Online Appendix for

"Double Robust Inference for Continuous Updating GMM"

FRANK KLEIBERGEN and ZHAOGUO ZHAN

October 15, 2024

Contents

I Lemmas and Proofs **2**

A. Lemmas 2

B. Proof of Theorem 1 18

C. Proof of Theorem 2 24

D. Proof of Theorem 3 25

E. Proof of Theorem 4 27

II Additional Results **28**

A. Simulation: Nonlinear GMM 28

B. Simulation: Linear IV 32

C. Application: Epstein-Zin 32

D. Application: Linear IV regression using Card (1995) data 33

E. DRLM using conditional critical values 35

F. Subvector inference 35

I Lemmas and Proofs

A. Lemmas

Lemma 1. The estimators $\bar{R}$ and $\hat{\beta}$ in the linear regression model:

$$R_t = c + \beta F_t + u_t,$$

with c an N-dimensional vector of constants, $F_t = G_t - \bar{G}$, with G_t an m-dimensional vector of factors and $\bar{G} = \frac{1}{T}\sum_{t=1}^{T} G_t$, so $\bar{F} = 0$, and u_t an N-dimensional vector which contains the errors which are i.i.d. distributed with mean zero and covariance matrix Ω, are independently distributed in large samples.

Proof: Since $\bar{R} = \hat{c} + \hat{\beta}\bar{F} = \hat{c}$, and the joint limit behavior of $\hat{c}$ and $\hat{\beta}$ accords with

$$\sqrt{T}\left[\begin{pmatrix} \hat{c} \\ \text{vec}(\hat{\beta}) \end{pmatrix} - \begin{pmatrix} c \\ \text{vec}(\beta) \end{pmatrix}\right] \underset{d}{\to} \begin{pmatrix} \psi_c \\ \psi_\beta \end{pmatrix},$$

with

$$\begin{pmatrix} \psi_c \\ \psi_\beta \end{pmatrix} \sim N\left(0, (Q^{-1} \otimes I_N)\Sigma(Q^{-1} \otimes I_N)\right),$$

since $\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} 1 \\ F_t \end{pmatrix}\begin{pmatrix} 1 \\ F_t \end{pmatrix}' \underset{p}{\to} Q = \begin{pmatrix} 1 & \mu_F' \\ \mu_F & Q_{FF} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & Q_{FF} \end{pmatrix}$, $\mu_F = 0$, $Q_{FF} = E(F_t F_t') = Q_{\bar{F}\bar{F}} + \mu_F \mu_F'$, and $\frac{1}{T}\sum_{t=1}^{T}\left(\begin{pmatrix} 1 \\ F_t \end{pmatrix}\begin{pmatrix} 1 \\ F_t \end{pmatrix}' \otimes u_t u_t'\right) \underset{p}{\to} \Sigma$. When u_t is i.i.d., $\Sigma = (Q \otimes \Omega)$, with $\Omega = var(u_t)$, so

$$\begin{pmatrix} \psi_c \\ \psi_\beta \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & 0 \\ 0 & Q_{FF}^{-1} \end{pmatrix} \otimes \Omega\right),$$

and the limit behaviors of $\bar{R} = \hat{c}$ and $\hat{\beta}$ are thus independent.

Lemma 2. **a.** When $\hat{V}_{ff}(\theta)^{-1} = \hat{V}_{ff}(\theta)^{-\frac{1}{2}\prime}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}$, $\theta : 1 \times 1$, $f_T(\theta, X)$ a linear function of θ, so $\frac{\partial}{\partial\theta}q_T(\theta, X) = 0$, it holds that

$$\frac{\partial}{\partial\theta}\hat{V}_{ff}(\theta)^{-\frac{1}{2}} = -\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}.$$

b. $\frac{\partial}{\partial\theta}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}f_T(\theta, X) = \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}(\theta)$.

c. $\frac{\partial}{\partial\theta}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}(\theta) = -2\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) - \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X)$.

d.

$$\begin{aligned}\frac{\partial}{\partial\theta}f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) = \;& \hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) - 2f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) - \\ & f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X).\end{aligned}$$

e.

$$\begin{aligned}\frac{\partial}{\partial\theta}\left(\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)\right) = \;& -4\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) - \\ & 2\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X).\end{aligned}$$

f.

$$\frac{\partial}{\partial\theta}\hat{V}_{\theta\theta}(\theta) = \; -\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)' - \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta).$$

g.

$$\begin{aligned}&\frac{\partial}{\partial\theta}\left(f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X)\right) \\ = \;& 2\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X) - 4f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X).\end{aligned}$$

h.

$$\begin{aligned}&\frac{\partial}{\partial\theta}\left(\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) + f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X)\right) \\ = \;& -4\left[\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) + f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X)\right].\end{aligned}$$

i. For θ^* satisfying the population FOC and $V(\theta) = \begin{pmatrix} V_{ff}(\theta) & V_{f\theta}(\theta) \\ V_{\theta f}(\theta) & V_{\theta\theta}(\theta)\end{pmatrix} = \left(\begin{pmatrix} 1 & 0 \\ \theta & 1\end{pmatrix}'\Sigma\begin{pmatrix} 1 & 0 \\ \theta & 1\end{pmatrix} \otimes \Omega\right)$, with Ω $k_f \times k_f$ and $\Sigma = I_2$, **d** above implies that

$$\begin{aligned}\tfrac{1}{2}\tfrac{\partial^2}{\partial\theta^2}Q_p(\theta)|_{\theta^*} = \left(\tfrac{1}{1+(\theta^*)^2}\right)^3 \times \Bigg[&\begin{pmatrix} -\theta^* \\ 1\end{pmatrix}'\left(\mu_f(0) \vdots J(0)\right)'\Omega^{-1}\left(\mu_f(0) \vdots J(0)\right)\begin{pmatrix} -\theta^* \\ 1\end{pmatrix} - \\ &\begin{pmatrix} 1 \\ \theta^*\end{pmatrix}'\left(\mu_f(0) \vdots J(0)\right)'\Omega^{-1}\left(\mu_f(0) \vdots J(0)\right)\begin{pmatrix} 1 \\ \theta^*\end{pmatrix}\Bigg].\end{aligned}$$

Proof: **a.** Because $\hat{V}_{ff}(\theta)^{-1} = \hat{V}_{ff}(\theta)^{-\frac{1}{2}\prime}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}$, $\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{ff}(\theta)\hat{V}_{ff}(\theta)^{-\frac{1}{2}\prime} = I_{k_f}$ and

$$\left(\frac{\partial \hat{V}_{ff}(\theta)^{-\frac{1}{2}}}{\partial\theta}\right)\hat{V}_{ff}(\theta)\hat{V}_{ff}(\theta)^{-\frac{1}{2}\prime} + \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\left(\frac{\partial \hat{V}_{ff}(\theta)}{\partial\theta}\right)\hat{V}_{ff}(\theta)^{-\frac{1}{2}\prime} + \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{ff}(\theta)\left(\frac{\partial \hat{V}_{ff}(\theta)^{-\frac{1}{2}}}{\partial\theta}\right)' = 0.$$

This equation implies that $\frac{\partial}{\partial\theta}\hat{V}_{ff}(\theta)^{-\frac{1}{2}} = -\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}$, since $\frac{\partial \hat{V}_{ff}(\theta)}{\partial\theta} = \hat{V}_{qf}(\theta) + \hat{V}_{qf}(\theta)'$ which results from the definition of $q_T(\theta, X) = \frac{\partial}{\partial\theta} f_T(\theta, X)$.

b. Using the product rule of differentiation:

$$\begin{aligned}
\frac{\partial}{\partial\theta}\hat{V}_{ff}(\theta)^{-\frac{1}{2}} f_T(\theta, X) = & \left(\frac{\partial}{\partial\theta}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\right) f_T(\theta, X) + \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\left(\frac{\partial}{\partial\theta} f_T(\theta, X)\right) \\
= & -\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X) + \hat{V}_{ff}(\theta)^{-\frac{1}{2}} q_T(\theta, X) \\
= & \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}(\theta).
\end{aligned}$$

c. The specification of $\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}(\theta)$ is $\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}(\theta) = \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\left[q_T(\theta, X) - \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X)\right]$:

$$\begin{aligned}
& \frac{\partial}{\partial\theta}\left(\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}(\theta)\right) \\
= & \left(\frac{\partial}{\partial\theta}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\right)\hat{D}(\theta) + \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\left(\frac{\partial}{\partial\theta}\left[q_T(\theta, X) - \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X)\right]\right) \\
= & -\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) + \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\left[\frac{\partial}{\partial\theta} q_T(\theta, X) - \left(\frac{\partial}{\partial\theta}\hat{V}_{qf}(\theta)\right)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X) - \right. \\
& \left.\hat{V}_{qf}(\theta)\left(\frac{\partial}{\partial\theta}\hat{V}_{ff}(\theta)^{-1}\right) f_T(\theta, X) - \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\left(\frac{\partial}{\partial\theta} f_T(\theta, X)\right)\right] \\
= & -\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) - \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{qq}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X) + \\
& \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X) + \\
& \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)'\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X) - \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1} q_T(\theta, X) \\
= & -2\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) - \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X).
\end{aligned}$$

since $\frac{\partial}{\partial\theta}\hat{V}_{qf}(\theta) = \hat{V}_{qq}(\theta)$ and $\frac{\partial}{\partial\theta} q_T(\theta, X) = 0$.

d.

$$\begin{aligned}
& \frac{\partial}{\partial\theta} f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) \\
= & \left(\frac{\partial}{\partial\theta}\hat{V}_{ff}(\theta)^{-\frac{1}{2}} f_T(\theta, X)\right)'\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}(\theta) + f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\left(\frac{\partial}{\partial\theta}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}(\theta)\right) \\
= & \hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) - 2 f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) - \\
& f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X).
\end{aligned}$$

e.

$$\begin{aligned}
\frac{\partial}{\partial\theta}\left(\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)\right) = & 2\left(\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\right)\left(\frac{\partial}{\partial\theta}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}(\theta)\right) \\
= & -4\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) - \\
& 2\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X).
\end{aligned}$$

f. The specification of $V_{\theta\theta}(\theta) = V_{qq}(\theta) - V_{qf}(\theta)V_{ff}(\theta)^{-1}V_{qf}(\theta)'$ is such that:

$$\begin{aligned}
&\tfrac{\partial}{\partial\theta}\hat{V}_{\theta\theta}(\theta) \\
= \;& \left(\tfrac{\partial}{\partial\theta}\hat{V}_{qq}(\theta)\right) - \left(\tfrac{\partial}{\partial\theta}\hat{V}_{qf}(\theta)\right)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)' - \hat{V}_{qf}(\theta)\left(\tfrac{\partial}{\partial\theta}\hat{V}_{ff}(\theta)^{-1}\right)\hat{V}_{qf}(\theta)' - \\
& \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\left(\tfrac{\partial}{\partial\theta}\hat{V}_{qf}(\theta)\right)' \\
= \;& -\hat{V}_{qq}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)' + \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)' + \\
& \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)' - \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qq}(\theta) \\
= \;& -\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)' - \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta).
\end{aligned}$$

g. The specification of $f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)$ is such that:

$$\begin{aligned}
&\tfrac{\partial}{\partial\theta}\left(f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)\right) \\
= \;& 2\left(\tfrac{\partial}{\partial\theta}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}f_T(\theta,X)\right)'\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X) + \\
& 2f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\left(\tfrac{\partial}{\partial\theta}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\right)\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X) + \\
& f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\left(\tfrac{\partial}{\partial\theta}\hat{V}_{\theta\theta}(\theta)\right)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X) \\
= \;& 2\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X) - \\
& 2f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X) - \\
& f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)'\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X) - \\
& f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X) \\
= \;& 2\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X) - \\
& 4f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X).
\end{aligned}$$

h. It follows from **e** and **g** above.

i. Use **d** above for the special case with $V_{ff}(\theta) = (1+\theta^2)\Omega$, $V_{qf}(\theta) = \theta\Omega$, $\mu_f(\theta) = \left(\mu_f(0) \vdots J(0)\right)\binom{1}{\theta}$, $D(\theta) = \left(\mu_f(0) \vdots J(0)\right)\binom{-\theta}{1}\frac{1}{1+\theta^2}$, $V_{\theta\theta}(\theta) = (1+\theta^2)^{-1}\Omega$ and since θ^* satisfies the population FOC:

$$\tfrac{1}{2}\tfrac{\partial}{\partial\theta}Q_p(\theta)|_{\theta^*} = \binom{1}{\theta^*}'\left(\mu_f(0) \vdots J(0)\right)'\Omega^{-1}\left(\mu_f(0) \vdots J(0)\right)\binom{-\theta^*}{1}\left(\frac{1}{1+(\theta^*)^2}\right)^2 = 0.$$

Lemma 3. The FOC (divided by two) for a stationary point θ^s of the population objective function reads:

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

with

$$D(\theta) = \quad J(\theta) - \left[V_{q_1 f}(\theta)V_{ff}(\theta)^{-1}\mu_f(\theta)\dots V_{q_m f}(\theta)V_{ff}(\theta)^{-1}\mu_f(\theta)\right] \tag{2}$$

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

$$V_{q_i f}(\theta) = \quad \lim_{T\to\infty} E\left[T(\tfrac{\partial}{\partial\theta_i}(f_T(\theta,X) - \mu_f(\theta)))\left(f_T(\theta,X) - \mu_f(\theta)\right)'\right], \qquad i = 1,\dots,m. \tag{3}$$

Proof: The derivative of $Q_p(\theta)$ with respect to θ consists of two parts. The derivative of $\mu_f(\theta)$ with respect to $\theta : J(\theta) = \frac{\partial}{\partial\theta'}\mu_f(\theta)$, and the derivative of $V_{ff}(\theta)^{-1}$ with respect to θ. To obtain the derivative of $V_{ff}(\theta)^{-1}$ with respect to θ, we start out with the derivative of $V_{ff}(\theta)$ with respect to $\theta :$

$$\begin{aligned}
\text{vec}(V_{ff}(\theta)) &= \lim_{T\to\infty} \text{vec}(\text{var}\left(\sqrt{T}f_T(\theta,X)\right)) \\
&= \lim_{T\to\infty} \text{vec}\left(E\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\left(f_t(\theta) - \mu_f(\theta)\right)\left(f_j(\theta) - \mu_f(\theta)\right)'\right)\right) \\
&= \lim_{T\to\infty} E\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\left[\left(f_j(\theta) - \mu_f(\theta)\right)\otimes\left(f_t(\theta) - \mu_f(\theta)\right)\right]\right) \\
\frac{\partial}{\partial\theta'}\text{vec}(V_{ff}(\theta)) &= \lim_{T\to\infty} \frac{\partial}{\partial\theta'}E\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\left[\left(f_j(\theta) - \mu_f(\theta)\right)\otimes\left(f_t(\theta) - \mu_f(\theta)\right)\right]\right) \\
&= \lim_{T\to\infty} E\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\left[\left(\frac{\partial}{\partial\theta'}f_j(\theta) - \frac{\partial}{\partial\theta'}\mu_f(\theta)\right)\otimes\left(f_t(\theta) - \mu_f(\theta)\right)\right]\right) + \\
&\quad \lim_{T\to\infty} E\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\left[\left(f_j(\theta) - \mu_f(\theta)\right)\otimes\left(\frac{\partial}{\partial\theta'}f_t(\theta) - \frac{\partial}{\partial\theta'}\mu_f(\theta)\right)\right]\right) \\
&= \lim_{T\to\infty} E\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\left[\left(q_j(\theta) - J(\theta)\right)\otimes\left(f_t(\theta) - \mu_f(\theta)\right)\right]\right) + \\
&\quad \lim_{T\to\infty} E\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\left[\left(f_j(\theta) - \mu_f(\theta)\right)\otimes\left(q_t(\theta) - J(\theta)\right)\right]\right) \\
&= \left(\text{vec}\left(V_{q_1 f}(\theta)\right)\dots\text{vec}\left(V_{q_m f}(\theta)\right)\right) + \left(\text{vec}\left(V_{q_1 f}(\theta)'\right)\dots\text{vec}\left(V_{q_m f}(\theta)'\right)\right)
\end{aligned}$$

with $q_j(\theta) = \frac{\partial}{\partial\theta'}f_j(\theta) = (q_{1,j}(\theta)\dots q_{m,j}(\theta))$ and

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

We can now specify the derivative of the objective function with respect to θ:

$$
\begin{aligned}
\tfrac{1}{2}\tfrac{\partial}{\partial\theta'}Q_p(\theta) = \quad & \mu_f(\theta)'V_{ff}(\theta)^{-1}\tfrac{\partial\mu_f(\theta)}{\partial\theta'} - \tfrac{1}{2}((\mu_f(\theta)\otimes\mu_f(\theta))'(V_{ff}(\theta)^{-1}\otimes V_{ff}(\theta)^{-1})\tfrac{\partial}{\partial\theta'}\text{vec}(V_{ff}(\theta)) \\
= \quad & \mu_f(\theta)'V_{ff}(\theta)^{-1}J(\theta) - \tfrac{1}{2}((\mu_f(\theta)\otimes\mu_f(\theta))'(V_{ff}(\theta)^{-1}\otimes V_{ff}(\theta)^{-1}) \\
& (\text{vec}\,(V_{q_1f}(\theta))\ldots\text{vec}\,(V_{q_mf}(\theta))) + (\text{vec}\,(V_{q_1f}(\theta)')\ldots\text{vec}\,(V_{q_mf}(\theta)')) \\
= \quad & \mu_f(\theta)'V_{ff}(\theta)^{-1}J(\theta) - \\
& \tfrac{1}{2}\left[\left(\mu_f(\theta)'V_{ff}(\theta)^{-1}V_{q_1f}(\theta)V_{ff}(\theta)^{-1}\mu_f(\theta)\ldots\mu_f(\theta)'V_{ff}(\theta)^{-1}V_{q_mf}(\theta)V_{ff}(\theta)^{-1}\mu_f(\theta)\right)+\right. \\
& \left.\left(\mu_f(\theta)'V_{ff}(\theta)^{-1}V_{q_1f}(\theta)'V_{ff}(\theta)^{-1}\mu_f(\theta)\ldots\mu_f(\theta)'V_{ff}(\theta)^{-1}V_{q_mf}(\theta)'V_{ff}(\theta)^{-1}\mu_f(\theta)\right)\right] \\
= \quad & \mu_f(\theta)'V_{ff}(\theta)^{-1}\left[J(\theta) - \left(V_{q_1f}(\theta)V_{ff}(\theta)^{-1}\mu_f(\theta)\ldots V_{q_mf}(\theta)V_{ff}(\theta)^{-1}\mu_f(\theta)\right)\right] \\
= \quad & \mu_f(\theta)'V_{ff}(\theta)^{-1}D(\theta),
\end{aligned}
$$

with $D(\theta) = J(\theta) - \left[V_{q_1f}(\theta)V_{ff}(\theta)^{-1}\mu_f(\theta)\ldots V_{q_mf}(\theta)V_{ff}(\theta)^{-1}\mu_f(\theta)\right]$. See also Kleibergen (2005).

Lemma 4. The FOC (divided by two) for a stationary point $\hat{\theta}^s$ of the CUE sample objective function reads:

$$
\frac{1}{2}\frac{\partial}{\partial\theta'}\hat{Q}_s(\hat{\theta}^s) = 0 \Leftrightarrow f_T(\hat{\theta}^s, X)'\hat{V}_{ff}(\hat{\theta}^s)^{-1}\hat{D}(\hat{\theta}^s) = 0, \tag{4}
$$

with

$$
\hat{D}(\theta) = \quad q_T(\theta, X) - \left[\hat{V}_{q_1f}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X)\ldots\hat{V}_{q_mf}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X)\right] \tag{5}
$$

and

$$
\hat{V}(\theta) = \begin{pmatrix} \hat{V}_{ff}(\theta) & \hat{V}_{fq}(\theta) \\ \hat{V}_{qf}(\theta) & \hat{V}_{qq}(\theta) \end{pmatrix}, \tag{6}
$$

with $\hat{V}_{qf}(\theta) = \hat{V}_{fq}(\theta)' = (\hat{V}_{q_1f}(\theta)'\ldots\hat{V}_{q_mf}(\theta)')'$, $\hat{V}_{qq}(\theta) = (\hat{V}_{q_iq_j}(\theta)) : i,j = 1,\ldots,m$; and $\hat{V}_{ff}(\theta)$, $\hat{V}_{q_if}(\theta)$, $\hat{V}_{q_iq_j}(\theta)$ are $k_f \times k_f$ dimensional matrices for $i,j = 1,\ldots,m$.

Proof: See Lemma 3 above and also Kleibergen (2005).

Lemma 5. When Assumptions 1 and 2 hold and for θ^* the pseudo-true value minimizing the population continuous updating objective function:

$$
\begin{aligned}
\sqrt{T}\left(f_T(\theta^*, X) - \mu_f(\theta^*)\right) \quad & \underset{d}{\rightarrow} \psi_f(\theta^*), \\
\sqrt{T}\text{vec}\left(\hat{D}(\theta^*) - D(\theta^*)\right) \quad & \underset{d}{\rightarrow} \psi_\theta(\theta^*),
\end{aligned} \tag{7}
$$

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

$$\begin{aligned}
\hat{D}(\theta) = \quad & q_T(\theta, X) - \left[\hat{V}_{q_1 f}(\theta) \hat{V}_{ff}(\theta)^{-1} f_T(\theta, X) \ldots \hat{V}_{q_m f}(\theta) \hat{V}_{ff}(\theta)^{-1} f_T(\theta, X) \right] \\
D(\theta) = \quad & J(\theta) - \left[V_{q_1 f}(\theta) V_{ff}(\theta)^{-1} \mu_f(\theta) \ldots V_{q_m f}(\theta) V_{ff}(\theta)^{-1} \mu_f(\theta) \right], \\
V_{\theta_i f}(\theta) = \quad & \lim_{T \to \infty} E \left[T (\tfrac{\partial}{\partial \theta_i} (f_T(\theta, X) - \mu_f(\theta))) \left(f_T(\theta, X) - \mu_f(\theta) \right)' \right], \ i = 1, \ldots, m, \\
\hat{V}(\theta) = \quad & \begin{pmatrix} \hat{V}_{ff}(\theta) & \hat{V}_{fq}(\theta) \\ \hat{V}_{qf}(\theta) & \hat{V}_{qq}(\theta) \end{pmatrix},
\end{aligned} \tag{8}$$

where $\hat{V}_{qf}(\theta) = \hat{V}_{fq}(\theta)' = (\hat{V}_{q_1 f}(\theta)' \ldots \hat{V}_{q_m f}(\theta)')'$, $\hat{V}_{qq}(\theta) = (\hat{V}_{q_i q_j}(\theta)) : i, j = 1, \ldots, m$; $\hat{V}_{ff}(\theta)$, $\hat{V}_{q_i f}(\theta)$, $\hat{V}_{q_i q_j}(\theta)$ are $k_f \times k_f$ dimensional matrices, $\psi_\theta(\theta^*) = \psi_q(\theta^*) - V_{qf}(\theta^*) V_{ff}(\theta^*)^{-1} \psi_f(\theta^*)$ independent of $\psi_f(\theta^*)$, and

$$\begin{aligned}
\psi_f(\theta^*) &\sim N(0, V_{ff}(\theta^*)), \\
\psi_\theta(\theta^*) &\sim N(0, V_{\theta\theta}(\theta^*)),
\end{aligned} \tag{9}$$

with $V_{\theta\theta}(\theta) = V_{qq}(\theta) - V_{qf}(\theta) V_{ff}(\theta)^{-1} V_{fq}(\theta)$.

Proof: The joint limit behavior of $f_T(\theta, X)$ and $q_T(\theta, X)$ at the pseudo-true value θ^* reads:

$$\sqrt{T} \begin{pmatrix} f_T(\theta^*, X) - \mu_f(\theta^*) \\ vec(q_T(\theta^*, X) - J(\theta^*)) \end{pmatrix} \underset{d}{\to} \begin{pmatrix} \psi_f(\theta) \\ \psi_q(\theta) \end{pmatrix}.$$

We pre-multiply it by

$$\hat{R}(\theta^*) = \begin{pmatrix} I_{k_f} & 0 \\ -\hat{V}_{qf}(\theta^*) \hat{V}_{ff}(\theta^*)^{-1} & I_{k_f m} \end{pmatrix} \underset{p}{\to} \begin{pmatrix} I_{k_f} & 0 \\ -V_{qf}(\theta^*) V_{ff}(\theta^*)^{-1} & I_{k_f m} \end{pmatrix} = R(\theta^*),$$

to obtain

$$\begin{aligned}
\sqrt{T} \left[\hat{R}(\theta^*) \begin{pmatrix} f_T(\theta^*, X) \\ vec(q_T(\theta^*, X)) \end{pmatrix} - R(\theta^*) \begin{pmatrix} \mu_f(\theta^*) \\ vec(J(\theta^*)) \end{pmatrix} \right] \underset{d}{\to} \quad R(\theta^*) \begin{pmatrix} \psi_f(\theta^*) \\ \psi_q(\theta^*) \end{pmatrix} \quad \Leftrightarrow \\
\sqrt{T} \begin{pmatrix} f_T(\theta^*, X) - \mu_f(\theta^*) \\ vec(\hat{D}(\theta^*) - D(\theta^*)) \end{pmatrix} \underset{d}{\to} \quad \begin{pmatrix} \psi_f(\theta^*) \\ \psi_\theta(\theta^*) \end{pmatrix},
\end{aligned}$$

with $\psi_\theta(\theta^*) = \psi_q(\theta^*) - V_{qf}(\theta^*)V_{ff}(\theta^*)^{-1}\psi_f(\theta^*)$, which is independent of $\psi_f(\theta^*)$ since

$$R(\theta^*)V(\theta^*)R(\theta^*)' = \begin{pmatrix} V_{ff}(\theta^*) & 0 \\ 0 & V_{\theta\theta}(\theta^*) \end{pmatrix},$$

where $V_{\theta\theta}(\theta^*) = V_{qq}(\theta^*) - V_{qf}(\theta^*)V_{ff}(\theta^*)^{-1}V_{qf}(\theta^*)'$, so $\psi_f(\theta^*)$ and $\psi_\theta(\theta^*)$ are uncorrelated and independent since they are normally distributed random variables. See also Lemma 1 in Kleibergen (2005).

Lemma 6. When Assumptions 1 and 2 hold, θ^* is the minimizer of the population continuous updating objective function and

$$\begin{aligned} \tilde{\mu}_f(\theta^*) = & \lim_{T\to\infty} \sqrt{T}\mu_f(\theta^*) \\ \tilde{D}(\theta^*) = & \lim_{T\to\infty} \sqrt{T}D(\theta^*) \end{aligned} \tag{10}$$

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

$$Ts(\theta^*) \underset{d}{\to} \tilde{\mu}_f(\theta^*)'V_{ff}(\theta^*)^{-1}\Psi_\theta(\theta^*) + \psi_f(\theta^*)'V_{ff}(\theta^*)^{-1}\tilde{D}(\theta^*) + \psi_f(\theta^*)'V_{ff}(\theta^*)^{-1}\Psi_\theta(\theta^*), \tag{11}$$

with $\text{vec}(\Psi_\theta(\theta^*)) = \psi_\theta(\theta^*)$, 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^* :

$$\lim_{T\to\infty} E\left[T \times s(\theta^*)\right] = 0. \tag{12}$$

Proof: The joint limit behaviors of $f_T(\theta^*, X)$, $\hat{D}(\theta^*)$ and $\hat{V}_{ff}(\theta^*)$ are such that:

$$\begin{aligned} Ts(\theta^*) = & \left(\sqrt{T}f_T(\theta^*, X)\right)' \hat{V}_{ff}(\theta^*)^{-1} \left(\sqrt{T}\hat{D}(\theta^*)\right) \\ \underset{d}{\to} & \left[\tilde{\mu}_f(\theta^*) + \psi_f(\theta^*)\right]' V_{ff}(\theta^*)^{-1} \left[\tilde{D}(\theta^*) + \Psi_\theta(\theta^*)\right] \\ = & \tilde{\mu}_f(\theta^*)'V_{ff}(\theta^*)^{-1}\Psi_\theta(\theta^*) + \psi_f(\theta^*)'V_{ff}(\theta^*)^{-1} \left[\tilde{D}(\theta^*) + \Psi_\theta(\theta^*)\right] \\ = & \left(\tilde{\mu}_f(\theta^*) + \psi_f(\theta^*)\right)' V_{ff}(\theta^*)^{-1}\Psi_\theta(\theta^*) + \psi_f(\theta^*)'V_{ff}(\theta^*)^{-1}\tilde{D}(\theta^*), \end{aligned}$$

where $\text{vec}(\Psi_\theta(\theta^*)) = \psi_\theta$, since $\tilde{\mu}_f(\theta^*)'V_{ff}(\theta^*)^{-1}\tilde{D}(\theta^*) = 0$. Because $\psi_f(\theta^*)$ and $\psi_\theta(\theta^*)$ are independently distributed, this shows that the expected value of the limit of the score of the CUE sample objective function equals zero at the pseudo-true value θ^*.

Lemmas 7-10: The general proof of the robustness of the DRLM test under strong misspecification is conducted in a sequence of steps. We start with proving Lemma 7. Thereafter we work in Lemmas 8-10 towards a higher order expansion of the sample score. Lemma 8 therefore constructs a higher order expansion for the covariance matrix estimator of the sample moment vector, $\hat{V}_{ff}(\theta)$, while Lemma 9 does so for the centered Jacobian estimator $\hat{D}(\theta)$. Lemma 10 combines these results to construct a higher order expansion of the sample score, which we use to show its limit behavior under different strengths of misspecification and identification. Next, based on Lemma 10, the proof of Theorem 4 shows that the DRLM test in Theorem 4 is size correct.

Lemma 7. Under Assumption 1^*, for θ equal to the pseudo-true value θ^*, we have:

$$\sqrt{T}\begin{pmatrix} f_T(\theta, X) - \mu_f(\theta) \\ \text{vec}(\hat{D}(\theta) - D(\theta)) \\ \text{vech}(\hat{V}_{ff}(\theta) - V_{ff}(\theta)) \end{pmatrix} \underset{d}{\rightarrow} \begin{pmatrix} \psi_f(\theta) \\ \psi_\theta(\theta) \\ \psi_{ff}(\theta) \end{pmatrix} \sim N(0, \bar{\mathcal{V}}(\theta)), \tag{13}$$

with

$$\bar{\mathcal{V}}(\theta) = \begin{pmatrix} V_{ff}(\theta) & V_{\theta f}(\theta)' & V_{ff,f}(\theta)' \\ V_{\theta f}(\theta) & V_{\theta\theta}(\theta) & V_{ff,\theta}(\theta)' \\ V_{ff,f}(\theta) & V_{ff,\theta}(\theta) & V_{ff,ff}(\theta) \end{pmatrix}, \tag{14}$$

and we have the following specification for the covariance matrices:

$$\begin{aligned} V_{\theta f}(\theta) = &\ \lim_{T\to\infty} E\left[T\left(\text{vec}(\hat{D}(\theta) - D(\theta))\right)\left(f_T(\theta, X) - \mu_f(\theta)\right)'\right] = 0 \\ V_{\theta\theta}(\theta) = &\ \lim_{T\to\infty} E\left[T\left(\text{vec}(\hat{D}(\theta) - D(\theta))\right)\left(\text{vec}(\hat{D}(\theta) - D(\theta))\right)'\right] \\ = &\ V_{qq}(\theta) - V_{qf}(\theta)V_{ff}(\theta)^{-1}V_{qf}(\theta)' \end{aligned} \tag{15}$$

Proof of Lemma 7: To construct an appropriate weight matrix for the sample score which also incorporates the effect of strong misspecification, we redefine Assumption 1^* in Lemma 7, which is the analog of Lemma 5, so it concerns the three different elements of the sample score. We pre-multiply the expression in Assumption 1^* by

$$\mathcal{R}(\theta) = \begin{pmatrix} I_{k_f} & 0 & 0 \\ -V_{qf}(\theta)V_{ff}(\theta)^{-1} & I_{mk_f} & 0 \\ 0 & 0 & I_{\frac{1}{2}k_f(k_f+1)} \end{pmatrix}$$

and note that

$$\hat{\mathcal{R}}(\theta)\begin{pmatrix} f_T(\theta,X) \\ \operatorname{vec}(q_T(\theta,X)) \\ \operatorname{vech}(\hat{V}_{ff}(\theta)) \end{pmatrix} = \begin{pmatrix} f_T(\theta,X) \\ \operatorname{vec}(\hat{D}(\theta)) \\ \operatorname{vech}(\hat{V}_{ff}(\theta)) \end{pmatrix}, \mathcal{R}(\theta)\begin{pmatrix} \mu_f(\theta) \\ \operatorname{vec}(J_T(\theta)) \\ \operatorname{vech}(V_{ff}(\theta)) \end{pmatrix} = \begin{pmatrix} \mu_f(\theta) \\ \operatorname{vec}(D(\theta)) \\ \operatorname{vech}(V_{ff}(\theta)) \end{pmatrix},$$

so when θ equals the pseudo-true value θ^* :

$$\begin{array}{cc} & \sqrt{T}\left[\hat{\mathcal{R}}(\theta)\begin{pmatrix} f_T(\theta,X) \\ \operatorname{vec}(q_T(\theta,X)) \\ \operatorname{vech}(\hat{V}_{ff}(\theta)) \end{pmatrix} - \mathcal{R}(\theta)\begin{pmatrix} \mu_f(\theta) \\ \operatorname{vec}(J(\theta)) \\ \operatorname{vech}(V_{ff}(\theta)) \end{pmatrix}\right] \\ = & \sqrt{T}\begin{pmatrix} f_T(\theta,X)-\mu_f(\theta) \\ \operatorname{vec}(\hat{D}(\theta)-D(\theta)) \\ \operatorname{vech}(\hat{V}_{ff}(\theta)-V_{ff}(\theta)) \end{pmatrix} \\ \underset{d}{\rightarrow} & \begin{pmatrix} \psi_f(\theta) \\ \psi_\theta(\theta) \\ \psi_{ff}(\theta) \end{pmatrix} \sim N(0,\bar{\mathcal{V}}(\theta)). \end{array}$$

with $\bar{\mathcal{V}}(\theta)=\mathcal{R}(\theta)\mathcal{V}(\theta)\mathcal{R}(\theta)'$.

Lemma 8. When Assumptions 1^* and 2 hold and θ equals the pseudo-true value θ^*, a higher order decomposition of $\hat{V}_{ff}(\theta)^{-1}$ reads

$$\begin{aligned} \hat{V}_{ff}(\theta)^{-1} &= V_{ff}(\theta)^{-1} - V_{ff}(\theta)^{-1}\left(\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right)V_{ff}(\theta)^{-1} + \\ & V_{ff}(\theta)^{-1}\left(\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right)V_{ff}(\theta)^{-1}\left(\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right)V_{ff}(\theta)^{-1} + o_p(T^{-1}). \end{aligned} \tag{16}$$

Proof: To obtain the higher order specification of $\hat{V}_{ff}(\theta)^{-1}$, we specify it as

$$\begin{aligned} \hat{V}_{ff}(\theta)^{-1} = & \left[V_{ff}(\theta)+\left(\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right)\right]^{-1} \\ = & V_{ff}(\theta)^{-\frac{1}{2}}\left[I_{k_f}+V_{ff}(\theta)^{-\frac{1}{2}}\left(\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right)V_{ff}(\theta)^{-\frac{1}{2}}\right]^{-1}V_{ff}(\theta)^{-\frac{1}{2}} \\ = & V_{ff}(\theta)^{-1} - V_{ff}(\theta)^{-1}\left(\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right)V_{ff}(\theta)^{-1} + V_{ff}(\theta)^{-1} \\ & \left(\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right)V_{ff}(\theta)^{-1}\left(\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right)V_{ff}(\theta)^{-1} + o_p(T^{-1}), \end{aligned}$$

where the $o_p(T^{-1})$ order of the remainder term results from the $\sqrt{T}$ convergence rate of the covariance matrix estimator.

Lemma 9. When Assumptions 1* and 2 hold and θ equals the pseudo-true value θ^*, the higher order specification of $\hat{D}(\theta)$ reads:

$$\hat{D}(\theta) = D(\theta) + \Psi_\theta(\theta)/\sqrt{T} + o_p(T^{-\frac{1}{2}}). \tag{17}$$

with $\text{vec}(\Psi_\theta(\theta)) = \psi_\theta(\theta)$ as in Lemma 6.

Proof: Results from Lemma 5.

Lemma 10. When Assumptions 1* and 2 hold, θ equals the pseudo-true value θ^* and depending on the convergence rate of $D(\theta)$ and $\mu_f(\theta)$, the specification of the score vector $\hat{s}(\theta)' = \hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X)$ reads for:

1. $\tilde{D}(\theta^*) = \lim_{T\to\infty}\sqrt{T}D(\theta^*), \tilde{\mu}_f(\theta^*) = \lim_{T\to\infty}\sqrt{T}\mu_f(\theta^*)$ both finite and non-negligible:
$$T\hat{s}(\theta^*)' = \begin{pmatrix} V_{ff}(\theta^*)^{-1}(\tilde{D}(\theta^*) + \Psi_\theta(\theta^*)) \\ (I_m \otimes V_{ff}(\theta^*)^{-1}\tilde{\mu}_f(\theta^*)) \\ 0 \end{pmatrix}' \begin{pmatrix} \psi_f(\theta^*) \\ \psi_\theta(\theta^*) \\ \psi_{ff}(\theta^*) \end{pmatrix} + O_p(T^{-\frac{1}{2}})$$

2. $D(\theta^*), \tilde{\mu}_f(\theta^*) = \lim_{T\to\infty}\sqrt{T}\mu_f(\theta^*)$ both finite and non-negligible:
$$\sqrt{T}\hat{s}(\theta^*)' = \begin{pmatrix} V_{ff}(\theta^*)^{-1}D(\theta^*) \\ 0 \\ 0 \end{pmatrix}' \begin{pmatrix} \psi_f(\theta^*) \\ \psi_\theta(\theta^*) \\ \psi_{ff}(\theta^*) \end{pmatrix} + O_p(T^{-\frac{1}{2}})$$

3. $\tilde{D}(\theta^*) = \lim_{T\to\infty}\sqrt{T}D(\theta^*),\ \mu_f(\theta^*)$ both finite and non-negligible:
$$\sqrt{T}\hat{s}(\theta^*)' = \begin{pmatrix} 0 \\ (I_m \otimes V_{ff}(\theta^*)^{-1}\mu_f(\theta^*)) \\ 0 \end{pmatrix}' \begin{pmatrix} \psi_f(\theta^*) \\ \psi_\theta(\theta^*) \\ \psi_{ff}(\theta^*) \end{pmatrix} + O_p(T^{-\frac{1}{2}})$$

4. $D(\theta^*),\ \mu_f(\theta^*)$ both finite and non-negligible:
$$\sqrt{T}\hat{s}(\theta^*)' = \begin{pmatrix} V_{ff}(\theta^*)^{-1}D(\theta^*) \\ (I_m \otimes V_{ff}(\theta^*)^{-1}\mu_f(\theta^*)) \\ -D_{k_f}'\left(V_{ff}(\theta^*)^{-1}\mu_f(\theta^*) \otimes V_{ff}(\theta^*)^{-1}D(\theta^*)\right) \end{pmatrix}' \begin{pmatrix} \psi_f(\theta^*) \\ \psi_\theta(\theta^*) \\ \psi_{ff}(\theta^*) \end{pmatrix} + O_p(T^{-\frac{1}{2}}).$$

with D_{k_f} the $k_f^2 \times \frac{1}{2}k_f(k_f+1)$ dimensional duplication matrix so $\text{vec}(V_{ff}(\theta)) = D_{k_f}\text{vech}(V_{ff}(\theta))$.

Proof: We consider the four cases one by one.

1. For $\tilde{D}(\theta^*) = \lim_{T\to\infty} \sqrt{T} D(\theta^*)$, $\tilde{\mu}_f(\theta^*) = \lim_{T\to\infty} \sqrt{T}\mu_f(\theta^*)$ both finite, $\tilde{D}(\theta^*)' V_{ff}(\theta^*)^{-1}\tilde{\mu}_f(\theta^*) = 0$, the higher order specification of the score at the pseudo-true value θ^* then reads:

$$
\begin{aligned}
T\hat{s}(\theta^*)' &= \left(\sqrt{T}\hat{D}(\theta^*)\right)' \hat{V}_{ff}(\theta^*)^{-1} \left(\sqrt{T} f_T(\theta^*, X)\right) \\
&= \tilde{D}(\theta^*)' V_{ff}(\theta^*)^{-1}\psi_f(\theta^*) + \Psi_\theta(\theta^*)' V_{ff}(\theta^*)^{-1}\tilde{\mu}_f(\theta^*) + \Psi_\theta(\theta^*)' V_{ff}(\theta^*)^{-1}\psi_f(\theta^*) + O_p(T^{-\frac{1}{2}}) \\
&= \begin{pmatrix} V_{ff}(\theta^*)^{-1}(\tilde{D}(\theta^*) + \Psi_\theta(\theta^*)) \\ (I_m \otimes V_{ff}(\theta^*)^{-1}\tilde{\mu}_f(\theta^*)) \\ 0 \end{pmatrix}' \begin{pmatrix} \psi_f(\theta^*) \\ \psi_\theta(\theta^*) \\ \psi_{ff}(\theta^*) \end{pmatrix} + O_p(T^{-\frac{1}{2}})
\end{aligned}
$$

where we used that $K_{k_f m}(V_{ff}(\theta^*)^{-1}\tilde{\mu}_f(\theta^*) \otimes I_m) = (I_m \otimes V_{ff}(\theta^*)^{-1}\tilde{\mu}_f(\theta^*))$ for K_{pr} the $pr \times pr$ dimensional commutation matrix so for the $p \times r$ dimensional matrix A : $\text{vec}(A) = K_{rp}\text{vec}(A')$, $\text{vec}(A') = K_{pr}\text{vec}(A)$, $\text{vec}(\Psi_\theta(\theta^*)) = \psi_\theta(\theta^*)$.

2. For $D(\theta^*)$, $\tilde{\mu}_f(\theta^*) = \lim_{T\to\infty} \sqrt{T}\mu_f(\theta^*)$ both finite and non-negligible and $D(\theta^*)' V_{ff}(\theta^*)^{-1}\tilde{\mu}_f(\theta^*) = 0$, the higher order specification of the score at the pseudo-true value θ^* then reads:

$$
\begin{aligned}
&\sqrt{T}\hat{s}(\theta^*)' \\
= \;& \hat{D}(\theta^*)'\hat{V}_{ff}(\theta^*)^{-1} \left(\sqrt{T} f_T(\theta^*, X)\right) \\
= \;& \left[D(\theta^*) + (\hat{D}(\theta^*) - D(\theta^*))\right]' \\
& [V_{ff}(\theta^*)^{-1} - V_{ff}(\theta^*)^{-1}\left(\hat{V}_{ff}(\theta^*) - V_{ff}(\theta^*)\right) V_{ff}(\theta^*)^{-1} + V_{ff}(\theta^*)^{-1}\left(\hat{V}_{ff}(\theta^*) - V_{ff}(\theta^*)\right) V_{ff}(\theta^*)^{-1} \\
& \left.\left(\hat{V}_{ff}(\theta^*) - V_{ff}(\theta^*)\right) V_{ff}(\theta^*)^{-1}\right] \left[\sqrt{T}\mu_f(\theta^*) + \sqrt{T}(f_T(\theta^*, X) - \mu_f(\theta^*))\right] + o_p(T^{-1}) \\
= \;& D(\theta^*)' \left[V_{ff}(\theta^*)^{-1} - V_{ff}(\theta^*)^{-1}\left(\hat{V}_{ff}(\theta^*) - V_{ff}(\theta^*)\right) V_{ff}(\theta^*)^{-1} + \right. \\
& \left. V_{ff}(\theta^*)^{-1}\left(\hat{V}_{ff}(\theta^*) - V_{ff}(\theta^*)\right) V_{ff}(\theta^*)^{-1}\left(\hat{V}_{ff}(\theta^*) - V_{ff}(\theta^*)\right) V_{ff}(\theta^*)^{-1}\right] \\
& \left[\sqrt{T}\mu_f(\theta^*) + \sqrt{T}(f_T(\theta^*, X) - \mu_f(\theta^*))\right] + O_p(T^{-\frac{1}{2}}) \\
= \;& D(\theta^*)' V_{ff}(\theta^*)^{-1}\tilde{\mu}_f(\theta^*) + D(\theta^*)' V_{ff}(\theta^*)^{-1}\psi_f(\theta^*) + O_p(T^{-\frac{1}{2}}) \\
= \;& D(\theta^*)' V_{ff}(\theta^*)^{-1}\psi_f(\theta^*) + O_p(T^{-\frac{1}{2}}) \\
= \;& \begin{pmatrix} V_{ff}(\theta^*)^{-1} D(\theta^*) \\ 0 \\ 0 \end{pmatrix}' \begin{pmatrix} \psi_f(\theta^*) \\ \psi_\theta(\theta^*) \\ \psi_{ff}(\theta^*) \end{pmatrix} + O_p(T^{-\frac{1}{2}})
\end{aligned}
$$

3. For $\tilde{D}(\theta^*) = \lim_{T\to\infty} \sqrt{T} D(\theta^*)$, $\mu_f(\theta^*)$ both finite and non-negligible and $\tilde{D}(\theta^*)' V_{ff}(\theta^*)^{-1}\mu_f(\theta^*) =$

0, the higher order specification of the score at the pseudo-true value θ^* then reads:

$$\begin{aligned}
& \sqrt{T}\hat{s}(\theta^*)' \\
= \; & \left(\sqrt{T}\hat{D}(\theta^*)\right)' \hat{V}_{ff}(\theta^*)^{-1}\left(f_T(\theta^*, X)\right) \\
= \; & \left[\sqrt{T}D(\theta^*) + \sqrt{T}(\hat{D}(\theta^*) - D(\theta^*))\right]' \\
& [V_{ff}(\theta^*)^{-1} - V_{ff}(\theta^*)^{-1}\left(\hat{V}_{ff}(\theta^*) - V_{ff}(\theta^*)\right)V_{ff}(\theta^*)^{-1} + V_{ff}(\theta^*)^{-1}\left(\hat{V}_{ff}(\theta^*) - V_{ff}(\theta^*)\right)V_{ff}(\theta^*)^{-1} \\
& \left(\hat{V}_{ff}(\theta^*) - V_{ff}(\theta^*)\right)V_{ff}(\theta^*)^{-1}\Big]\left[\mu_f(\theta^*) + (f_T(\theta^*, X) - \mu_f(\theta^*))\right] + o_p(T^{-1}) \\
= \; & \left[\tilde{D}(\theta^*) + \Psi_\theta(\theta^*)\right]' V_{ff}(\theta^*)^{-1}\mu_f(\theta^*) + O_p(T^{-\frac{1}{2}}) \\
= \; & \Psi_\theta(\theta^*)'V_{ff}(\theta^*)^{-1}\mu_f(\theta^*) + O_p(T^{-\frac{1}{2}}) \\
= \; & \begin{pmatrix} 0 \\ (I_m \otimes V_{ff}(\theta^*)^{-1}\mu_f(\theta^*)) \\ 0 \end{pmatrix}' \begin{pmatrix} \psi_f(\theta^*) \\ \psi_\theta(\theta^*) \\ \psi_{ff}(\theta^*) \end{pmatrix} + O_p(T^{-\frac{1}{2}}).
\end{aligned}$$

4. For $D(\theta^*)$, $\mu_f(\theta^*)$ both finite and non-negligible and $D(\theta^*)'V_{ff}(\theta^*)^{-1}\mu_f(\theta^*) = 0$, the higher order specification of the score at the pseudo-true value θ^* then reads:

$$\begin{aligned}
& \sqrt{T}\hat{s}(\theta^*)' \\
= \; & \sqrt{T}\hat{D}(\theta^*)'\hat{V}_{ff}(\theta^*)^{-1}f_T(\theta^*, X) \\
= \; & \sqrt{T}\left[D(\theta^*) + (\hat{D}(\theta^*) - D(\theta^*))\right]' \\
& \Big[V_{ff}(\theta^*)^{-1} - V_{ff}(\theta^*)^{-1}\left(\hat{V}_{ff}(\theta^*) - V_{ff}(\theta^*)\right)V_{ff}(\theta^*)^{-1} + V_{ff}(\theta^*)^{-1}\left(\hat{V}_{ff}(\theta^*) - V_{ff}(\theta^*)\right)V_{ff}(\theta^*)^{-1} \\
& \left(\hat{V}_{ff}(\theta^*) - V_{ff}(\theta^*)\right)V_{ff}(\theta^*)^{-1}\Big]\left[\mu_f(\theta^*) + (f_T(\theta^*, X) - \mu_f(\theta^*))\right] + O_p(T^{-\frac{1}{2}}) \\
= \; & \sqrt{T}D(\theta^*)'V_{ff}(\theta^*)^{-1}\mu_f(\theta^*) - D(\theta^*)'V_{ff}(\theta^*)^{-1}\left[\sqrt{T}\left(\hat{V}_{ff}(\theta^*) - V_{ff}(\theta^*)\right)\right]V_{ff}(\theta^*)^{-1}\mu_f(\theta^*) + \\
& \left[\sqrt{T}(\hat{D}(\theta^*) - D(\theta^*))\right]' V_{ff}(\theta^*)^{-1}\mu_f(\theta^*) + D(\theta^*)'V_{ff}(\theta^*)^{-1}\left[\sqrt{T}(f_T(\theta^*, X) - \mu_f(\theta^*))\right] + O_p(T^{-\frac{1}{2}}) \\
= \; & -D(\theta^*)'V_{ff}(\theta^*)^{-1}\left[\sqrt{T}\left(\hat{V}_{ff}(\theta^*) - V_{ff}(\theta^*)\right)\right]V_{ff}(\theta^*)^{-1}\mu_f(\theta^*) + \\
& \Psi_\theta' V_{ff}(\theta^*)^{-1}\mu_f(\theta^*) + D(\theta^*)'V_{ff}(\theta^*)^{-1}\psi_f(\theta^*) + O_p(T^{-\frac{1}{2}}) \\
= \; & \begin{pmatrix} V_{ff}(\theta^*)^{-1}D(\theta^*) \\ (I_m \otimes V_{ff}(\theta^*)^{-1}\mu_f(\theta^*)) \\ -D_{k_f}'\left(V_{ff}(\theta^*)^{-1}\mu_f(\theta^*) \otimes V_{ff}(\theta^*)^{-1}D(\theta^*)\right) \end{pmatrix}' \begin{pmatrix} \psi_f(\theta^*) \\ \psi_\theta(\theta^*) \\ \psi_{ff}(\theta^*) \end{pmatrix} + O_p(T^{-\frac{1}{2}}).
\end{aligned}$$

Lemma 11. Under the regularity conditions that: (i) X_t is i.i.d.; (ii) $f(\theta, X_t)$, $q_{it}(\theta, X_t)$, and $f(\theta, X_t)q_{it}(\theta, X_t)'$ all have finite second moments:

$$\hat{V}(\theta) \underset{p}{\rightarrow} V(\theta) \;,$$

where $q_{it}(\theta, X_t) = \frac{\partial f_t(\theta, X_t)}{\partial \theta_i}$,

$$\hat{V}(\theta) = \begin{pmatrix} \hat{V}_{ff}(\theta) & \hat{V}_{fq}(\theta) \\ \hat{V}_{qf}(\theta) & \hat{V}_{qq}(\theta) \end{pmatrix},$$

with $\hat{V}_{qf}(\theta) = \hat{V}_{fq}(\theta)' = (\hat{V}_{q_1 f}(\theta)' \dots \hat{V}_{q_m f}(\theta)')', \hat{V}_{qq}(\theta) = (\hat{V}_{q_i q_i}(\theta))$, and

$$\begin{aligned}
\hat{V}_{ff}(\theta) &= \frac{1}{T}\sum_{t=1}^{T} \left(f(\theta, X_t) - f_T(\theta, X)\right)\left(f(\theta, X_t) - f_T(\theta, X)\right)' \\
\hat{V}_{q_i f}(\theta) &= \frac{1}{T}\sum_{t=1}^{T} \left(q_{it}(\theta, X) - q_{iT}(\theta, X)\right)\left(f(\theta, X_t) - f_T(\theta, X)\right)' \\
\hat{V}_{q_i q_i}(\theta) &= \frac{1}{T}\sum_{t=1}^{T} \left(q_{it}(\theta, X) - q_{iT}(\theta, X)\right)\left(q_{it}(\theta, X) - q_{iT}(\theta, X)\right)'.
\end{aligned}$$

Proof: Because of the similarities in $\hat{V}_{ff}(\theta), \hat{V}_{q_i f}(\theta), \hat{V}_{q_i q_i}(\theta)$, we just show $\hat{V}_{ff}(\theta) \underset{p}{\rightarrow} V_{ff}(\theta)$:

$$\begin{aligned}
\hat{V}_{ff}(\theta) = \; & \tfrac{1}{T}\textstyle\sum_{t=1}^{T} \left(f(\theta, X_t) - f_T(\theta, X)\right)\left(f(\theta, X_t) - f_T(\theta, X)\right)' \\
= \; & \tfrac{1}{T}\textstyle\sum_{t=1}^{T} f(\theta, X_t) f(\theta, X_t)' - f_T(\theta, X) f_T(\theta, X)' \\
\underset{p}{\rightarrow} \; & E(f(\theta, X_t) f(\theta, X_t)') - \mu_f(\theta)\mu_f(\theta)' \\
= \; & V_{ff}(\theta).
\end{aligned}$$

Remark on Assumption 2 For the sample score to converge to the population score, the derivative of the sample covariance estimator has to converge to the derivative of the population covariance matrix. Assumption 2(i) requests a consistent covariance estimator for $V(\theta)$. For Assumption 2(ii), Kleibergen (2005, p.1120) states that (see also the proof of Lemma 3 in the Online Appendix):

$$\frac{\partial \text{vec}(V_{ff}(\theta))}{\partial \theta'} = (\text{vec}(V_{q_1 f}(\theta)')...\text{vec}(V_{q_m f}(\theta)')) + (\text{vec}(V_{q_1 f}(\theta))...\text{vec}(V_{q_m f}(\theta))), \tag{18}$$

so Assumption 2(ii) calls for

$$\frac{\partial \text{vec}(\hat{V}_{ff}(\theta))}{\partial \theta'} = (\text{vec}(\hat{V}_{q_1 f}(\theta)')...\text{vec}(\hat{V}_{q_m f}(\theta)')) + (\text{vec}(\hat{V}_{q_1 f}(\theta))...\text{vec}(\hat{V}_{q_m f}(\theta))), \tag{19}$$

to hold as well. Hence, the covariance matrix estimator $\hat{V}_{ff}(\theta)$ has to be such that the $\hat{V}_{qf}(\theta)$, which results from the derivative of $\hat{V}_{ff}(\theta)$ with respect to θ, is also a consistent estimator of $V_{qf}(\theta)$. This puts further conditions on $\hat{V}_{ff}(\theta)$. If, for example, $\bar{f}_t(\theta)$ is a martingale difference series while $\bar{q}_t(\theta)$ is not, a heteroskedastic autocorrelation consistent covariance matrix estimator for $\hat{V}_{ff}(\theta)$ could just be a White covariance matrix estimator (see White (1980)), but it would not imply a consistent estimator for $V_{qf}(\theta)$ through its derivative with respect to θ. The same consistent covariance matrix estimator thus has to be involved in all elements of $\hat{V}(\theta)$ and Assumption 2(i) has to hold for it. If so, Assumption 2(ii) essentially results from Assumption 2(i). When using the same covariance matrix estimator for all elements of $V(\theta)$, Assumption 2(i) holds under the conditions for consistency of (heteroskedastic autocorrelation consistent) covariance matrix estimators; see, e.g., White (1980), Newey and West (1987). It is worth noting that for the purpose of conducting the double robust Lagrange multiplier test, the covariance estimator $\hat{V}(\theta)$ is only calculated under the hypothesized pseudo-true value, not the CUE. Thus, it is quite straightforward to have the consistency of $\hat{V}(\theta)$, as shown by Lemma 11, where the low-level regularity conditions under which Assumption 2(i) holds in the leading i.i.d. case are provided.

Lemma 12. For a given data set of realized values for linear moment equations, the sum of $f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)$ and $\text{vec}(\hat{D}(\theta))'\hat{V}_{\theta\theta}(\theta)^{-1}\text{vec}(\hat{D}(\theta))$ is a constant function of θ. In addition, the derivative of DRLM(θ) with respect to θ can be written as follows.

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

$$\begin{aligned}&\tfrac{1}{2}\tfrac{\partial}{\partial\theta}DRLM(\theta)=T\times\left(\tfrac{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)}{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)+\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)}\right)\times\\&\quad\Big\{\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)-f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)-\\&\quad 2f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)+2f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)\times\\&\quad\tfrac{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)+\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)}{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)+\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)}\Big\}.\end{aligned}\tag{20}$$

b. 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 $\hat{V}_{ff}(\theta)=\hat{v}_{ff}(\theta)\hat{V}$, $\hat{V}_{qf}(\theta)=\hat{v}_{qf}(\theta)\hat{V}$ and $\hat{V}_{\theta\theta}(\theta)=\hat{v}_{\theta\theta}(\theta)\hat{V}$, with $\hat{v}_{ff}(\theta)$, $\hat{v}_{qf}(\theta)$, $\hat{v}_{\theta\theta}(\theta)$ scalar functions of θ and $\hat{V}$ a $k_f\times k_f$ dimensional covariance matrix estimator, and the derivative of DRLM(θ) becomes:

$$\begin{aligned}&\tfrac{1}{2}\tfrac{\partial}{\partial\theta}DRLM(\theta)=\tfrac{\left(\hat{V}_{ff}(\theta)^{-\frac{1}{2}}f_T(\theta,X)\right)'\left(\hat{V}_{\theta\theta}(\theta)^{-\frac{1}{2}}\hat{D}(\theta)\right)}{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)+\hat{D}(\theta)'\hat{V}_{\theta\theta}(\theta)^{-1}\hat{D}(\theta)}\times\\&\quad\Big\{T\times\hat{D}(\theta)'\hat{V}_{\theta\theta}(\theta)^{-1}\hat{D}(\theta)-T\times f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)\Big\}\times\left(\tfrac{\hat{v}_{\theta\theta}(\theta)}{\hat{v}_{ff}(\theta)}\right)^{\frac{1}{2}}.\end{aligned}\tag{21}$$

Proof: Starting out from a linear moment equation with $\mu_f(\theta) = \mu_f(0) + J(0)\theta$, so $f_T(\theta, X) = \widehat{\mu}_f(0) + \widehat{J}(0)\theta$:

$$\begin{aligned} d &= \begin{pmatrix} \widehat{\mu}_f(0) \\ \text{vec}(\widehat{J}(0)) \end{pmatrix}' \widehat{\text{var}}\left(\sqrt{T}\begin{pmatrix} \widehat{\mu}_f(0) \\ \text{vec}(\widehat{J}(0)) \end{pmatrix}\right)^{-1} \begin{pmatrix} \widehat{\mu}_f(0) \\ \text{vec}(\widehat{J}(0)) \end{pmatrix} \\ &= \begin{pmatrix} \widehat{\mu}_f(0) + \widehat{J}(0)\theta \\ \text{vec}(\widehat{J}(0)) \end{pmatrix}' \widehat{\text{var}}\left(\sqrt{T}\begin{pmatrix} \widehat{\mu}_f(0) + \widehat{J}(0)\theta \\ \text{vec}(\widehat{J}(0)) \end{pmatrix}\right)^{-1} \begin{pmatrix} \widehat{\mu}_f(0) + \widehat{J}(0)\theta \\ \text{vec}(\widehat{J}(0)) \end{pmatrix} \\ &= \begin{pmatrix} f_T(\theta, X) \\ \text{vec}(\hat{D}(\theta)) \end{pmatrix}' \widehat{\text{var}}\left(\sqrt{T}\begin{pmatrix} f_T(\theta, X) \\ \text{vec}(\hat{D}(\theta)) \end{pmatrix}\right)^{-1} \begin{pmatrix} f_T(\theta, X) \\ \text{vec}(\hat{D}(\theta)) \end{pmatrix} \\ &= f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X) + \text{vec}(\hat{D}(\theta))'\hat{V}_{\theta\theta}(\theta)^{-1}\text{vec}(\hat{D}(\theta)), \end{aligned}$$

which shows that the sum of $f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X)$ and $\text{vec}(\hat{D}(\theta))'\hat{V}_{\theta\theta}(\theta)^{-1}\text{vec}(\hat{D}(\theta))$ does not depend on θ, since d does not depend on θ given a realized data set.

a. Given the specifications of the derivatives in Lemma 2, the derivative of DRLM(θ) when $m = 1$ and $f_T(\theta, X)$ is linear in θ reads:

$$\begin{aligned} &\tfrac{1}{2}\tfrac{\partial}{\partial\theta} DRLM(\theta) \\ = \; &\tfrac{1}{2}T\tfrac{\partial}{\partial\theta}\Big\{ f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)\Big[f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X) + \\ &\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)\Big]^{-1}\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X)\Big\} \\ = \; &T\Big[f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X) + \hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)\Big]^{-1} f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) \\ &\left(\tfrac{\partial}{\partial\theta}\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X)\right) - \tfrac{1}{2}T\left(\tfrac{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)}{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X) + \hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)}\right)^2 \\ &\left(\tfrac{\partial}{\partial\theta}\Big[f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X) + \hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)\Big]\right) \\ = \; &T\left(\tfrac{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)}{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X) + \hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)}\right)\Big\{\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) - \\ &2 f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) - f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1} f_T(\theta, X) + \\ &2\tfrac{\left[f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X) + \hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)\right]}{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X) + \hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)} f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)\Big\}. \end{aligned}$$

b. In case of i.i.d. data, $m = 1$, and $f_T(\theta, X)$ linear in θ, $\hat{V}(\theta)$ has a Kronecker product structure so $\hat{V}_{ff}(\theta) = \hat{v}_{ff}(\theta)\hat{V}$, $\hat{V}_{qf}(\theta) = \hat{v}_{qf}(\theta)\hat{V}$ and $\hat{V}_{\theta\theta}(\theta) = \hat{v}_{\theta\theta}(\theta)\hat{V}$, with $\hat{v}_{ff}(\theta)$, $\hat{v}_{qf}(\theta)$, $\hat{v}_{\theta\theta}(\theta)$ scalar and $\hat{V}$ a $k_f \times k_f$ matrix, the ratio in the last line of the above expression simplifies to $\frac{\hat{v}_{qf}(\theta)}{\hat{v}_{ff}(\theta)}$ so:

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

B. Proof of Theorem 1

We can specify the limit behavior of $Ts(\theta^*)$ as:

$$Ts(\theta^*)' \underset{d}{\to} \quad a+b+c,$$

with $a = \Psi_\theta(\theta^*)'V_{ff}(\theta^*)^{-1}\tilde{\mu}_f(\theta^*)$, $b = \tilde{D}(\theta^*)'V_{ff}(\theta^*)^{-1}\psi_f(\theta^*)$ and $c = \Psi_\theta(\theta^*)'V_{ff}(\theta^*)^{-1}\psi_f(\theta^*)$. To obtain the bound on the limiting distribution of the DRLM statistic, we next further characterize the limit behavior of the above components. We first do so for $m=1$.

m=1: We specify a, b and c as:

$$\begin{aligned} a =& \quad \Psi_\theta^{*\prime}G'\mu^*, \\ b =& \quad D^{*\prime}G'\psi_f^*, \\ c =& \quad \Psi_\theta^{*\prime}G'\psi_f^*, \end{aligned}$$

which results from a singular value decomposition of $V_{ff}(\theta^*)^{-\frac{1}{2}}V_{\theta\theta}(\theta^*)^{\frac{1}{2}}$:

$$V_{ff}(\theta^*)^{-\frac{1}{2}}V_{\theta\theta}(\theta^*)^{\frac{1}{2}} = LGK',$$

with L and K $k_f \times k_f$ dimensional orthonormal matrices and G a diagonal $k_f \times k_f$ dimensional matrix with the non-negative singular values in decreasing order on the main diagonal and we used that $\mu^* = L'V_{ff}(\theta^*)^{-\frac{1}{2}}\tilde{\mu}_f(\theta^*)$, $D^* = K'V_{\theta\theta}(\theta^*)^{-\frac{1}{2}}\tilde{D}(\theta^*)$, $\psi_f^* = L'V_{ff}(\theta^*)^{-\frac{1}{2}}\psi_f(\theta^*) \sim N(0, I_{k_f})$, $\Psi_\theta^* = K'V_{\theta\theta}(\theta^*)^{-\frac{1}{2}}\Psi_\theta(\theta^*) \sim N(0, I_{k_f})$ and independent of ψ_f^*.

Using the above, the limit behavior of the DRLM statistic can be specified as:

$$\begin{aligned} DRLM(\theta^*) \underset{d}{\to}& \quad \left[\Psi_\theta^{*\prime}G'\mu^* + D^{*\prime}G'\psi_f^* + \Psi_\theta^{*\prime}G'\psi_f^*\right]' \\ & \quad \left[\left(\mu^*+\psi_f^*\right)'GG'\left(\mu^*+\psi_f^*\right) + \left(D^*+\Psi_\theta^*\right)'G'G\left(D^*+\Psi_\theta^*\right)\right]^{-1} \\ & \quad \left[\Psi_\theta^{*\prime}G'\mu^* + D^{*\prime}G'\psi_f^* + \Psi_\theta^{*\prime}G'\psi_f^*\right] \\ =& \quad \frac{\left[\Psi_\theta^{*\prime}G\mu^* + D^{*\prime}G\psi_f^* + \Psi_\theta^{*\prime}G\psi_f^*\right]^2}{\left[\left(\mu^*+\psi_f^*\right)'G^2\left(\mu^*+\psi_f^*\right) + \left(D^*+\Psi_\theta^*\right)'G^2\left(D^*+\Psi_\theta^*\right)\right]}. \end{aligned}$$

The limiting distribution of the DRLM statistic only depends on the $3k_f$ parameters present in k_f, G, μ^* and D^*. The $3k_f$ results since the limiting distribution is invariant to multiplying G by a positive scalar so the largest element of G, G_{11}, can be set to one. This implies that G contains $k_f - 1$ non-negative elements which are not preset to 0 or 1. The number of elements in both μ^* and D^* equals k_f.

When μ^* and D^* equal zero, the limit behavior of $DRLM(\theta^*)$ becomes:

$$DRLM(\theta^*)|_{\mu^*=D^*=0} \underset{d}{\rightarrow} \quad \frac{\left[\Psi_\theta^{*\prime} G' \psi_f^*\right]^2}{\left[\psi_f^{*\prime} G^2 \psi_f^* + \Psi_\theta^{*\prime} G^2 \Psi_\theta^*\right]} \preceq \chi^2(1),$$

since both $\frac{\left[\Psi_\theta^{*\prime} G' \psi_f^*\right]^2}{\psi_f^{*\prime} G^2 \psi_f^*} \sim \chi^2(1)$ and $\frac{\left[\Psi_\theta^{*\prime} G' \psi_f^*\right]^2}{\Psi_\theta^{*\prime} G^2 \Psi_\theta^*} \sim \chi^2(1)$ and "$\preceq$" indicates stochastically dominated, i.e., for a continuous non-negative scalar random variable $u \preceq \chi^2(m) : \Pr\left[u > cv_{\chi^2(m)}(\alpha)\right] \leq \alpha$, for $\alpha \in (0,1]$ and with $cv_{\chi^2(m)}(\alpha)$ the $(1-\alpha) \times 100\%$ critical value for the $\chi^2(m)$ distribution.

Similarly, when the length of μ^* and/or D^* goes to infinity:

$$\left.\begin{array}{r} \lim_{\mu^{*\prime}\mu^* \to \infty} DRLM(\theta^*) \\ \lim_{D^{*\prime}D^* \to \infty} DRLM(\theta^*) \\ \lim_{\mu^{*\prime}\mu^* \to \infty, D^{*\prime}D^* \to \infty} DRLM(\theta^*) \end{array}\right\} \quad \underset{d}{\rightarrow} \chi^2(1).$$

The limit behavior is identical with respect to the different elements of μ^* and D^*. Figure A1 shows for a pre-specified fixed value of G that the distribution function associated with the limit behavior of $DRLM(\theta^*)$ is a non-increasing function of either the length of μ^* or D^*. Figure A1 also shows the difference with the $\chi^2(1)$ distribution function which makes it clear that the $\chi^2(1)$ distribution dominates the limiting distribution of the DRLM statistic for this specific value of G. Since G is a diagonal matrix with only non-negative elements, this behavior holds also for all other values of G so the limit behavior of $DRLM(\theta^*)$ is bounded by the $\chi^2(1)$ distribution:

$$\lim_{T \to \infty} \Pr\left[DRLM(\theta^*) > cv_{\chi^2(1)}(\alpha)\right] \leq \alpha.$$

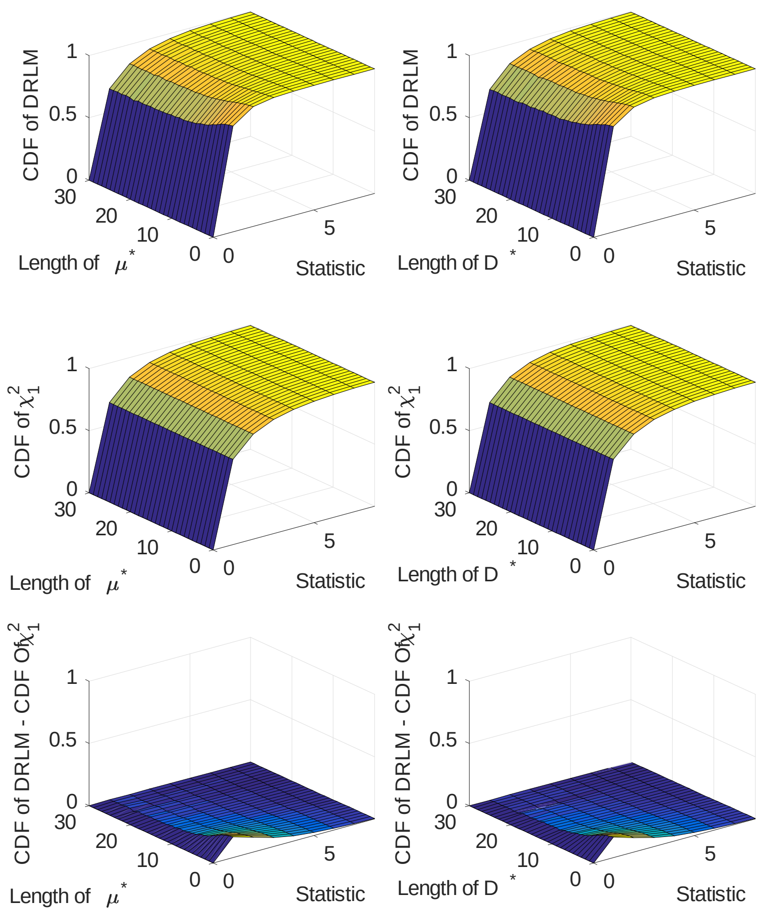
m > 1: We specify a, b and c as:

$$\begin{aligned} a =& \quad \Psi_\theta^{*\prime} \mu^*, \\ b =& \quad D^{*\prime} \psi_f^*, \\ c =& \quad \Psi_\theta^{*\prime} \psi_f^*, \end{aligned}$$

with $G = (I_m \otimes V_{ff}(\theta^*)^{-\frac{1}{2}}) V_{\theta\theta}(\theta^*)^{\frac{1}{2}}$, $\Psi_\theta^* = V_{ff}(\theta^*)^{-\frac{1}{2}} \Psi_\theta(\theta^*)$, $\text{vec}(\Psi_\theta^*) = G\psi_\theta^*$, $\psi_\theta^* \sim N(0, I_{k_f m})$,

$\psi_f^* = V_{ff}(\theta^*)^{-\frac{1}{2}}\psi_f(\theta^*) \sim N(0, I_{k_f})$ and independent of ψ_θ^*, $\mu^* = V_{ff}(\theta^*)^{-\frac{1}{2}}\tilde{\mu}_f(\theta^*)$, $D^* = V_{ff}(\theta^*)^{-\frac{1}{2}}\tilde{D}(\theta^*)$, $\text{vec}(D^*) = G\text{vec}(\bar{D}^*)$ and $\bar{D}^*$ is a $k_f \times m$ dimensional matrix.

Figure A1: Distribution function of the DRLM statistic for a fixed value of G as a function of the length of either μ^* or D^*.



Using the above, the limit behavior of the DRLM statistic can be specified as:

$$
\begin{aligned}
DRLM(\theta^*) \underset{d}{\rightarrow} \quad & \left[\Psi_\theta^{*\prime}\mu^* + D^{*\prime}\psi_f^* + \Psi_\theta^{*\prime}\psi_f^*\right]' \\
& \left[\left(I_m \otimes \left(\mu^* + \psi_f^*\right)\right)' GG' \left(I_m \otimes \left(\mu^* + \psi_f^*\right)\right) + \left(D^* + \Psi_\theta^*\right)' \left(D^* + \Psi_\theta^*\right)\right]^{-1} \\
& \left[\Psi_\theta^{*\prime}\mu^* + D^{*\prime}\psi_f^* + \Psi_\theta^{*\prime}\psi_f^*\right].
\end{aligned}
$$

The limiting distribution of the DRLM statistic depends on the $k_f^2 m^2 + k_f m + k_f + 1$ parameters present in: G, D^*, μ^*, k_f and m. Since the limiting distribution is invariant to multiplying G by a positive scalar, we normalize G such that one diagonal element of G, say G_{11}, is equal to one. This explains the number of parameters affecting the limiting distribution of the DRLM statistic.

When μ^* and D^* equal zero, the limit behavior of $DRLM(\theta^*)$ becomes:

$$DRLM(\theta^*)|_{\mu^*=D^*=0} \underset{d}{\to} \quad \psi_f^{*\prime}\Psi_\theta^*\left[\left(I_m\otimes\psi_f^*\right)' GG'\left(I_m\otimes\psi_f^*\right)+\Psi_\theta^{*\prime}\Psi_\theta^*\right]^{-1}\Psi_\theta^{*\prime}\psi_f^* \preceq \chi^2(m),$$

since $\psi_f^{*\prime}\Psi_\theta^*\left[\Psi_\theta^{*\prime}\Psi_\theta^*\right]^{-1}\Psi_\theta^{*\prime}\psi_f^*\sim\chi^2(m)$ and $\psi_f^{*\prime}\Psi_\theta^*\left[\left(I_m\otimes\psi_f^*\right)' GG'\left(I_m\otimes\psi_f^*\right)\right]^{-1}\Psi_\theta^{*\prime}\psi_f^*\sim\chi^2(m)$.

Similarly, when using a singular value decomposition of D^* :

$$D^* = L_D G_D K_D',$$

with L_D and K_D $k_f\times k_f$ and $m\times m$ dimensional orthonormal matrices and G_D a diagonal $k_f\times m$ dimensional matrix with the non-negative singular values in decreasing order on the main diagonal, we can specify the limit behavior of the DRLM statistic:

$$\begin{aligned}
DRLM(\theta^*) \underset{d}{\to} \quad & \left[K_D\bar\Psi_\theta' L_D'\mu^* + K_D G_D' L_D'\psi_f^* + K_D\bar\Psi_\theta' L_D'\psi_f^*\right]' \\
& \Big[\left(I_m\otimes\left(\mu^*+\psi_f^*\right)\right)' GG'\left(I_m\otimes\left(\mu^*+\psi_f^*\right)\right)+ \\
& \left(L_D G_D K_D' + L_D\bar\Psi_\theta K_D'\right)'\left(L_D G_D K_D' + L_D\bar\Psi_\theta K_D'\right)\Big]^{-1} \\
& \left[K_D\bar\Psi_\theta' L_D'\mu^* + K_D G_D' L_D'\psi_f^* + K_D\bar\Psi_\theta' L_D'\psi_f^*\right] \\
= \quad & \left[\bar\Psi_\theta'\bar\mu + G_D'\bar\psi_f + \bar\Psi_\theta'\bar\psi_f\right]'\Big[\left(I_m\otimes\left(\bar\mu+\bar\psi_f\right)\right)'\bar G\bar G'\left(I_m\otimes\left(\bar\mu+\bar\psi_f\right)\right)+ \\
& \left(G_D+\bar\Psi_\theta\right)'\left(G_D+\bar\Psi_\theta\right)\Big]^{-1}\left[\bar\Psi_\theta'\bar\mu + G_D'\bar\psi_f + \bar\Psi_\theta'\bar\psi_f\right],
\end{aligned}$$

where $\Psi_\theta^* = L_D\bar\Psi_\theta K_D'$, $\bar\psi_f = L_D'\psi_f^*$, $\bar\mu = L_D'\mu^*$, $\bar G = (K_D\otimes L_D V_{ff}(\theta^*)^{-\frac{1}{2}})V_{\theta\theta}(\theta^*)^{\frac{1}{2}}$, $\text{vec}(\bar\Psi_\theta) = \bar G\psi_\theta^*$, $\psi_\theta^*\sim N(0, I_{k_f m})$. The resulting limit behavior is such that when the length of μ^* or the m singular values in G_D go to infinity:

$$\left.\begin{array}{r}
\lim_{\mu^{*\prime}\mu^*\to\infty} DRLM(\theta^*) \\
\lim_{G_{D,ii}^*\to\infty,\ i=1,\ldots,m} DRLM(\theta^*) \\
\lim_{\mu^{*\prime}\mu^*\to\infty, G_{D,ii}^*\to\infty,\ i=1,\ldots,m} DRLM(\theta^*)
\end{array}\right\} \underset{d}{\to} \chi^2(m).$$

Since G is positive semi-definite, it can be verified numerically that for any fixed G, the distribution function associated with the limit behavior of $DRLM(\theta^*)$ is non-increasing when any element of μ^* or $G_{D,ii}^*\to\infty$, $i=1,\ldots,m$ increases. The limit behavior of $DRLM(\theta^*)$ is therefore bounded by the $\chi^2(m)$ distribution:

$$\lim_{T\to\infty} \Pr\left[DRLM(\theta^*) > cv_{\chi^2(m)}(\alpha)\right] \leq \alpha.$$

m > 1 and $V_{ff}(\theta) = v_{ff}(\theta)\bar{V}$, $V_{\theta\theta}(\theta) = \left(\Sigma_{\theta\theta}(\theta) \otimes \bar{V}\right)$, $\Sigma_{\theta\theta}(\theta) : m \times m$ **dimensional matrix:** We specify a, b and c as:

$$\begin{aligned} a = &\quad \Sigma_{\theta\theta}(\theta)^{\frac{1}{2}}\Psi_\theta^{*\prime}\mu^* v_{ff}(\theta)^{-\frac{1}{2}}, \\ b = &\quad \Sigma_{\theta\theta}(\theta)^{\frac{1}{2}}D^{*\prime}\psi_f^* v_{ff}(\theta)^{-\frac{1}{2}}, \\ c = &\quad \Sigma_{\theta\theta}(\theta)^{\frac{1}{2}}\Psi_\theta^{*\prime}\psi_f^* v_{ff}(\theta)^{-\frac{1}{2}}, \end{aligned}$$

with $\mu^* = v_{ff}(\theta^*)^{-\frac{1}{2}}\bar{V}_f^{-\frac{1}{2}}\tilde{\mu}_f(\theta^*)$, $D^* = \bar{V}^{-\frac{1}{2}}\tilde{D}(\theta^*)\Sigma_{\theta\theta}(\theta)^{-\frac{1}{2}}$, $\psi_f^* = v_{ff}(\theta^*)^{-\frac{1}{2}}\bar{V}^{-\frac{1}{2}}\psi_f(\theta^*) \sim N(0, I_{k_f})$, $\Psi_\theta^* = \bar{V}^{-\frac{1}{2}}\Psi_\theta(\theta^*)\Sigma_{\theta\theta}(\theta)^{-\frac{1}{2}} \sim N(0, I_{k_f m})$ and independent of ψ_f^*.

Using the above, the limit behavior of the DRLM statistic can be specified as:

$$\begin{aligned} DRLM(\theta^*) \underset{d}{\rightarrow} &\quad \left[\Psi_\theta^{*\prime}\mu^* + D^{*\prime}\psi_f^* + \Psi_\theta^{*\prime}\psi_f^*\right]' \\ &\quad \left[\left(\mu^* + \psi_f^*\right)'\left(\mu^* + \psi_f^*\right)I_m + \left(D^* + \Psi_\theta^*\right)'\left(D^* + \Psi_\theta^*\right)\right]^{-1} \\ &\quad \left[\Psi_\theta^{*\prime}\mu^* + D^{*\prime}\psi_f^* + \Psi_\theta^{*\prime}\psi_f^*\right]. \end{aligned}$$

The limiting distribution of the DRLM statistic only depends on the $2+k_f+k_f m$ parameters present in k_f, m, μ^* and D^*. To reduce this further, we conduct a singular value decomposition of D^* :

$$D^* = L_{D^*}G_{D^*}K_{D^*}',$$

with L_{D^*} and K_{D^*} $k_f \times k_f$ and $m \times m$ dimensional orthonormal matrices and G_{D^*} a diagonal $k_f \times m$ dimensional matrix with the non-negative singular values in decreasing order on the main diagonal. Using next that $\bar{\Psi}_\theta = L_{D^*}'\Psi_\theta^* K_{D^*} \sim N(0, I_{k_f m})$, $\bar{\mu} = L_{D^*}'\mu^*$ and $\bar{\psi}_f = L_{D^*}'\psi_f^* \sim N(0, I_{k_f})$, we can specify the limit behavior as:

$$\begin{aligned} DRLM(\theta^*) \underset{d}{\rightarrow} &\quad \left[\bar{\Psi}_\theta'\bar{\mu} + G_{D^*}'\bar{\psi}_f + \bar{\Psi}_\theta'\bar{\psi}_f\right]' \\ &\quad \left[\left(\bar{\mu} + \bar{\psi}_f\right)'\left(\bar{\mu} + \bar{\psi}_f\right)I_m + \left(G_{D^*} + \bar{\Psi}_\theta\right)'\left(G_{D^*} + \bar{\Psi}_\theta\right)\right]^{-1} \\ &\quad \left[\bar{\Psi}_\theta'\bar{\mu} + G_{D^*}'\bar{\psi}_f + \bar{\Psi}_\theta'\bar{\psi}_f\right], \end{aligned}$$

which only depends on the m singular values in G_{D^*} and the length of $\bar{\mu}$. The distribution function of the limit behavior is again a non-decreasing function of the length of $\bar{\mu}$ and the m singular values in G_{D^*}, so its limit behavior is bounded by the $\chi^2(m)$ distribution.

Definition of the parameter space In Andrews and Guggenberger (2017), the asymptotic size of the KLM test is proven to equal the nominal size and the accompanying parameter space on the distributions of the observations is stated for both i.i.d. and dependent data settings.

To start out with the i.i.d. setting, define for some κ, $\tau > 0$ and $M < \infty$, the parameter space:

$$\begin{aligned}\mathcal{F} = \; & \big\{F : \{X_t : t \geq 1\} \text{ are i.i.d. under } F,\; E(f_t(\theta^*)) = \mu_f(\theta^*),\; \text{for} \\ & \theta^* = \arg\min_{\theta\in\mathbb{R}^m} \mu_f(\theta)'V_{ff}(\theta)^{-1}\mu_f(\theta),\; V_{ff}(\theta) = E\left(\left(f_t(\theta) - \mu_f(\theta)\right)\left(f_t(\theta) - \mu_f(\theta)\right)'\right) \\ & E\left(\left\|\left(f_t(\theta^*)' \vdots \left(\operatorname{vec}\left(\tfrac{\partial}{\partial\theta'} f_t(\theta^*)\right)\right)'\right)^{2+\kappa}\right\|\right) \leq M \text{ and } \lambda_{\min}(V_{ff}(\theta^*)) \geq \tau\right\},\end{aligned}$$

where $\lambda_{\min}(A)$ is the smallest characteristic root of the matrix A. The parameter space above is identical to the one in Andrews and Guggenberger (2017) Equation (3.3) except that it is defined for the pseudo-true value θ^* defined as the minimizer of the population continuous updating objective function for which $\mu_f(\theta^*)$ is not necessarily equal to zero.

Since we are after proving the size correctness of the DRLM test which tests hypotheses specified on the pseudo-true value θ^*, we define the recentered Jacobian:

$$\begin{aligned}& D(\theta) = J(\theta) - \left[V_{q_1 f}(\theta)V_{ff}(\theta)^{-1}\mu_f(\theta)\ldots V_{q_m f}(\theta)V_{ff}(\theta)^{-1}\mu_f(\theta)\right],\; J(\theta) = \tfrac{\partial}{\partial\theta'}\mu_f(\theta), \\ & V_{\theta_i f}(\theta) = E\left[\left(\tfrac{\partial}{\partial\theta_i}(f_t(\theta) - \mu_f(\theta))\right)\left(f_t(\theta) - \mu_f(\theta)\right)'\right],\; i = 1,\ldots,m, \\ & V_{ff}(\theta) = E\left(\left(f_t(\theta) - \mu_f(\theta)\right)\left(f_t(\theta) - \mu_f(\theta)\right)'\right).\end{aligned}$$

The pseudo-true value is then such that

$$\mu_f(\theta^*)'V_{ff}(\theta^*)^{-1}D(\theta^*) = 0.$$

To guarantee with probability one, a non-singular value of the limit value of the sample analog of $V_{ff}(\theta^*)^{-1}D(\theta^*)$, $\hat{V}_{ff}(\theta^*)^{-1}\hat{D}(\theta^*)$, Andrews and Guggenberger (2017) provide a number of additional conditions on the parameter space $\mathcal{F}$. Since we allow for misspecification, these conditions have to hold when using the recentered Jacobian $D(\theta)$ instead of the Jacobian $J(\theta)$ as in Andrews and Guggenberger (2017). Taken together these conditions imply that the singular values of $V_{ff}(\theta^*)^{-1}D(\theta^*)$ should be bounded away from zero and the same applies for the quadratic form of the orthonormal vectors resulting from the singular value decomposition of $V_{ff}(\theta^*)^{-1}D(\theta^*)$ with respect to the covariance matrix of $\operatorname{vec}(\hat{D}(\theta^*))$. We refer to Andrews and Guggenberger (2017) for the definition of this reduced parameter space.

Parameter space The parameter spaces in Andrews and Guggenberger (2017) imply Lemma 10.2 in their Supplementary Appendix which coincides with our Lemma 5 except that Lemma 5 allows for a population mean function $\mu_f(\theta^*)$ different from zero. Jointly with some weak laws of large numbers, the limiting distributions resulting from Lemma 10.2 in the Supplementary Appendix of Andrews and Guggenberger (2017) provide the building blocks for their Theorem 11.1, which states that the asymptotic size of the KLM test equals the nominal size. Since the parameter spaces also imply our Lemma 5 whose resulting limiting distributions alongside some weak laws of large numbers imply Lemma 6 and Theorem 1, which states that the limiting distribution of the DRLM statistic is bounded by a $\chi^2(m)$ distribution, the parameter spaces thus also imply that the asymptotic size of the DRLM test equals the nominal size.

For the dependent times-series setting, κ, $\tau > 0$, $d > (2+\kappa)/\kappa$ and $M < \infty$, the space of distributions is defined by:

$$\begin{aligned}\mathcal{F}_{ts} = \; & \{F : \{X_t : t = 0, 1, \ldots\} \text{ are stationary and strong mixing under } F \text{ with strong} \\ & \text{mixing numbers } \{\alpha_F(p) : p \geq 1\} \text{ that satisfy } \alpha_F(p) \leq Cp^{-d},\ E(f_t(\theta^*)) = \mu_f(\theta^*), \\ & \theta^* = \arg\min_{\theta \in \mathbb{R}^m} \mu_f(\theta)' V_{ff}(\theta)^{-1} \mu_f(\theta), \\ & V_{ff}(\theta) = E\left[\lim_{T\to\infty} \tfrac{1}{T} \textstyle\sum_{t=1}^{T} \sum_{j=1}^{T} \left(f_t(\theta) - \mu_f(\theta)\right)\left(f_j(\theta) - \mu_f(\theta)\right)'\right], \\ & E\left(\left\| \left(f_t(\theta)' \vdots \left(\operatorname{vec}\left(\tfrac{\partial}{\partial\theta'} f_t(\theta)\right)\right)'\right)^{2+\kappa} \right\|\right) \leq M \text{ and } \lambda_{\min}(V_{ff}(\theta^*)) \geq \tau \Big\}\end{aligned}$$

which again, except for the usage of the pseudo-true value θ^* and a possibly non-zero mean of $f_t(\theta^*)$, is identical to Equation (7.2) in Andrews and Guggenberger (2017). Identical to the i.i.d. setting, Andrews and Guggenberger (2017) provide a number of additional conditions on the parameter space $\mathcal{F}_{ts}$, to guarantee with probability one, a non-singular value of the limit value of the sample analog of $V_{ff}(\theta^*)^{-1}D(\theta^*)$, $\hat{V}_{ff}(\theta^*)^{-1}\hat{D}(\theta^*)$. Replacing the value of the Jacobian, $J(\theta)$, by the recentered Jacobian, $D(\theta)$, in the conditions from Andrews and Guggenberger (2017) then implies that also for our setting the limit value of the $\hat{V}_{ff}(\theta^*)^{-1}\hat{D}(\theta^*)$ is non-singular with probability one. The resulting parameter space then again implies our Lemma 5 from which Theorem 1 follows, so the asymptotic size of the DRLM test coincides with the nominal size.

C. Proof of Theorem 2

We first construct the limit behavior of $\hat{D}(0)$ when the pseudo-true value of θ equals θ^*. We use that under the conditions imposed for this theorem,

$$
\begin{aligned}
D(\theta) = & \quad J(0) - \mu_f(\theta)\tfrac{\theta}{1+\theta^2} \\
D(0) = & \quad J(0)
\end{aligned}
$$

so

$$
\begin{aligned}
\sqrt{T}\left(\hat{D}(0) - D(0)\right) & \underset{d}{\rightarrow} \psi_\theta(0) & \Leftrightarrow \\
\sqrt{T}\left(\hat{D}(0) - D(\theta^*) - \mu_f(\theta^*)\tfrac{\theta^*}{1+(\theta^*)^2}\right) & \underset{d}{\rightarrow} \psi_\theta(0) & \Leftrightarrow \\
\sqrt{T}\hat{\Omega}^{-\frac{1}{2}}\hat{D}(0) & \underset{d}{\rightarrow} \bar{D}(1+(\theta^*)^2)^{-\frac{1}{2}} + \bar{\mu}(1+(\theta^*)^2)^{-\frac{1}{2}}\theta^* + \psi_\theta^*(0), &
\end{aligned}
$$

with $\psi_\theta(0) \sim N(0,\Omega)$, $\tilde{D} = \lim_{T\rightarrow\infty}\sqrt{T}D(\theta^*)$, $\bar{D} = \Omega^{-\frac{1}{2}}\tilde{D}(1+(\theta^*)^2)^{\frac{1}{2}}$, $\tilde{\mu}_f = \lim_{T\rightarrow\infty}\sqrt{T}\mu_f(\theta^*)$, $\bar{\mu} = \Omega^{-\frac{1}{2}}\tilde{\mu}_f(1+(\theta^*)^2)^{-\frac{1}{2}}$, and $\psi_\theta^*(0)$ a standard normal k_f dimensional random vector.

Next, we consider $\hat{\mu}_f(0)$:

$$
\begin{aligned}
\mu_f(\theta) = & \quad \mu_f(0) + J(0)\theta \\
= & \quad \mu_f(0) + D(0)\theta
\end{aligned}
$$

so

$$
\begin{aligned}
\sqrt{T}\left(\widehat{\mu}_f(0) - \mu_f(0)\right) & \underset{d}{\rightarrow} \psi_f(0) & \Leftrightarrow \\
\sqrt{T}\left(\widehat{\mu}_f(0) - \mu_f(\theta^*) + D(0)\theta^*\right) & \underset{d}{\rightarrow} \psi_f(0) & \Leftrightarrow \\
\sqrt{T}\left(\widehat{\mu}_f(0) - \mu_f(\theta^*)\tfrac{1}{1+(\theta^*)^2} + D(\theta^*)\theta^*\right) & \underset{d}{\rightarrow} \psi_f(0) & \Leftrightarrow \\
\sqrt{T}\hat{\Omega}^{-\frac{1}{2}}\widehat{\mu}_f(0) & \underset{d}{\rightarrow} \bar{\mu}(1+(\theta^*)^2)^{-\frac{1}{2}} - \bar{D}(1+(\theta^*)^2)^{-\frac{1}{2}}\theta^* + \psi_f^*(0), &
\end{aligned}
$$

with $\psi_f(0) \sim N(0,\Omega)$ and independent of $\psi_\theta(0)$, and $\psi_f^*(0)$ a standard normal k_f dimensional random vector.

D. Proof of Theorem 3

We prove Theorem 3 in two parts: Part **a.** deals with i.i.d. data and Part **b** is for a general covariance matrix setting.

a. The derivative of the DRLM statistic is:

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

Lemma 12 shows that the denominator of the first component is constant over θ. The numerator of the first component equals ($\frac{1}{2}\times$) the score of the sample CUE objective function. The expression of the derivative of the DRLM statistic thus shows that it equals zero when either the score of the sample CUE objective function equals zero, or the two statistics in the second part of the expression are equal.

The derivative of the DRLM statistic shows that the closed set of non-significant values contains a stationary point of the CUE sample objective function different from the CUE. When considering a line from the CUE to the closed set, the DRLM statistic reaches its maximal value in between the CUE and the closed set. Since the DRLM statistic is significant for values outside the closed set, it is significant at its maximal value. On the line from its maximal value to the closed set, the DRLM statistic is declining. The expression of the derivative of the DRLM statistic shows that this decline results from the increase of the CUE objective function, $Tf_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X)$, towards the stationary point inside the closed set. Since the CUE objective function is only increasing from the maximal value of the DRLM statistic towards the stationary point of the CUE objective function inside the closed set, all values of the CUE objective function inside the closed set exceed its value at the maximizer of the DRLM statistic. Since the latter is significant, we can therefore consider all values inside the closed set to be significant as well without altering the size of the test.

b. For the non-Kronecker case a similar argument applies. The derivative is now:

$$\tfrac{1}{2}\tfrac{\partial}{\partial\theta}DRLM(\theta) = T\left(\tfrac{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)}{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)+\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)}\right)\left\{\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)-\right.$$
$$2f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) - f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X)+$$
$$\left.2\tfrac{\left[f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)+\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)\right]}{f_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)+\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)}f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)\right\}.$$

We first note that

$$\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta) = \left(\hat{V}_{\theta\theta}(\theta)^{-\frac{1}{2}}\hat{D}(\theta)\right)' A'A\left(\hat{V}_{\theta\theta}(\theta)^{-\frac{1}{2}}\hat{D}(\theta)\right)$$

and

$$f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X) = \left(\hat{V}_{ff}(\theta)^{-\frac{1}{2}}f_T(\theta, X)\right)' AA'\left(\hat{V}_{ff}(\theta)^{-\frac{1}{2}}f_T(\theta, X)\right)$$

for $A = \hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{\theta\theta}(\theta)^{\frac{1}{2}}$, so AA' and $A'A$ are both positive definite matrices. An increase of $f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X)$ resulting from a change in θ then also implies an increase in $f_T(\theta, X)'\hat{V}_{ff}(\theta)^{-1}\hat{V}_{\theta\theta}(\theta)\hat{V}_{ff}(\theta)^{-1}f_T(\theta, X)$ and decreases, because of Lemma 12, in both $\hat{D}(\theta)'\hat{V}_{\theta\theta}(\theta)^{-1}\hat{D}(\theta)$ and $\hat{D}(\theta)'\hat{V}_{ff}(\theta)^{-1}\hat{D}(\theta)$. This allows us to extend the argument from the proof of Theorem 3a:

The expression of the derivative of the DRLM statistic shows that the closed set of non-significant values contains a stationary point of the CUE sample objective function different from the CUE. On the line from the CUE to the closed set, the DRLM statistic is maximal and significant outside the closed set. The expression of the derivative of the DRLM statistic shows that its decline from its maximal value to the closed set largely results from the increase of the CUE objective function, $Tf_T(\theta,X)'\hat{V}_{ff}(\theta)^{-1}f_T(\theta,X)$. Since the CUE objective function is only increasing from the maximal value of the DRLM statistic towards the stationary point of the CUE objective function inside the closed set, all values of the CUE objective function inside the closed set exceed its value at the maximizer of the DRLM statistic. Since the latter is significant, we can therefore consider all values inside the closed set to be significant as well without altering the size of the test. Thus, the power improvement rule does not affect the size of the test.

E. Proof of Theorem 4

1. For $\tilde{D}(\theta^*) = \lim_{T\to\infty}\sqrt{T}D(\theta^*)$, $\tilde{\mu}_f(\theta^*) = \lim_{T\to\infty}\sqrt{T}\mu_f(\theta^*)$, the top two elements of

$$\begin{pmatrix} \hat{V}_{ff}(\theta^*)^{-1}\hat{D}(\theta^*) \\ (I_m \otimes \hat{V}_{ff}(\theta^*)^{-1}f_T(\theta^*,X)) \\ -D'_{k_f}\left(\hat{V}_{ff}(\theta^*)^{-1}f_T(\theta^*,X)\otimes\hat{V}_{ff}(\theta^*)^{-1}\hat{D}(\theta^*)\right) \end{pmatrix}$$

are of a larger order of magnitude than the bottom element. Hence, the resulting specification of the DRLM statistic corresponds with the one in Definition 1 plus an $o_p(1)$ term, so Theorem 1 proves that it is size-correct.

2. For $D(\theta^*)$, $\tilde{\mu}_f(\theta^*) = \lim_{T\to\infty}\sqrt{T}\mu_f(\theta^*)$ are both finite and non-negligible, the top element of

$$\begin{pmatrix} \hat{V}_{ff}(\theta^*)^{-1}\hat{D}(\theta^*) \\ (I_m \otimes \hat{V}_{ff}(\theta^*)^{-1}f_T(\theta^*,X)) \\ -D'_{k_f}\left(\hat{V}_{ff}(\theta^*)^{-1}f_T(\theta^*,X)\otimes\hat{V}_{ff}(\theta^*)^{-1}\hat{D}(\theta^*)\right) \end{pmatrix}$$

is of a larger order of magnitude that the remaining elements. The resulting specification of the DRLM statistic therefore corresponds with the one in Definition 1 plus an $o_p(1)$ term for which Theorem 1 proves that it is size-correct.

3. For $\tilde{D}(\theta^*) = \lim_{T\to\infty} \sqrt{T}D(\theta^*)$, $\mu_f(\theta^*)$ both finite and non-negligible, the second element of

$$\begin{pmatrix} \hat{V}_{ff}(\theta^*)^{-1}\hat{D}(\theta^*) \\ (I_m \otimes \hat{V}_{ff}(\theta^*)^{-1} f_T(\theta^*, X)) \\ -D'_{k_f}\left(\hat{V}_{ff}(\theta^*)^{-1} f_T(\theta^*, X) \otimes \hat{V}_{ff}(\theta^*)^{-1}\hat{D}(\theta^*)\right) \end{pmatrix}$$

is of a larger order of magnitude that the remaining elements. The resulting specification of the DRLM statistic therefore corresponds with the one in Definition 1 plus an $o_p(1)$ term for which Theorem 1 proves that it is size-correct.

4. For $D(\theta^*)$, $\mu_f(\theta^*)$ both finite and non-negligible, all three elements of

$$\begin{pmatrix} \hat{V}_{ff}(\theta^*)^{-1}\hat{D}(\theta^*) \\ (I_m \otimes \hat{V}_{ff}(\theta^*)^{-1} f_T(\theta^*, X)) \\ -D'_{k_f}\left(\hat{V}_{ff}(\theta^*)^{-1} f_T(\theta^*, X) \otimes \hat{V}_{ff}(\theta^*)^{-1}\hat{D}(\theta^*)\right) \end{pmatrix}$$

are of the same order of magnitude. Also $f_T(\theta^*, X) \underset{p}{\to} \mu_f(\theta^*)$, $\hat{D}(\theta^*) \underset{p}{\to} D(\theta^*)$ and

$$\sqrt{T}\hat{s}(\theta^*)' \underset{d}{\to} \begin{pmatrix} V_{ff}(\theta^*)^{-1}D(\theta^*) \\ (I_m \otimes V_{ff}(\theta^*)^{-1}\mu_f(\theta^*)) \\ -D'_{k_f}\left(V_{ff}(\theta^*)^{-1}\mu_f(\theta^*) \otimes V_{ff}(\theta^*)^{-1}D(\theta^*)\right) \end{pmatrix}' \begin{pmatrix} \psi_f(\theta^*) \\ \psi_\theta(\theta^*) \\ \psi_{ff}(\theta^*) \end{pmatrix}$$

so combined with Lemma 7, $\widehat{\mathcal{W}}(\theta_0^*) = \hat{W}(\theta_0^*) + \hat{W}_s(\theta_0^*)$ provides a consistent estimator of the covariance matrix. Hence, the limit behavior of the DRLM statistic is $\chi^2(m)$ so the DRLM test is size-correct.

II Additional Results

A. Simulation: Nonlinear GMM

We use a log-normal data generating process to simulate consumption growth and asset returns in accordance with the CRRA moment condition:

$$E\left[\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(\iota_{k_f} + R_{t+1}) - \iota_{k_f}\right] = \mu_f(\delta, \gamma). \tag{22}$$

Let $\triangle c_{t+1} = \ln\left(\frac{C_{t+1}}{C_t}\right)$ and $r_{t+1} = \ln(\iota_{k_f} + R_{t+1})$, which are i.i.d. normally distributed:

$$\begin{bmatrix} \triangle c_{t+1} \\ r_{t+1} \end{bmatrix} \sim NID(\mu, V) \equiv NID\left(\begin{bmatrix} 0 \\ \mu_{2,0} \end{bmatrix}, \begin{bmatrix} V_{cc,0} & V_{cr,0} \\ V_{rc,0} & V_{rr,0} \end{bmatrix}\right),$$

with $\mu_{2,0} = (\mu_{2,1,0} \dots \mu_{2,k_f,0})'$ the mean of r_{t+1}, $V_{cc,0}$ the (scalar) variance of $\triangle c_{t+1}$, $V_{rc,0} = V'_{cr,0} = (V_{rc,1,0} \dots V_{rc,k_f,0})'$ the $k_f \times 1$ dimensional covariance between r_{t+1} and $\triangle c_{t+1}$, and $V_{rr,0} = (V_{rr,ij,0}) : i, j = 1, \dots, k_f$, the $k_f \times k_f$ dimensional covariance matrix of r_{t+1}. This DGP has also been used in Kleibergen and Zhan (2020), where the covariance matrix $V = [V_{cc,0}, V_{cr,0}; V_{rc,0}, V_{rr,0}]$ is calibrated to data. We will change the value of $\mu_{2,0}$ to vary the magnitude of the misspecification through a constant c as detailed below. We will also alter the correlation coefficient of $\triangle c_{t+1}$ and r_{t+1} through multiplying a constant $\tilde{c}$ to vary identification.

We use the data generating process described above to jointly simulate consumption growth and asset returns. When the discount factor δ is fixed at its true value, γ is the single structural parameter of interest; see, for example, Savov (2011) and Kroencke (2017).

Given pre-set values of δ_0, $\mu_{2,0}$, $V_{cc,0}$, $V_{rc,0}$ and $V_{rr,0}$, the CRRA moment equation is such that:

$$\begin{aligned}
&\mu_f(\gamma) \\
&= E\left[\delta_0 \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} (\iota_{k_f} + R_{t+1}) - \iota_{k_f}\right] \\
&= E\left[\begin{pmatrix} \exp\left(\ln(\delta_0) - \gamma \triangle c_{t+1} + r_{t+1,1}\right) \\ \vdots \\ \exp\left(\ln(\delta_0) - \gamma \triangle c_{t+1} + r_{t+1,k_f}\right) \end{pmatrix} - \iota_{k_f}\right] \\
&= \begin{pmatrix} \exp\left(\ln(\delta_0) + \mu_{2,1,0} + \frac{1}{2}\left(V_{rr,11,0} + \gamma^2 V_{cc,0} - 2\gamma V_{rc,1,0}\right)\right) \\ \vdots \\ \exp\left(\ln(\delta_0) + \mu_{2,k_f,0} + \frac{1}{2}\left(V_{rr,k_f k_f,0} + \gamma^2 V_{cc,0} - 2\gamma V_{rc,k_f,0}\right)\right) \end{pmatrix} - \iota_{k_f}.
\end{aligned}$$

We also need the explicit expression of $V_{ff}(\gamma)$:

$$V_{ff}(\gamma) = E\left[(f_t(\gamma) - \mu_f(\gamma))(f_t(\gamma) - \mu_f(\gamma))'\right] = Var\left(e^{\ln(\delta) - \gamma \triangle c_{t+1} + r_{t+1}}\right)$$
$$= \left(\begin{pmatrix} \exp\left(\ln(\delta_0) + \mu_{2,1,0} + \frac{1}{2}\left(V_{rr,11,0} + \gamma^2 V_{cc,0} - 2\gamma V_{rc,1,0}\right)\right) \\ \vdots \\ \exp\left(\ln(\delta_0) + \mu_{2,k_f,0} + \frac{1}{2}\left(V_{rr,k_f k_f,0} + \gamma^2 V_{cc,0} - 2\gamma V_{rc,k_f,0}\right)\right) \end{pmatrix} \right.$$
$$\left. \begin{pmatrix} \exp\left(\ln(\delta_0) + \mu_{2,1,0} + \frac{1}{2}\left(V_{rr,11,0} + \gamma^2 V_{cc,0} - 2\gamma V_{rc,1,0}\right)\right) \\ \vdots \\ \exp\left(\ln(\delta_0) + \mu_{2,k_f,0} + \frac{1}{2}\left(V_{rr,k_f k_f,0} + \gamma^2 V_{cc,0} - 2\gamma V_{rc,k_f,0}\right)\right) \end{pmatrix}' \right) \odot$$
$$\left(\exp\left(\left(-\gamma \iota_{k_f} \vdots I_{k_f}\right) \begin{bmatrix} V_{cc,0} & V_{cr,0} \\ V_{rc,0} & V_{rr,0} \end{bmatrix} \left(-\gamma \iota_{k_f} \vdots I_{k_f}\right)' \right) - \iota_{k_f} \iota_{k_f}' \right),$$

where $\odot$ stands for the element-by-element multiplication.

The population moment function $\mu_f(\gamma)$ and the population covariance matrix $V_{ff}(\gamma)$ provided above are employed to compute the pseudo-true value γ^*:

$$\gamma^* = arg\min\nolimits_{\gamma} \mu_f(\gamma)' V_{ff}(\gamma)^{-1} \mu_f(\gamma). \tag{23}$$

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.[1]

For correctly specified GMM, $\mu_f(\gamma) = 0$ holds so we solve for $\mu_{2,0}$:

$$\mu_{2,0} = -\iota_{k_f} \ln(\delta_0) - \frac{1}{2} \left[\begin{pmatrix} V_{rr,11,0} \\ \vdots \\ V_{rr,k_f k_f,0} \end{pmatrix} + \iota_{k_f} \gamma^2 V_{cc,0} - 2\gamma V_{rc,0} \right]. \tag{24}$$

For the misspecified setting, we use an an auxiliary $\tilde{\mu}_2$ that makes $\mu_f(\gamma) = 0$ and then subtract

[1] We note that in other models of interest, the pseudo-true value does not necessarily change as the magnitude of misspecification varies; see Hansen and Lee (2021) for an IV model where their defined pseudo-true value is invariant to misspecification.

a vector of constants, $c\iota_{k_f}$, to introduce misspecification in the DGP:

$$\mu_{2,0} = \tilde{\mu}_2 - c\iota_{k_f}. \tag{25}$$

The sample moment function and its derivative now only depend on γ :

$$\begin{array}{llll} f_T(\gamma, X) = & \frac{1}{T}\sum_{t=1}^T f_t(\gamma), & f_t(\gamma) = & \delta_0 \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(\iota_{k_f} + R_{t+1}) - \iota_{k_f}, \\ q_T(\gamma, X) = & \frac{1}{T}\sum_{t=1}^T q_t(\gamma), & q_t(\gamma) = & -\delta_0 \ln\left(\frac{C_{t+1}}{C_t}\right)\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(\iota_{k_f} + R_{t+1}). \end{array} \tag{26}$$

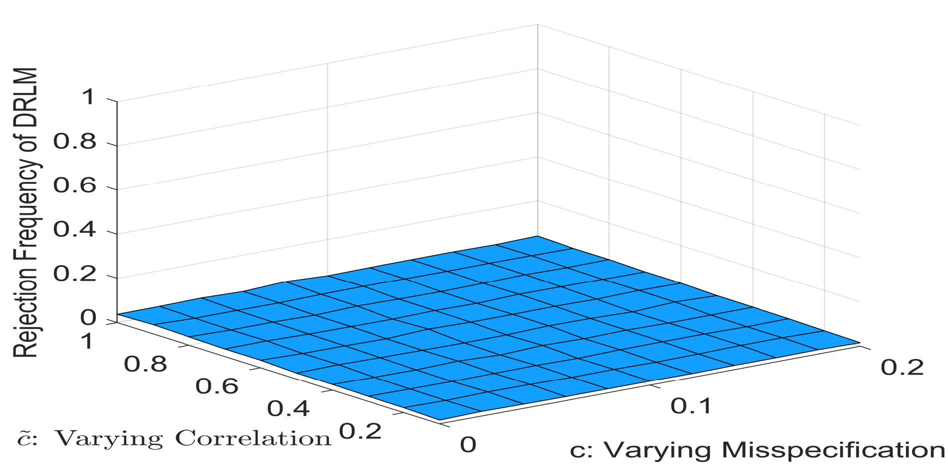
We use the Eicker-White covariance matrix estimators, see White (1980):

$$\begin{array}{ll} \hat{V}_{ff}(\gamma) = & \frac{1}{T}\sum_{t=1}^T (f_t(\gamma) - f_T(\gamma, X))(f_t(\gamma) - f_T(\gamma, X))', \\ \hat{V}_{qf}(\gamma) = & \frac{1}{T}\sum_{t=1}^T (q_t(\gamma) - q_T(\gamma, X))(f_t(\gamma) - f_T(\gamma, X))', \\ \hat{V}_{qq}(\gamma) = & \frac{1}{T}\sum_{t=1}^T (q_t(\gamma) - q_T(\gamma, X))(q_t(\gamma) - q_T(\gamma, X))', \\ \hat{V}_{\theta\theta}(\gamma) = & \hat{V}_{qq}(\gamma) - \hat{V}_{qf}(\gamma)\hat{V}_{ff}(\gamma)^{-1}\hat{V}_{qf}(\gamma)', \end{array} \tag{27}$$

which are employed for the compuation of the DRLM statistic.

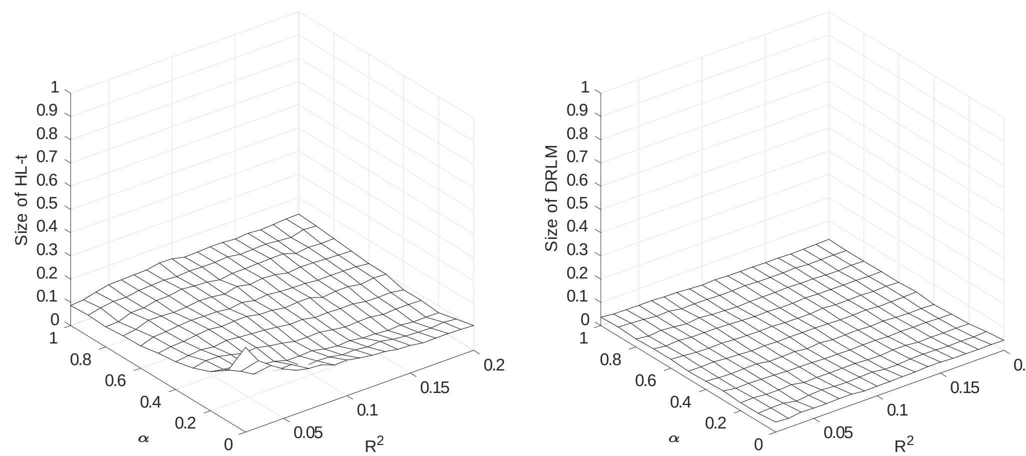
For the simulation studies with multiple parameters, we further jointly test the pseudo-trues of δ and γ and illustrate the size of the DRLM test in Figure A2. Thus, δ is treated as an extra parameter instead of the fixed δ_0 in the aforementioned expressions. Similar to the single parameter case presented in the paper, Figure A2 shows that the rejection frequencies of the DRLM test do not exceed the nominal 5% for varying strengths of identification and misspecification.

Figure A2: Rejection frequencies of 5% significance DRLM tests of $\text{H}_0 : \delta^* = \delta_0^*, \gamma^* = \gamma_0^*$ with $m = 2$, $k_f = 5$ as a function of misspecification c and the strength of identification $\tilde{c}$.



B. Simulation: Linear IV

Figure A3: Rejection frequencies of 5% significance Hansen and Lee (2021) t-test and DRLM of $\mathrm{H}_0 : \theta^* = \theta_0^*$ with $m = 1$, $k_f = 4$ as a function of the strengths of identification R^2, and misspecification α .



We take the data generating process for the linear IV model in Hansen and Lee (2021) to illustrate the proposed DRLM test in Figure A3, where the first-stage R^2 is a measure of the identification strength, and α is a measure of misspecification. We consider a simulation setting as in Hansen and Lee (2021) with $m = 1$, $k_f = 4$, and the sample size is set to 250; for further details of the data generating process, we refer to Hansen and Lee (2021). As expected, Figure A3 shows that the DRLM test remains size-correct regardless of the magnitudes of misspecification and identification while the Hansen and Lee (2021) t-test is over rejecting when identification is weak.

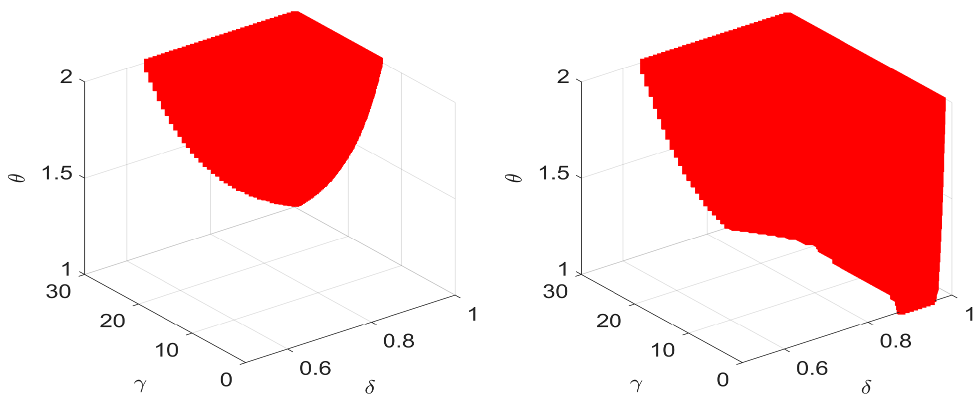
C. Application: Epstein-Zin

We extend our empirical analysis from Figure 12 to a similar setting with Epstein-Zin (1989) preferences in Figure A4, where there are three parameters of interest, δ, γ, and θ, with δ the discount rate, γ the relative rate of risk aversion and θ the elasticity of intertemporal substitution. Panels A4.1 and A4.2 contain the 95% confidence sets that result from the GMM-AR and DRLM tests, respectively.

Similar to Figure 12 in the paper, Panel A4.1 shows that the GMM-AR test rejects plausible values of the parameters, so its 95% confidence set only consists of unrealistically large values of γ.

In contrast, Panel A4.2 shows that the DRLM test yields a joint confidence set that does contain economically meaningful values.

Figure A4: 95% confidence regions of (δ, γ, θ) for Epstein-Zin utility.



Panel A4.1: GMM-AR   Panel A4.2: DRLM

Notes: The three parameters δ, γ, and θ result from the Epstein-Zin preferences, see Epstein and Zin (1989), with $E\left[\delta^{\theta} R_{m,t+1}^{\theta-1}\left(\frac{C_{t+1}}{C_t}\right)^{-\theta\gamma}(\iota_N + R_{t+1}) - \iota_N\right] = \mu_f(\delta, \gamma, \theta)$.

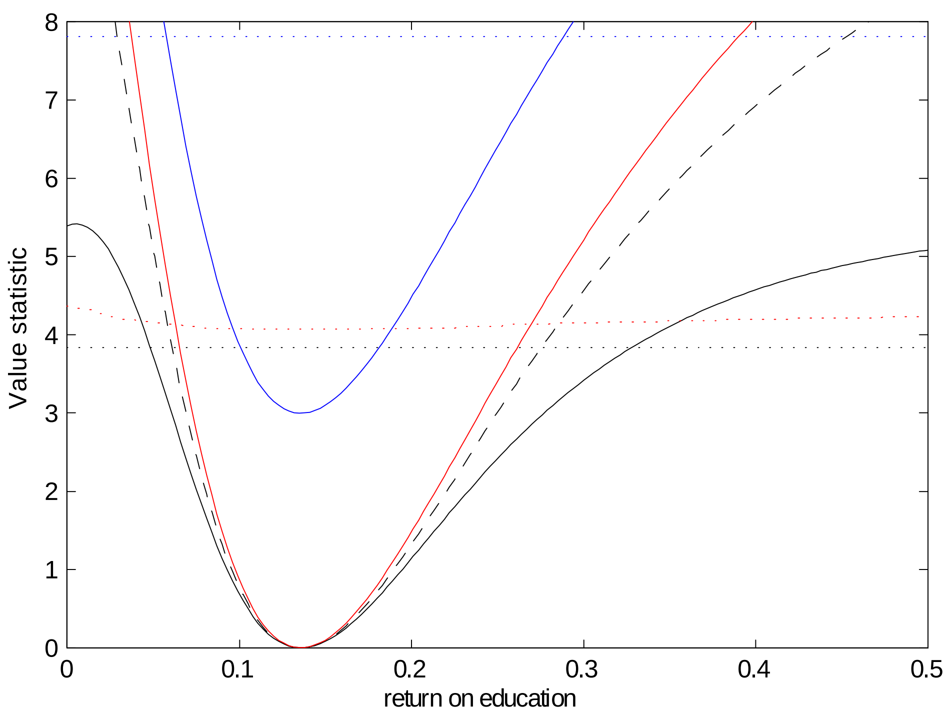
D. Application: Linear IV regression using Card (1995) data

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 linear IV regression model since it considers them to be identical, see Imbens and Angrist (1994).

Figure A5 shows the values of the GMM-AR, LR, KLM and DRLM statistics around the CUE. It also shows their critical value functions at the 5% level. The other area of small values of the DRLM statistic is left out, since it would be discarded by the power enhancement rule. The J-statistic, which equals the minimal value of the GMM-AR statistic, is 2.99 with a p-value of 0.22. The first stage F-statistic is 7.01. This suggests that the return on education is likely weakly identified, see

Stock and Yogo (2005), which then also implies that the J-test does not have much power. Its quite low p-value can thus as well indicate misspecification, which results from distinct local average treatment effects for the different instruments. Since there are three instruments, the IS statistic is about 21 ($\approx 3 \times 7.01$), so the LR no-identification statistic equals 18 ($= IS - J$) and is significant at the 6% level (6% conditional critical value is 17.9, conditioning statistic=32.5) which explains the bounded 95% confidence sets. 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 likely weak identification of the return on education indicated by the small first stage F-statistic and the value of the LR no-identification statistic. 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 A5 then shows that allowing for misspecification further enlarges the identification-robust confidence set for the return on education.

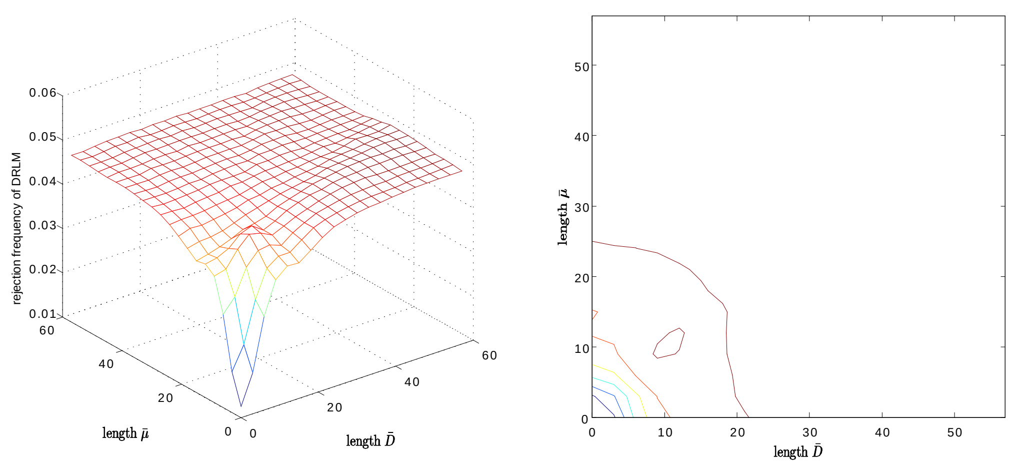
Figure A5: Tests of the return on education using Card (1995) data with the DRLM (solid black), KLM (dashed black), LR (solid red) and GMM-AR (solid blue) statistics and their 95% (conditional) critical value lines (dotted in the color of the test they refer to).



E. DRLM using conditional critical values

Figure 1 in the paper 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 can calibrate a feasible conditional critical value function following Guggenberger et al. (2019), who propose data-dependent conditional critical values to improve the performance of the subvector Anderson and Rubin test. The data-dependent conditional critical values of Guggenberger et al. (2019) adapt to the strength of identification, while for the DRLM test, we consider the conditional critical values based on the maximum of $\bar{\mu}'\bar{\mu}$ and $\bar{D}'\bar{D}$, since we study the joint setting of both misspecification and weak identification. Taking Figure 1 for example, when the maximum of $\bar{\mu}'\bar{\mu}$ and $\bar{D}'\bar{D}$ is less than two-hundred and fifty, we computed a feasible 95% conditional critical value function and used it to generate the corresponding Figure A6 for the size-improved DRLM test.[2] The contour lines in Figure A6 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 A6: Rejection frequency of 5% significance DRLM tests of $\mathrm{H}_0 : \theta^* = \theta_0^*$ using a conditional 95% critical value as a function of the lengths of $\bar{\mu}$ and $\bar{D}$, $m = 1$, $k_f = 25$.



F. Subvector inference

$\theta^* = (\theta_1^* \dots \theta_m^*)'$ with $m \geq 1$. Without loss of generality, consider, e.g., the 95% confidence set of θ_1^* when $m > 1$. We briefly discuss how to use the DRLM test for constructing such a confidence set.

[2]The conditional critical value function we calibrated for Figure A6 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.

One straightforward approach is to use the projection method, see also Dufour and Taamouti (2005). It starts with constructing the joint 95% confidence set of θ^* by inverting the DRLM test, and then projects the joint confidence set of θ^* on the axis of θ_1^*. The resulting confidence set of θ_1^* has at least 95% coverage, so it is valid but conservative.

A second approach is to substitute the CUE for the parameters $\theta_2^* \dots \theta_m^*$ under the hypothesized value $\theta_{1,0}^*$ of θ_1^*. In other words, under $\mathrm{H}_0 : \theta_1^* = \theta_{1,0}^*$, first compute the restricted CUE of $\theta_2^* \dots \theta_m^*$, and then use the hypothesized $\theta_{1,0}^*$ and the restricted CUE of $\theta_2^* \dots \theta_m^*$ for calculating the DRLM statistic. This approach, however, is not necessarily size correct for the DRLM test since Guggenberger et al. (2012) show that it does not control the size of the KLM test. Kleibergen (2021) shows that it does control the size of the conditional likelihood ratio test in the correctly specified linear IV regression model, but this test is not size correct under misspecification.

A third approach is to orthogonalize the parameters using the reparametrization proposed in Han and McCloskey (2019). Since the identification of the parameters involves both the misspecification and the Jacobian, as further discussed in Kleibergen and Zhan (2024), it would not be straightforward to do so. We therefore relegate this to future work.

References

[1] Andrews, D.W.K. and P. Guggenberger. Asymptotic size of Kleibergen's LM and conditional LR tests for moment condition models. *Econometric Theory*, **33**:1046–1080, 2017.

[2] 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)).

[3] Dufour, J.-M. and M. Taamouti. Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica*, **73**:1351–1365, 2005.

[4] Epstein, L. and S. Zin. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. *Econometrica*, 57:937–969, 1989.

[5] 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.

[6] Guggenberger, P., F. Kleibergen, and S. Mavroeidis. A more powerful Anderson Rubin test in linear instrumental variables regression. *Quantitative Economics*, **10**:487–526, 2019.

[7] Han, S. and A. McCloskey. Estimation and Inference with a (Nearly) Singular Jacobian. *Quantitative Economics*, **10**:1019–1068, 2019.

[8] Hansen, B. E. and Lee, S. Inference for iterated gmm under misspecification. *Econometrica*, 89(3):1419–1447, 2021.

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

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

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

[12] Kleibergen, F. Efficient size correct subset inference in homoskedastic linear instrumental variables regression. *Journal of Econometrics*, **221**:78–96, 2021.

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

[14] Kleibergen, F. and Z. Zhan. Testing for identification in potentially misspecified linear GMM. *Unpublished Working paper, University of Amsterdam*, 2024.

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

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

[17] Newey, W. K and West, K. D. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation. *Econometrica*, 55(3):703–708, 1987.

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

[19] Stock, J.H. and M. Yogo. Testing for Weak Instruments in Linear IV Regression. In D.W.K. Andrews and J.H. Stock, editor, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 80–108. Cambridge: Cambridge University Press, 2005. Prepared for Festschrift in honor of Thomas Rothenberg.

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