Figure 21.2: AR

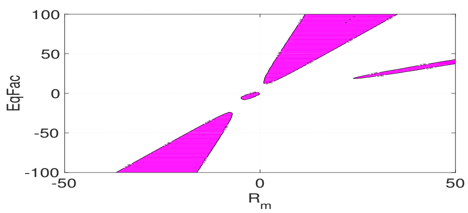


Figure 21.3: KLM

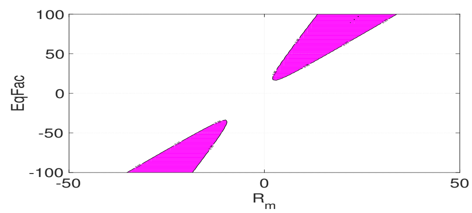


Figure: 21.4: LR

Panel 22: R_m and SMB. 95% confidence sets from DRLM, AR, KLM and LR.

J-statistic (minimum of AR) equals 59.34, with p-value of 0.00 resulting from $\chi^2(N-3)$.

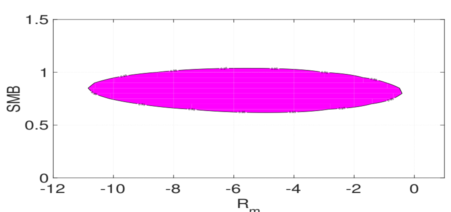


Figure 22.1: DRLM

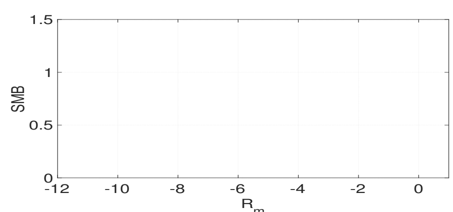


Figure 22.2: AR

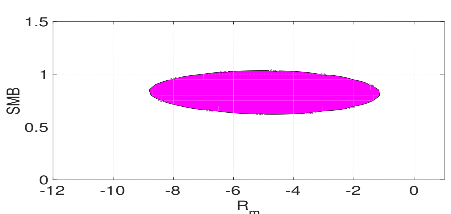


Figure 22.3: KLM

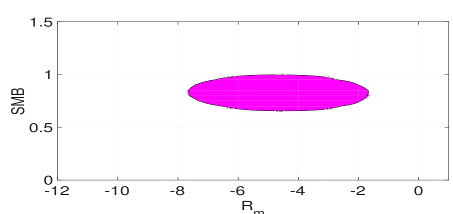


Figure: 22.4: LR

Running example 2: Linear instrumental variables regression for the return on education using Card (1995) data To further show the ease of implementing the DRLM test for applied work, we use the return on education data from Card (1995). Card (1995) uses proximity to college as the instrument in an IV regression of (the log) wage on (length of) education. For more details on the data, we refer to Card (1995). The instruments used in our specification are three binary indicator variables which show the proximity to a two-year college, a four-year college and a four-year public college, respectively. The included exogenous variables are a constant term, age, age^2, and racial, metropolitan, family and regional indicator variables. All three binary instruments have their own local average treatment effects, which in case of heterogeneous treatment effects leads to misspecification of the IV regression model since it considers them to be identical, see Imbens and Angrist (1994).

Figure 23: Tests of the return on education using Card (1995) data with the DRLM (solid black), KLM (dashed black), LR (solid red) and AR (solid blue) tests and their 95% (conditional) critical value lines (dotted in the color of the test they refer to).

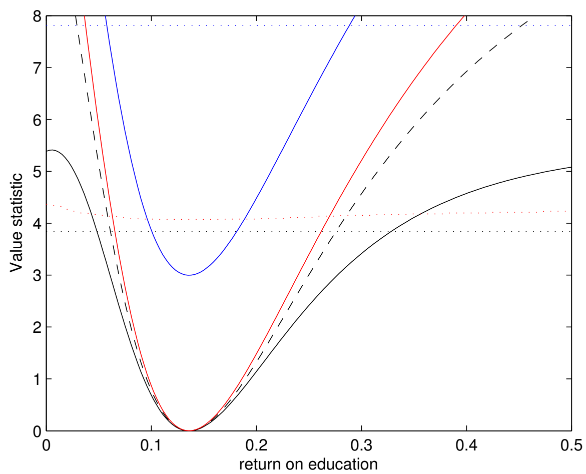


Figure 23 shows the values of the AR, LR, KLM and DRLM tests around the CUE. It also shows their 5% critical value functions. The other area of small values of the DRLM test is left out since it would be discarded by the power enhancement rule. The J-test, which equals the minimal value of the AR statistic, is 2.99 with a p-value of 0.22. Since the return on education is not strongly identified, the J-test does not have much power. Its quite low p-value can indicate misspecification which then results from distinct local average treatment effects of the different instruments. Lee (2018) constructs misspecification-robust standard errors for the two stage least squares estimator when the local average treatment effects differ but the resulting t-test is not valid here because of the weak identification of the return on education. This makes the DRLM test more appealing since it is robust to both misspecification and weak identification. Kitagawa (2015) further shows that the validity of the instruments for the Card data depends on the specification of the model. Figure 23 then shows that allowing for misspecification further enlarges the identification-robust confidence set for the return on education.

8 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 using the DRLM test. While settings of weak identification paired with misspecification are empirically relevant, it was so far not possible to conduct reliable inference in these settings. This holds since weak identification robust tests are size distorted when the model is misspecified while the misspecification tests which are typically used to detect misspecification, are virtually powerless under weak identification. Hence, it is not possible to test for the settings where weak identification robust tests falter, in a powerful manner. We propose some straightforward power improvements for the DRLM test which make it work well. We hope to conduct further power improvements in future work. We also used the DRLM test to analyze data from three studies which are plagued by both weak identification and misspecification issues: Card (1995), Adrian et al. (2014), and He et al. (2017). It shows that other inference procedures can seriously underestimate the uncertainty concerning the structural parameters when both misspecification and weak identification matter.

References

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

[2] Anderson, T.W. and H. Rubin. Estimation of the Parameters of a Single Equation in a Complete Set of Stochastic Equations. *The Annals of Mathematical Statistics*, **21**:570–582, 1949.

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

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

[5] Andrews, I. Conditional Linear Combination Tests for Weakly Identified Models. *Econometrica*, **84**:2155–2182, 2016.

[6] Andrews, I. and A. Mikusheva. A Geometric Approach to Nonlinear Econometric Models. *Econometrica*, **84**:1249–1264, 2016a.

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

[8] Andrews, I., J. Stock and L. Sun. Weak instruments in IV Regression: Theory and Practice. *Annual Review of Economics*, **11**:727–753, 2019.

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

[10] Fama, E.F. and J.D. MacBeth. Risk, Return and Equillibrium: Empirical Tests. *Journal of Political Economy*, **81**:607–636, 1973.

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

[12] Gospodinov, N., R. Kan and C. Robotti. Spurious inference in Reduced-Rank Regresson Models. *Econometrica*, **85**:1613–1628, 2017.

[13] Guggenberger, P. , F. Kleibergen, S. Mavroeidis, and L. Chen. On the asymptotic sizes of subset anderson-rubin and lagrange multiplier tests in linear instrumental variables regression. *Econometrica*, **80**:2649–2666, 2012.

[14] Hall, A. and A. Inoue. The large sample behaviour of the generalized method of moments estimator in misspecified models. *Journal of Econometrics*, 114(2):361–394, 2003.

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

[16] Hansen, L.P. and K. Singleton. Generalized Instrumental Variable Estimation of Nonlinear Rational Expectations Models. *Econometrica*, **50**:1269–1286, 1982.

[17] Hansen, L.P., J. Heaton and A. Yaron. Finite Sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics*, **14**:262–280, 1996.

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

[19] Imbens G. and J. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994.

[20] Kan, R. and C. Zhang. Two-Pass Tests of Asset Pricing Models with Useless Factors. *Journal of Finance*, **54**:203–235, 1999.

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

[22] Kitagawa, T. A Test for Instrument Validity. *Econometrica*, **83**:2043–2063, 2015.

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

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

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

[26] Kleibergen, F. Tests of Risk Premia in Linear Factor Models. *Journal of Econometrics*, **149**:149–173, 2009.

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

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

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

[30] Kolesár, M., R. Chetty, J. Friedman, E. Glaeser and G.W. Imbens. Identification and Inference With Many Invalid Instruments. *Journal of Business and Economic Statistics*, **33**:474–484, 2015.

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

[32] Lee, S. A consistent variance estimator for 2SLS when instruments identify different LATEs. *Journal of Business & Economic Statistics*, 36(3):400–410, 2018.

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

[34] Mavroeidis, S., M. Plagborg-Moller and J. Stock. Empirical Evidence on Inflation Expectations in the New-Keynesian Phillips Curve. *Journal of Economic Literature*, **52**:124–188, 2014.

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

[36] Sargan, J.D. The Estimation of Economic Relationships using Instrumental Variables. *Econometrica*, **26**:393–415, 1958.

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

[38] Shanken, J. On the Estimation of Beta-Pricing Models. *Review of Financial Studies*, **5**:1–33, 1992.

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

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

[41] White, H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica*, **48**:817–838, 1980.