Appendix to

# Conditional Linear Combination Tests for Weakly Identified Models

Isaiah Andrews

June 8, 2016

This appendix contains asymptotic results and proofs for the paper "Conditional Linear Combination Tests for Weakly Identified Models," by Isaiah Andrews.

## Appendix 1: Asymptotic Properties of CLC Tests

The results of Sections 3-7 of the main text treat the limiting random variables $(g, \Delta g, \gamma)$ as observed and consider the problem of testing $H_0 : m = 0, \mu \in \mathbb{M}$ against $H_1 : m \in \mathcal{M}(\mu)\backslash\{0\}, \mu \in \mathbb{M}$. In this appendix, we show that under mild assumptions our results for the limit problem (2) imply asymptotic results along sequences of models satisfying (1). We first introduce a useful invariance condition for the weight function $a$ and then prove results concerning the asymptotic size and power of CLC tests.

We previously wrote the weight functions $a$ of CLC tests as functions of $D$ alone, since in the limit problem the parameter $\gamma$ is fixed and known. In this appendix, however, it is helpful to instead write $a(D, \gamma)$. Likewise, since the estimator $\hat{\mu}_D$ used in plug-in tests may depend on $\gamma$, we will write it as $\hat{\mu}_D(D, \gamma)$.

### Appendix 1.1 Postmultiplication Invariant Weight Functions

Our weak convergence assumption (1), together with the continuous mapping theorem, implies that $D_T \rightarrow_d D$ for $D$ normally distributed, where we assume that $D$ is full rank almost surely for all $(\theta, \gamma) \in \Theta \times \Gamma$. In many applications such convergence will only hold if we choose an appropriate normalization when defining $\Delta g_T$, which may seem like an obstacle to applying our approach. In the linear IV model for instance, the appropriate definition for $\Delta g_T$ will depend on the strength of identification.

**Example I: Weak IV (Continued)** In Section 2 we assumed that the instruments were weak, with $\pi_T = \frac{c}{\sqrt{T}}$, and showed that $\Delta g_T = \sqrt{T}\hat{\Omega}_{ff}^{-\frac{1}{2}} \frac{\partial}{\partial \beta} f_T(\beta_0)$ converged in distribution. If on the other hand the instruments are strong, $\pi_T = \pi_1$ and $||\pi_1|| > 0$, then $\frac{\partial}{\partial \beta} f_T(\beta) \rightarrow_p E[X_t Z_t] \neq 0$ so $\sqrt{T}\hat{\Omega}_{ff}^{-\frac{1}{2}} \frac{\partial}{\partial \beta} f_T(\beta_0)$ diverges and we should instead take $\Delta g_T = \hat{\Omega}_{ff}^{-\frac{1}{2}} \frac{\partial}{\partial \beta} f_T(\beta_0)$.$\square$

This apparent dependence on normalization is not typically a problem, however, since many CLC tests are invariant to renormalization of $(g_T, \Delta g_T, \hat{\gamma})$. In particular, for $A$ any full rank $p \times p$ matrix consider the transformations

$$h_{\Delta g}(\Delta g_T; A) = \Delta g_T A$$

$$h_\Sigma(\Sigma; A) = \left( \begin{bmatrix} 1 & 0 \\ 0 & A \end{bmatrix} \otimes I_k \right)' \Sigma \left( \begin{bmatrix} 1 & 0 \\ 0 & A \end{bmatrix} \otimes I_k \right)$$

and let $h_\gamma(\gamma; A)$ be the transformation of $\gamma$ such that $\Sigma(h_\gamma(\gamma; A)) = h_\Sigma(\Sigma(\gamma); A)$. Let

$$h(g_T, \Delta g_T, \hat{\gamma}; A) = (g_T, h_{\Delta g}(\Delta g_T; A), h_\gamma(\hat{\gamma}; A)) \tag{21}$$

and note that the statistics $J_T$ and $K_T$ are invariant to this transformation for all full rank matrices $A$, in the sense that their values based on $(g_T, \Delta g_T, \hat{\gamma})$ are the same as those based on $h(g_T, \Delta g_T, \hat{\gamma}; A)$. Thus if we choose a weight function $a(D, \gamma)$ which is invariant, the CLC test $\phi_{a(D_T, \hat{\gamma})}$ will be invariant as well. Formally, we say that the weight function $a(D, \gamma)$ is invariant to postmultiplication if for all full-rank $p \times p$ matrices $A$ we have

$$a(D, \gamma) = a(h_{\Delta g}(D; A), h_\gamma(\gamma; A)),$$

where we have used the fact that $D$ calculated using $h(g, \Delta g, \gamma; A)$ is equal to $h_{\Delta g}(D; A)$.

Invariance to postmultiplication is useful since to obtain results for invariant tests based on $(\tilde{g}_T, \Delta\tilde{g}_T, \tilde{\gamma})$ it suffices that there exist some sequence $A_T$ such that

$$(g_T, \Delta g_T, \hat{\gamma}) = h(\tilde{g}_T, \Delta\tilde{g}_T, \tilde{\gamma}; A_T)$$

satisfies the weak convergence assumption (1), without any need to know the correct sequence $A_T$ for a given application. Thus, in the linear IV example discussed above we can take $\Delta g_T$ as originally defined and make use of results derived under the convergence assumption (6) without knowing identification strength in a given context.

The class of postmultiplication-invariant weight functions $a$ is quite large, and includes all the weight functions discussed above. In particular we can choose the minimax regret weight function $a_{MMR}$ to be invariant to postmultiplication. Likewise, provided we take the estimator $\hat{\mu}_D(D,\gamma)$ to be equivariant under transformation by $h$, so that $h_{\Delta g}\left(\hat{\mu}_D(D,\gamma);A\right)=\hat{\mu}_D\left(h_{\Delta g}\left(D;A\right),h_\gamma\left(\gamma;A\right)\right)$, the plug-in weight function $a_{PI}$ will be invariant as well.

## Appendix 1.2 Asymptotic Size and Power of CLC Tests

Let $F(g,\Delta g,\gamma)$ denote the distribution of $(g,\Delta g,\gamma)$ in the limit problem, noting that the marginal distribution for $\gamma$ in the limit problem is a point mass. Since we have assumed that $D$ is full rank almost surely, $J$ and $K$ are $F$-almost-everywhere continuous functions of $(g,\Delta g,\gamma)$ and the continuous mapping theorem implies

$$(J_T,K_T,D_T)\to_d (J,K,D)\,.$$

To obtain asymptotic size control for the CLC test

$$\phi_{a(D_T,\hat{\gamma})}=1\left\{\left(1-a\left(D_T,\hat{\gamma}\right)\right)\cdot K_T+a\left(D_T,\hat{\gamma}\right)\cdot S_T>c_\alpha\left(a\left(D_T,\hat{\gamma}\right)\right)\right\}$$

all we require is that $a$ be almost-everywhere continuous. Indeed, this test is asymptotically conditionally similar in the sense discussed by Jansson and Moreira (2006).

**Proposition 1** *Assume $(g_T,\Delta g_T,\hat{\gamma})$ satisfies the weak convergence assumption (1) and let $a(D,\gamma)$ be $F\left(g,\Delta g,\gamma\right)$-almost-everywhere continuous for $(\theta_0,\gamma)\in\{\theta_0\}\times\Gamma$. Then under $(\theta_0,\gamma)$ we have that*

$$\lim_{T\to\infty}E_{T,(\theta_0,\gamma)}\left[\phi_{a(D_T,\hat{\gamma})}\right]=\alpha. \tag{22}$$

*Moreover, for $\mathcal{F}$ the set of bounded functions $f(D)$ which are $F\left(g,\Delta g,\gamma\right)$-almost-everywhere continuous under $(\theta_0,\gamma)$,*

$$\lim_{T\to\infty}E_{T,(\theta_0,\gamma)}\left[\left(\phi_{a(D_T,\hat{\gamma})}-\alpha\right)f\left(D_T\right)\right]=0\ \forall f\in\mathcal{F}. \tag{23}$$

It is important to note that Proposition 1 only establishes sequential size control, and depending on the underlying model establishing uniform size control over some base parameter space may require substantial further restrictions. In Example I, however, we can use results from D. Andrews et al. (2011, henceforth ACG) to prove that a large class of CLC tests based on postmultiplication-invariant weight functions control size uniformly in heteroskedastic linear IV with a single endogenous regressor. Unfortunately, however, matters are less clear in the case with multiple endogenous regressors. In that context, D. Andrews and Guggenberger (2014) show that while $K$ tests have uniformly correct asymptotic size over a large parameter space, the asymptotic size of QCLR tests depends on the construction of the weighting function $r(D)$. Correspondingly, only a subset of conditional linear combination tests will have correct asymptotic size in that context.

**Example I: Weak IV (Continued)** Define $\hat{\Omega}$ and $\hat{\Sigma}$ in the usual way (detailed in the proof of Proposition 2 below). Define a parameter space $\Lambda$ of null distributions as in ACG Section 3, noting that $\gamma$ consists of the elements of $(\Omega_F, \Gamma_F, \Sigma_F)$ in the notation of ACG. Building on results in ACG it is straightforward to prove the following proposition:

**Proposition 2** *Consider the CLC test $\phi_{a(D_T,\hat{\gamma})}$ based on a postmultiplication-invariant weight function $a(D,\gamma)$ which is continuous in $D$ and $\gamma$ at all points with $||D|| > 0$ and satisfies*

$$\lim_{\delta\to 0}\left(\sup_{(D,\gamma):||D||>\varepsilon, maxeig(\Sigma_D)\leq\delta} a(D,\gamma)\right) = \lim_{\delta\to 0}\left(\inf_{(D,\gamma):||D||>\varepsilon, maxeig(\Sigma_D)\leq\delta} a(D,\gamma)\right) = a_0 \tag{24}$$

*for some constant $a_0 \in [0,1]$, $maxeig(A)$ the maximal eigenvalue of $A$, and all $\varepsilon > 0$. The test $\phi_{a(D_T,\hat{\gamma})}$ is uniformly asymptotically similar on $\Lambda$:*

$$\lim_{T\to\infty}\inf_{\lambda\in\Lambda} E_{T,\lambda}\left[\phi_{a(D_T,\hat{\gamma})}\right] = \lim_{T\to\infty}\sup_{\lambda\in\Lambda} E_{T,\lambda}\left[\phi_{a(D_T,\hat{\gamma})}\right] = \alpha.$$

The assumption (24), together with the assumed postmultiplication invariance of $a(D,\gamma)$ and the restrictions on the parameter space $\Lambda$, ensures that under sequences with $\sqrt{T}||\pi_T|| \to \infty$ we have that $a(D_T,\hat{\gamma}) \to_p a_0$ asymptotically, and hence that under all strongly identified sequences the test converges to the linear combination test $\phi_{a_0}$. We show in the next section that for $a_0 = 0$ this condition plays an important role in establishing asymptotic efficiency of CLC tests in linear IV under strong identification, and will verify this condition for PI tests $\phi_{PI}$ in linear IV. The conditions needed to ensure that $a_{PI}$ satisfies the continuity conditions in Proposition 2 are much less clear, but we can always create a sufficiently continuous weight function $\tilde{a}$ which approximates $a_{PI}$ arbitrarily well by calculating $a_{PI}$ on a grid of values for $(D, \gamma)$ and taking $\tilde{a}$ to continuously interpolate between these values.[18]$\square$

Power results in the limit problem (2) also imply asymptotic power results under (1). In particular, for $a(D, \gamma)$ almost everywhere continuous with respect to $F(g, \Delta g, \gamma)$, the asymptotic power of $\phi_{a(D_T, \hat{\gamma})}$ is simply the power of $\phi_{a(D,\gamma)}$ in the limit problem.

**Proposition 3** *Assume $(g_T, \Delta g_T, \hat{\gamma})$ satisfies the weak convergence assumption (1) and let $a(D, \gamma)$ be $F(g, \Delta g, \gamma)$-almost-everywhere continuous for some $(\theta, \gamma) \in \Theta \times \Gamma$. Then under $(\theta, \gamma)$*

$$\lim_{T \to \infty} E_{T,(\theta,\gamma)} \left[\phi_{a(D_T,\hat{\gamma})}\right] = E_{m,\mu_D,\gamma} \left[\phi_{a(D,\gamma)}\right]$$

*where $m = m(\theta, \theta_0, \gamma)$ and $\mu_D$ are the parameters in the limit problem.*

Thus, under mild continuity conditions on $a(D, \gamma)$, the asymptotic size and power of tests under (1) are just their size and power in the limit problem. Moreover, sufficiently continuous postmultiplication invariant weight functions $a(D, \gamma)$ which select a fixed weight $a_0$ under strong identification yield uniformly asymptotically similar tests in heteroskedastic linear IV.

## Appendix 1.3 Asymptotic Efficiency Under Strong Identification

The power results above concern the asymptotic properties of CLC tests under general conditions that allow for weak identification, but since the commonly-used non-robust tests are efficient under strong identification we may particularly want to ensure that our CLC tests share this property.

As noted in Section 3, under strong identification we typically have that $\Sigma_{\theta\theta} = 0$, $\Sigma_{\theta g} = 0$, that $\mu$ is full rank, and that $\mathcal{M}(\mu) = \{\mu \cdot c : c \in \mathbb{R}^p\}$. We say that $(g_T, \Delta g_T, \hat{\gamma})$ converges to a Gaussian shift model under $(\theta, \gamma)$ if $(g_T, \Delta g_T, \hat{\gamma}) \to_d (g, \Delta g, \gamma)$ for

$$\begin{pmatrix} g \\ vec(\Delta g) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \cdot b \\ vec(\mu) \end{pmatrix}, \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}\right) \tag{25}$$

[18] To ensure that $\tilde{a}$ is invariant to postmultiplication we can fix $||D|| = 1$ in the grid used to calculate $\tilde{a}$ and evaluate $\tilde{a}$ for other values by rescaling the problem to $||D|| = 1$ using the transformation (21).

where $\mu$ is full rank and $b \in \mathbb{R}^p$. Under strong identification, general GMM models parametrized in terms of local alternatives converge to Gaussian shift models. In many cases strong identification is not necessary to obtain convergence to (25), however, and sequences of models between the polar cases of weak and strong identification, like the "semi-strong" case discussed in D. Andrews and Cheng (2012), often yield Gaussian shift limit problems under appropriately defined sequences of local alternatives.

**Example I: Weak IV (Continued)** Suppose that $\pi_T = r_T c$ for $c \in \mathbb{R}^p$ with $||c|| > 0$ for any sequence $\{r_T\}_{T=1}^{\infty}$ such that $r_T \to r$ as $T \to \infty$ and $\sqrt{T} r_T \to \infty$. For $0 < r < \infty$ this is the usual, strongly identified case, while for $r = 0$ this is falls into the "semi-strong" category of D. Andrews and Cheng (2012): the first stage converges to zero, but at a sufficiently slow rate that many standard asymptotic results are preserved. Let $\tilde{\Omega}$ be a consistent estimator for $\lim_{T\to\infty} Var\left(\left(\sqrt{T} f_T\left(\beta_0\right)', r_T^{-1} f_T\left(\beta_0\right)'\right)'\right)$ and define $\tilde{g}_T(\beta) = \sqrt{T}\tilde{\Omega}_{ff}^{-\frac{1}{2}} f_T(\beta)$ and $\tilde{\gamma} = vec\left(\tilde{\Omega}\right)$ as before. Consider sequences of local alternatives with $\beta_T = \beta_0 + \frac{b^*}{r_T\sqrt{T}}$ and let $\Delta\tilde{g}_T = r_T^{-1}\tilde{\Omega}_{ff}^{-\frac{1}{2}} \frac{\partial}{\partial\beta} f_T(\beta)$. As $T \to \infty$, $(\tilde{g}_T, \Delta\tilde{g}_T, \tilde{\gamma})$ converges to the Gaussian shift limit problem (25) with $\mu = E\left[Z_t Z_t'\right] c$ and $b = b^*$.$\square$

In the Gaussian shift limit problem (25), the Neyman Pearson Lemma implies that the uniformly most powerful level $\alpha$ test based on $(J, K, D)$ is $\phi_K$ as defined in (16). Further, under the weak convergence assumption (1) for $(g, \Delta g)$ as in (25) the test $\phi_{K_T} = 1\left\{K_T > \chi^2_{p,1-\alpha}\right\}$ is asymptotically efficient in the sense of Mueller (2011) for a family of elliptically-contoured weight functions.[19] Under strong identification $\phi_{K_T} = 1\left\{K_T > \chi^2_{p,1-\alpha}\right\}$ is also generally equivalent to the usual Wald tests, though we will need conditions beyond (1) to establish this. It is straightforward to show that a CLC test based on the weight function $a\left(D, \gamma\right)$ will share these properties, and so be asymptotically efficient under sequences converging to (25), if and only if $a\left(D_T, \hat{\gamma}\right) \to_p 0$ under such sequences.

**Proposition 4** *Denote by $\mathcal{A}_c$ the class of weight functions functions $a(D, \gamma)$ that are continuous in both $D$ and $\gamma$ for all full-rank $D$. Fix $(\theta, \gamma) \in \Theta \times \Gamma$ with $\theta \neq \theta_0$ and*

[19] Formally, in the limit problem $\mu = \Delta g$ is known so to derive weighted average power optimal tests we need only consider weights on $b$. For any weights $G\left(b\right)$ with density $g\left(b\right)$ that depends on $b$ only through $\|\mu b\|$, so that $g\left(b\right) \propto \tilde{g}\left(\|\mu b\|\right)$, $\phi_K$ is weighted average power maximizing in the limit problem and by Mueller (2011) $\phi_{K_T}$ is asymptotically optimal over the class of tests with correct size under (1) and (25).

*suppose that* $(g_T, \Delta g_T, \hat{\gamma})$ *converges weakly to the Gaussian shift limit problem (25) with* $b \neq 0$*. For* $a(D,\gamma)$ *almost-everywhere continuous with respect to the limiting measure* $F(g,\Delta g,\gamma)$ *under* $(\theta,\gamma)$*,*

$$\lim_{T\to\infty} E_{T,(\theta,\gamma)}\left[\phi_{a(D_T,\hat{\gamma})}\right] = \sup_{\tilde{a}\in\mathcal{A}_c} \lim_{T\to\infty} E_{T,(\theta,\gamma)}\left[\phi_{\tilde{a}(D_T,\hat{\gamma})}\right]$$

*if and only if* $a(D,\gamma) = 0$ *almost surely with respect to* $F(g,\Delta g,\gamma)$*. Thus*

$$\lim_{T\to\infty} E_{T,(\theta,\gamma)}\left[\phi_{K_T}\right] = \sup_{\tilde{a}\in\mathcal{A}_c} \lim_{T\to\infty} E_{T,(\theta,\gamma)}\left[\phi_{\tilde{a}(D_T,\hat{\gamma})}\right].$$

Using this proposition, it is easy to see that the condition (24) that we used to ensure uniformly correct size for CLC tests in linear IV Example I will also ensure asymptotic efficiency under strong and semi-strong identification provided $a_0 = 0$.

It is straightforward to give conditions under which MMRU tests select $a(\mu_D,\gamma) = 0$ asymptotically in sequences of models converging to Gaussian shift experiments:

**Theorem 5** *Suppose that for some pair* $(\mu_D,\gamma) \in \mathbb{M}_D \times \Gamma$ *with* $\mu_D$ *full-rank and* $\Sigma_{\theta g}(\gamma) = \Sigma_{\theta\theta}(\gamma) = 0$*, for all* $C > 0$ *and all sequences* $(\mu_{D,n},\gamma_n) \in \mathbb{M}_D \times \Gamma$ *such that* $(\mu_{D,n},\gamma_n) \to (\mu_D,\gamma)$ *we have*

$$d_H\left(\mathcal{M}_D(\mu_{D,n},\gamma_n) \cap B_C, \{\mu_D \cdot b : b \in \mathbb{R}^p\} \cap B_C\right) \to 0$$

*where* $B_C = \{m : ||m|| \leq C\}$ *and* $d_H(A_1,A_2)$ *is the Hausdorff distance between the sets* $A_1$ *and* $A_2$*,*

$$d_H(A_1,A_2) = \max\left\{\sup_{x_1\in A_1}\inf_{x_2\in A_2}\|x_1 - x_2\|, \sup_{x_2\in A_2}\inf_{x_1\in A_1}\|x_1 - x_2\|\right\}.$$

*Then for* $\beta^u_{m,\mu_{D,n},\gamma_n} = \sup_{a\in[0,1]} E_{m,\mu_{D,n},\gamma_n}[\phi_a]$ *and all* $(\mu_{D,n},\gamma_n) \to (\mu_D,\gamma)$ *the MMRU weight*

$$a_{MMRU}(\mu_{D,n},\gamma_n) = \arg\min_{a\in[0,1]} \sup_{m\in\mathcal{M}_D(\mu_{D,n},\gamma_n)} \left(\beta^u_{m,\mu_{D,n},\gamma_n} - E_{m,\mu_{D,n},\gamma_n}[\phi_a]\right)$$

*satisfies* $a_{MMRU}(\mu_{D,n},\gamma_n) \to 0$.

Using Theorem 5 we can show that PI tests will be efficient under strong and semi-strong identification in Example I, while MMR tests will be efficient under strong and semi-strong identification in Example II, where the MMR and MMRU tests coincide.

**Example I: Weak IV (Continued)** Define $(g_T, \Delta g_T, \hat{\gamma})$ as in Section 1, and as above let $\pi_T = r_T c$ for $c \in \mathbb{R}^p$ with $||c|| > 0$. For simplicity we take $\hat{\mu}_D = D_T$ but the extension to other estimators is straightforward.

**Corollary 2** *Provided $\sqrt{T} r_T \to \infty$, we have that in the linear IV model $a_{PI}(\hat{\mu}_D, \hat{\gamma}) \to_p 0$ and thus that the PI test based on $(g_T, \Delta g_T, \hat{\gamma})$ is efficient under strong and semi-strong identification.*$\square$

**Example II: Minimum Distance (Continued)** We can model semi-strong identification in this example by taking $\Omega_\eta = r_T \Omega_{\eta,0}$ where $r_T \to 0$ and $r_T^{-1}\hat{\Omega}_\eta \to_p \Omega_{\eta,0}$, noting that $r_T = \frac{1}{T}$ is the typical strongly identified case. Again define $\hat{\gamma} = vec\left(\hat{\Omega}_\eta\right)$ and note that $\mathcal{M}(\hat{\gamma}) = \left\{\hat{\Omega}_\eta^{-\frac{1}{2}}\left(f(\theta) - f(\theta_0)\right)\right\}$. Defining $g_T(\theta) = \hat{\Omega}_\eta^{-\frac{1}{2}}\left(\hat{\eta} - f(\theta)\right)$ and $\Delta g_T(\theta) = \frac{\partial}{\partial\theta} g_T(\theta) = \hat{\Omega}_\eta^{-\frac{1}{2}} \frac{\partial}{\partial\theta} f(\theta)$ as before, a global identification assumption yields that PI tests are asymptotically efficient.

**Corollary 3** *Assume that $\theta$ is in the interior of $\Theta$ and that for all $\delta > 0$ there exists $\varepsilon(\delta) > 0$ such that $||f(\tilde{\theta}) - f(\theta)|| < \varepsilon(\delta)$ implies $||\tilde{\theta} - \theta|| < \delta$. Provided $r_T \to 0$, the MMR weight function $a_{MMR}$ satisfies $a_{MMR}(\hat{\gamma}) \to_p 0$ and the MMR test is efficient under strong and semi-strong identification.*$\square$

Hence, in our examples the plug-in test $\phi_{PI}$ is asymptotically efficient under strong and semi-strong identification.

# Appendix 2: Proofs

## Proof of Theorem 2

Statement (1) follows from results in Monti and Sen (1976) and Koziol and Perlman (1978). Specifically, both papers note that if $(A, B) \sim \left(\chi^2_{k-p}(\tau_A), \chi^2_p(\tau_B)\right)$ and $(\tau_A, \tau_B) = \lambda \cdot (t_A, t_B)$ for $t_A$, $t_B \geq 0$ then for $\phi$ any size $\alpha$ test for $H_0 : \tau_A = \tau_B = 0$ based on $(A, B)$ there exists some $\bar{\lambda} > 0$ such that for $0 < \lambda < \bar{\lambda}$,

$$E_{(\tau_A,\tau_B)}[\phi] \leq E_{(\tau_A,\tau_B)}\left[1\left\{\frac{t_A}{k-p}A + \frac{t_B}{p}B > c\right\}\right]$$

for $c$ the $1-\alpha$ quantile of a $\frac{t_A}{k-p}\chi^2_{k-p}+\frac{t_B}{p}\chi^2_p$ distribution. Statement (1) then follows immediately by the fact that $(J,K)\,|D=d\sim\left(\chi^2_{k-p}\left(\tau_J\right),\chi^2_p\left(\tau_K\right)\right)$.

Establishing statement (2) is similarly straightforward. In particular for $F_{t_K,t_J}$ as described in Theorem 2, Koziol and Perlman note that we can use the Neyman Pearson Lemma to establish that the weighted average power maximizing level $\alpha$ test based on $(A,B)\sim\left(\chi^2_{k-p}\left(\tau_A\right),\chi^2_p\left(\tau_B\right)\right)$ is $\phi^*_F=1\left\{\frac{t_K}{t_K+1}A+\frac{t_J}{t_J+1}B>c\right\}$, where $c$ is the $1-\alpha$ quantile of a $\frac{t_K}{t_K+1}\chi^2_p+\frac{t_J}{t_J+1}\chi^2_{k-p}$ distribution. In particular, for $\Phi_\alpha$ the class of level $\alpha$ tests based on $(A,B)$,

$$\phi^*_F\in\arg\max_{\phi\in\Phi_\alpha}\int_{\mathcal{T}(d)}E_{\tau_A,\tau_B}\left[\phi\right]dF(\tau_A,\tau_B).$$

Statement (2) again follows from the fact that $(J,K)\,|D=d\sim\left(\chi^2_{k-p}\left(\tau_J\right),\chi^2_p\left(\tau_K\right)\right)$.

## Proof of Theorem 3

By the independence of $J$, $K$, and $D$ under the null, conditional on the event $D=d$

$$K+a(D)\cdot J|D=d\sim\chi^2_p+a(d)\cdot\chi^2_{k-p}.$$

Hence

$$Pr\left\{K+a(D)\cdot J>c_\alpha\left(a\left(D\right)\right)\right|D=d\}=\alpha$$

so $E_{m=0,\mu_D}\left[\phi_{a(D)}\middle|D=d\right]=\alpha$ for all $d$ in the support of $D$ and all values $\mu_D$. $E_{m=0,\mu_D}\left[\phi_{a(D)}\right]$ can then be written as

$$\int E_{m=0,\mu_D}\left[\phi_{a(D)}\middle|D=d\right]dF_D=\int\alpha dF_D=\alpha$$

for $F_D$ the distribution of $D$, proving the theorem.

## Proof of Lemma 1

We prove that $E_{m,\mu_D}\left[\phi_{a(D)}|D\right]\geq\alpha$ almost surely, from which $E_{m,\mu_D}\left[\phi_{a(D)}\right]\geq\alpha$ follows immediately. Fix some CLC weight function $a\left(D\right)$. Recall that conditional on $D=d$, $J$ and $K$ are independently distributed $\chi^2_{k-p}\left(\tau_J\right)$ and $\chi^2_k\left(\tau_k\right)$, respectively, and

$$K+a(D)\cdot J|D=d\sim\chi^2_p\left(\tau_K\right)+a(d)\cdot\chi^2_{k-p}\left(\tau_J\right).$$

The CLC test $\phi_{a(D)}$ will reject if and only if $K+a(d)\cdot J > c_\alpha\left(a\left(d\right)\right)$, and the conditional probability of this event under $\tau_J = \tau_K = 0$ is $\alpha$. To establish the result, we need only show that the probability of this event is at least $\alpha$ under any pair $(\tau_J, \tau_K) \neq 0$. However, this follows from the form of the CLC test statistic and the observation that a noncentral $\chi^2$ distribution is increasing in its noncentrality parameter (in the sense of first-order stochastic dominance).

## Proof of Theorem 4

We first argue that conditional on $D = d$ the test $\phi_{QCLR_r}$ is exactly equivalent to the level $\alpha$ test that rejects for large values of the statistic $K + \frac{q_\alpha(r(d))}{q_\alpha(r(d))+r(d)} \cdot J$. This result is trivial for $r\left(d\right) = \infty$. For $r\left(d\right) < \infty$ and $K > 0$ or $J - r\left(D\right) > 0$, note first that for fixed $d$ the QCLR statistic is strictly increasing in $(J, K)$. Further, for any $L > 0$, the $L$ level set of the $QCLR_r$ statistic is of the form $L = K + \frac{L}{L+r(d)} \cdot J$ so that fixing $D = d$,

$$\left\{(J,K) \in \mathbb{R}^2_+ : QCLR_r = L\right\} = \left\{(J,K) \in \mathbb{R}^2_+ : L = K + \frac{L}{L + r(d)} \cdot J\right\}.$$

To verify that this is the case, note that if we plug $K = L - \frac{L}{L+r(d)} \cdot J$ into the $QCLR_r$ statistic and collect terms we have

$$QCLR_r = \frac{1}{2}\left(L + \frac{r\left(d\right)}{L + r\left(d\right)} \cdot J - r(d) + \sqrt{\left(L + r\left(d\right) + \frac{r\left(d\right)}{L + r\left(d\right)} \cdot J)\right)^2 - 4J \cdot r(d)}\right).$$

However,

$$\left(L + r(d) + \frac{r\left(d\right)}{L + r\left(d\right)} \cdot J\right)^2 - 4J \cdot r(d) = \left(L + r\left(d\right) - \frac{r\left(d\right)}{L + r\left(d\right)} \cdot J\right)^2$$

and thus for $K = L - \frac{L}{L+r(d)} \cdot J$,

$$QCLR = \frac{1}{2}\left(L + \frac{r\left(d\right)}{L + r\left(d\right)} \cdot J - r(d) + \sqrt{\left(L + r\left(d\right) - \frac{r\left(d\right)}{L + r\left(d\right)} \cdot J\right)^2}\right).$$

Since we've taken $K = L - \frac{L}{L+r(D)} \cdot J$ and we know $K \geq 0$, we have that $J \leq L + r\left(d\right)$. Thus $L + r\left(d\right) - \frac{r(d)}{L+r(d)} \cdot J \geq 0$ and we can open the square root and collect terms to obtain $QCLR_r = L$ on the set $\left\{(J,K)\in\mathbb{R}_+^2 : L = K + \frac{L}{L+r(d)}\cdot J\right\}$, as we claimed.

Conditional on $D=d$ the rejection region of $\phi_{QCLR_r}$ is

$$\left\{(J,K)\in\mathbb{R}_+^2 : q_\alpha\left(r(d)\right) < K + \frac{q_\alpha\left(r(d)\right)}{q_\alpha\left(r(d)\right)+r(d)}\cdot J\right\}.$$

Since $J$ and $K$ are pivotal under the null,

$$Pr_{m=0,\mu_D}\left\{q_\alpha\left(r(d)\right) < K + \frac{q_\alpha\left(r(d)\right)}{q_\alpha\left(r(d)\right)+r(d)}\cdot J \middle| D=d\right\} = \alpha,$$

so since $K+\frac{q_\alpha(r(d))}{q_\alpha(r(d))+r(d)}\cdot J$ is continuously distributed with support equal $\mathbb{R}_+$, $q_\alpha\left(r(d)\right)$ must be the $1-\alpha$ quantile of this random variable. Hence, if we define the test $\phi_{\tilde{a}(D)}$ as in (18) with $\tilde{a}(D)=\frac{q_\alpha(r(D))}{q_\alpha(r(D))+r(D)}$, we can see that $c_\alpha\left(\tilde{a}\left(d\right)\right)=q_\alpha\left(d\right)$ and thus that $\phi_{QCLR_r}=\phi_{\tilde{a}(d)}$ conditional on $D=d$. Since this holds for all $d$, $\phi_{QCLR_r}\equiv\phi_{\tilde{a}(D)}$. Thus, for any function $r:\mathcal{D}\to\mathbb{R}_+\cup\{\infty\}$ there is a function $\tilde{a}:\mathcal{D}\to[0,1]$ such that $\phi_{QCLR_r}\equiv\phi_{\tilde{a}(D)}$.

To prove the converse, that for any CLC test $\phi_{a(D)}$ for $a:\mathcal{D}\to[0,1]$ we can find a function $r:\mathcal{D}\to\mathbb{R}_+\cup\{\infty\}$ yielding the same test, fix the function $a(D)$ and note that $q_\alpha(r(D))$ is a continuous function of $r(D)$ which is deceasing in $r(D)$ and is bounded below by $\chi^2_{p,1-\alpha}$ and above by $\chi^2_{k,1-\alpha}$ (see Moreira (2003)). Hence for any value $d$, as $r(d)$ goes from zero to infinity $\frac{q_\alpha(r(d))}{q_\alpha(r(d))+r(d)}$ varies continuously between zero and one, with $\lim_{r(d)\to0}\frac{q_\alpha(r(d))}{q_\alpha(r(d))+r(d)}=1$ and $\lim_{r(d)\to\infty}\frac{q_\alpha(r(d))}{q_\alpha(r(d))+r(d)}=\frac{q_\alpha(\infty)}{q_\alpha(\infty)+\infty}=0$. If $a\left(d\right)=0$ define $\tilde{r}\left(d\right)=\infty$. If $a\left(d\right)>0$, note that there exists a value $r^*<\infty$ such that $a(d)>\frac{q_\alpha(r^*)}{q_\alpha(r^*)+r^*}$, so by the intermediate value theorem we can pick $\tilde{r}(d)\in[0,r^*]$ such that $a(d)=\frac{q_\alpha(\tilde{r}(d))}{q_\alpha(\tilde{r}(d))+\tilde{r}(d)}$. Repeating this exercise for all values $d$ we can construct a function $\tilde{r}:\mathcal{D}\to\mathbb{R}_+\cup\{\infty\}$ such that $\phi_{a(D)}\equiv\phi_{QCLR_{\tilde{r}}}$, completing the proof.

## Proof of Proposition 1

The discussion preceding Proposition 1 establishes that under $(\theta_0,\gamma)$, $(J_T,K_T,D_T)\to_d(J,K,D)$ and $\hat{\gamma}\to_p\gamma$. Since we assume that $a\left(D,\gamma\right)$ is almost everywhere continuous with respect to the limiting distribution $F$ and $c_\alpha\left(a\right)$ is a continuous function of $a$, the Continuous Mapping Theorem establishes that

$$K_T + a\left(D_T,\hat{\gamma}\right)J_T - c_\alpha\left(a\left(D_T,\hat{\gamma}\right)\right)\to_d K + a\left(D,\gamma\right)J - c_\alpha\left(a\left(D,\gamma\right)\right).$$

Since zero is a point of continuity of the distribution of the right hand side this implies that

$$Pr_{T,(\theta_0,\gamma)}\left\{K_T + a\left(D_T,\hat{\gamma}\right)J_T > c_\alpha\left(a\left(D_T,\hat{\gamma}\right)\right)\right\} \to$$
$$Pr_{m=0,\mu_D}\left\{K + a\left(D,\gamma\right)J > c_\alpha\left(a\left(D,\gamma\right)\right)\right\} = \alpha$$

which proves (22). To prove (23) note that the results above establish that $\phi_{a(D,\gamma)}$ is almost-everywhere continuous with respect to $F$, and hence for $f \in \mathcal{F}$

$$\left(\phi_{a(D_T,\hat{\gamma})} - \alpha\right) f\left(D_T\right) \to_d \left(\phi_{a(D,\gamma)} - \alpha\right) f\left(D\right).$$

Since the left hand side is bounded, convergence in distribution implies convergence in expectation, proving (23).

## Proof of Proposition 2

Let us take the estimator $\hat{\Omega}$ to be

$$\hat{\Omega} = \begin{pmatrix} \hat{\Omega}_{ff} & \hat{\Omega}_{f\beta} \\ \hat{\Omega}_{\beta f} & \hat{\Omega}_{\beta\beta} \end{pmatrix} =$$
$$\frac{1}{T}\sum_t \begin{pmatrix} f_t(\beta_0) - f_T\left(\beta_0\right) \\ \frac{\partial}{\partial\beta} f_t(\beta_0) - \frac{\partial}{\partial\beta} f_T\left(\beta_0\right) \end{pmatrix} \begin{pmatrix} f_t(\beta_0)' - f_T\left(\beta_0\right)' & \frac{\partial}{\partial\beta} f_t(\beta_0)' - \frac{\partial}{\partial\beta} f_T\left(\beta_0\right)' \end{pmatrix}$$

and

$$\hat{\Sigma} = \begin{pmatrix} I_k & \hat{\Omega}_{ff}^{-\frac{1}{2}}\hat{\Omega}_{f\beta}\hat{\Omega}_{ff}^{-\frac{1}{2}} \\ \hat{\Omega}_{ff}^{-\frac{1}{2}}\hat{\Omega}_{\beta f}\hat{\Omega}_{ff}^{-\frac{1}{2}} & \hat{\Omega}_{ff}^{-\frac{1}{2}}\hat{\Omega}_{\beta\beta}\hat{\Omega}_{ff}^{-\frac{1}{2}} \end{pmatrix}.$$

These choices imply that our $S_T$ and $K_T$ coincide exactly with $AR$ and $LM$ in ACG, and that our $D_T$ is $\sqrt{T}\hat{\Omega}_{ff}^{-\frac{1}{2}}\hat{D}$ for $\hat{D}$ as in ACG. To prove the proposition we will rely heavily on their results. ACG consider two cases: sequences $\lambda_T$ for which $\sqrt{T}||\pi_T||$ converges to a constant and those for which it diverges to infinity.

Let us begin by considering the case where $\sqrt{T}||\pi_T||$ converges. ACG establish that for this case their $(LM, AR, \hat{D})$ converges in distribution to $\left(\chi_1^2, \chi_{k-1}^2, \tilde{D}\right)$ where all three random variables are independent and $\tilde{D}$ has a non-degenerate Gaussian distribution. Since $\hat{\Omega}_{ff} \to_p \Omega_{ff}$ which is full-rank by assumption, this proves that $(K_T, S_T, D_T) \to_d \left(\chi_1^2, \chi_{k-1}^2, D\right)$ where again all the variables on the RHS are mutually independent and $D$ has a non-degenerate Gaussian distribution. Thus, by the Continuous Mapping Theorem and consistency of $\hat{\Sigma}_{\theta g}$ and $\hat{\Sigma}_{\theta\theta}$, which under the null follows from (6.7) and (6.9) in ACG, we have that

$$\begin{gathered}(1-a(D_T,\hat{\gamma}))\,K_T + a(D_T,\hat{\gamma})S_T - c_\alpha\left(a(D_T,\hat{\gamma})\right) \to_d \\ (1-a(D,\gamma))\,K + a(D,\gamma)S - c_\alpha\left(a(D,\gamma)\right)\end{gathered}$$

which establishes correct asymptotic size under sequences with $\sqrt{T}||\pi_T||$ converging.

Next, consider the case where $\sqrt{T}||\pi_T||$ diverges. Let

$$(\tilde{g}_T, \Delta\tilde{g}_T, \tilde{\gamma}) = h\left(g_T, \Delta g_T, \hat{\gamma}; ||\pi_T||^{-1}\right),$$

and define the random variables $\tilde{D}_T$, $\tilde{\Sigma}$, and $\tilde{\Sigma}_D$ accordingly. ACG equation (6.22) establishes that in this case $\tilde{D}_T \to_p D^*$ for $||D^*|| > 0$, and equations (6.7) and (6.21) together establish that $\tilde{\Sigma}_D \to_p 0$. Our assumption on $a(\tilde{D}_T, \tilde{\gamma})$ thus implies that $a(\tilde{D}_T, \tilde{\gamma}) \to_p a_0$. Since ACG establish the convergence in distribution of $(LM, AR)$ under sequences of this type, we have that

$$\left(1-a(\tilde{D}_T,\tilde{\gamma})\right)K_T + a(D_T,\xi_T)S_T - c_\alpha\left(a(D_T,\xi_T)\right) \to_d (1-a_0)\,K + a_0 S - c_\alpha\left(a_0\right)$$

and thus that the CLC test $\phi_{a\left(\tilde{D}_T,\tilde{\gamma}\right)}$ has asymptotic rejection probability equal to $\alpha$ under these sequences. By the assumed invariance the postmultiplication, however, this implies that $\phi_{a(D_T,\gamma)}$ has asymptotic rejection probability $\alpha$ as well.

To complete the proof, following ACG we can note that the above argument verifies their Assumption $B^*$ and that we can thus use ACG Corollary 2.1 to establish the result.

## Proof of Proposition 3

Follows by the same argument as the first part of Proposition 1.

## Proof of Proposition 4

As discussed in the text, $\phi_K$ is efficient in the limit problem (25) by the Neyman-Pearson Lemma, and $\phi_{K_T} = \phi_{\tilde{a}(D_T,\hat{\gamma})}$ for $\tilde{a}\,(D,\gamma) \equiv 0$, $\tilde{a} \in \mathcal{A}_c$, so

$$\lim_{T\to\infty} E_{T,(\theta,\gamma)}\left[\phi_{K_T}\right] = \sup_{\tilde{a}\in\mathcal{A}_c} \lim_{T\to\infty} E_{T,(\theta,\gamma)}\left[\phi_{\tilde{a}(D_T,\hat{\gamma})}\right]$$

follows from Proposition 3.

If $a(D,\gamma) = 0$ almost surely, then we have that $\lim_{T\to\infty} E_{T,(\theta,\gamma)}\left[\phi_{a(D_T,\hat{\gamma})}\right] = \lim_{T\to\infty} E_{T,(\theta,\gamma)}\left[\phi_{K_T}\right]$ by Proposition 3. If, on the other hand, $Pr\{a(D,\gamma) \neq 0\} = \delta > 0$, note that $D = \mu$ is non-random in the limit problem, so this implies that $a(\mu,\gamma) = a^* \neq 0$. Note, however, that the test $\phi_{a^*}$ does not satisfy the necessary condition for a most powerful test given in Theorem 3.2.1 in Lehmann and Romano and thus has strictly lower power than the test $\phi_K$ in the limit problem, which together with Proposition 3 implies that $\lim_{T\to\infty} E_{T,(\theta,\gamma)}\left[\phi_{a(D_T,\hat{\gamma})}\right] < \lim_{T\to\infty} E_{T,(\theta,\gamma)}\left[\phi_{K_T}\right]$.

## Proof of Theorem 5

Define $\mathcal{M}_L = \{\mu_D \cdot b : b \in \mathbb{R}^p\}$. Note that for any $\zeta > 0$, there exists $C_\zeta > 0$ such that

$$\inf_{a\in[0,1]} \inf_{m\in\mathcal{M}_L:||m||>C_\zeta} E_{m,\mu_D,\gamma}\left[\phi_a\right] > 1-\zeta.$$

Note further that $C_\zeta \to \infty$ as $\zeta \to 0$. Since the test $\phi_K$ is UMP over the class of tests depending on $(J,K,D)$ against $m \in \mathcal{M}_L$ for $\Sigma_D = 0$, we can see that for $\beta^u_{m,\mu_D,\gamma} = \sup_{a\in[0,1]} E_{m,\mu_D,\gamma}\left[\phi_a\right]$ we have $\beta^u_{m,\mu_D,\gamma} = E_{m,\mu_D,\gamma}\left[\phi_K\right]\ \forall m \in \mathcal{M}_L$. Thus,

$$\sup_{m\in\mathcal{M}_L}\left(\beta^u_{m,\mu_D,\gamma} - E_{m,\mu_D,\gamma}\left[\phi_a\right]\right) = \sup_{m\in\mathcal{M}_L}\left(E_{m,\mu_D,\gamma}\left[\phi_K\right] - E_{m,\mu_D,\gamma}\left[\phi_a\right]\right).$$

Next note that, as discussed in the proof of Proposition 4, none of the tests $\phi_a : a \in (0,1]$ satisfy the necessary condition for an optimal test against $m \in \mathcal{M}_L$ for $\Sigma_D = 0$ given in Lehman and Romano Theorem 3.2.1. Thus if we define

$$\varepsilon(a) = \sup_{m\in\mathcal{M}_L}\left(E_{m,\mu_D,\gamma}\left[\phi_K\right] - E_{m,\mu_D,\gamma}\left[\phi_a\right]\right)$$

we have that $\varepsilon(a) > 0\ \forall a \in (0,1]$. Moreover for all $a$ there is some $m^* \in \mathcal{M}_L$ such that

$$\varepsilon(a) = E_{m^*,\mu_D,\gamma}\left[\phi_K\right] - E_{m^*,\mu_D,\gamma}\left[\phi_a\right],$$

which can be seen by noting that for $\zeta = \frac{\varepsilon(a)}{2}$, $B_C = \{m : ||m|| \leq C\}$, and $A = \mathcal{M}_L \cap B^C_{C_\zeta}$ (for $B^C_{C_\zeta}$ the complement of $B_{C_\zeta}$)

$$\sup_{m\in A}\left(E_{m,\mu_D,\gamma}\left[\phi_K\right] - E_{m,\mu_D,\gamma}\left[\phi_a\right]\right) \leq 1 - \frac{\varepsilon(a)}{2}$$

by the definition of $C_\zeta$. Thus, for $\tilde{A} = \mathcal{M}_L \cap B_{C_\zeta}$,

$$\varepsilon(a) = \sup_{m \in \tilde{A}} \left( E_{m,\mu_D,\gamma}\left[\phi_K\right] - E_{m,\mu_D,\gamma}\left[\phi_a\right] \right).$$

Since $\tilde{A}$ is compact and $E_{m,\mu_D,\gamma}\left[\phi_K\right] - E_{m,\mu_D,\gamma}\left[\phi_a\right]$ is continuous in $m$, the sup must be attained at some $m^* \in \tilde{A}$.

Since $E_{m,\mu_D,\gamma}\left[\phi_a\right]$ is continuous in $a$ for all $m$, the fact that

$$\varepsilon\left(a\right) = \sup_{m \in \mathcal{M}_L} \left( E_{m,\mu_D,\gamma}\left[\phi_K\right] - E_{m,\mu_D,\gamma}\left[\phi_a\right] \right)$$

is achieved implies that $\varepsilon\left(a\right)$ is continuous in $a$. We know that $\varepsilon(0) = 0$ by definition, so 0 is the unique minimizer of $\varepsilon(a)$ over $[0, 1]$. By the compactness of $[0, 1]$, this implies that for any $\delta > 0$ there exists $\bar{\varepsilon}\left(\delta\right) > 0$ such that $\varepsilon(a) < \bar{\varepsilon}\left(\delta\right)$ only if $a < \delta$. Further, by the intermediate value theorem there exists $a(\delta) > 0$ such that $\varepsilon\left(a(\delta)\right) = \frac{\bar{\varepsilon}(\delta)}{2}$.

To prove Theorem 5 we want to show that under the assumptions of the theorem, for all $\nu > 0$ there exists $N$ such that $n > N$ implies

$$\arg\min_{a \in [0,1]} \sup_{m \in \mathcal{M}_D\left(\mu_{D,n},\gamma_n\right)} \left( \beta^u_{m,\mu_{D,n},\gamma_n} - E_{m,\mu_{D,n},\gamma_n}\left[\phi_{a^*}\right] \right) < \nu.$$

Fixing $\nu$, let $\bar{\varepsilon}^* = \bar{\varepsilon}\left(\nu\right)$, $a^* = a(\nu)$, for $\bar{\varepsilon}\left(\cdot\right)$ and $a\left(\cdot\right)$ as defined above. Let $\zeta^* = \frac{\bar{\varepsilon}^*}{4}$, and take $C^*$ to be such that

$$\inf_{m \in \mathbb{R}^k : ||m|| > C^*} E_{m,\mu_D,\gamma}\left[\phi_{a^*}\right] > 1 - \zeta^*.$$

Under our assumptions and the continuity of $E_{m,\mu_D,\gamma}\left[\phi_a\right]$ in $(m, \mu_D, \gamma, a)$, there exists some $N$ such that for $n > N$,

$$\inf_{a \in [\nu,1]} \sup_{m \in \mathcal{M}_D\left(\mu_{D,n},\gamma_n\right) \cap B_{C^*}} \left( \beta^u_{m,\mu_{D,n},\gamma_n} - E_{m,\mu_{D,n},\gamma_n}\left[\phi_a\right] \right) > 3\zeta^*$$

while

$$\sup_{m \in \mathcal{M}_D\left(\mu_{D,n},\gamma_n\right) \cap B_{C^*}} \left( \beta^u_{m,\mu_{D,n},\gamma_n} - E_{m,\mu_{D,n},\gamma_n}\left[\phi_{a^*}\right] \right) < 3\zeta^*$$

and

$$\sup_{m \in \mathcal{M}_D\left(\mu_{D,n},\gamma_n\right) \cap B^C_{C^*}} \left( \beta^u_{m,n} - E_{m,\mu_{D,n},\gamma_n}\left[\phi_{a^*}\right] \right) < 2\zeta^*.$$

Thus, for $n > N$ we have

$$\sup_{m \in \mathcal{M}_D\left(\mu_{D,n},\gamma_n\right)} \left(\beta^u_{m,\mu_{D,n},\gamma_n} - E_{m,\mu_{D,n},\gamma_n}\left[\phi_{a^*}\right]\right) <$$
$$\inf_{a\in[\nu,1]} \sup_{m \in \mathcal{M}_D\left(\mu_{D,n},\gamma_n\right)\cap B_{C^*}} \left(\beta^u_{m,\mu_{D,n},\gamma_n} - E_{m,\mu_{D,n},\gamma_n}\left[\phi_{a}\right]\right)$$

and thus that $a\left(\mu_{D,n},\gamma_n\right) < \nu$ since $a^* < \nu$. Since we can repeat this argument for all $\nu > 0$ we obtain that $a\left(\mu_{D,n},\gamma_n\right) \to 0$ as desired.

## Proof of Corollary 2

Let $\left(\tilde{g}_T, \Delta\tilde{g}_T, \tilde{\gamma}\right) = h\left(g_T, \Delta g_T, \hat{\gamma}; r_T^{-1}/\sqrt{T}\right)$ for $h$ as defined in (21), and note that this is the same definition of $\left(\tilde{g}_T, \Delta\tilde{g}_T, \tilde{\gamma}\right)$ given near the beginning of Appendix 1. By the postmultiplication-invariance of plug-in tests with equivariant $\hat{\mu}_D$, tests based on $\left(\tilde{g}_T, \Delta\tilde{g}_T, \tilde{\gamma}\right)$ with plug-in estimate $\tilde{\mu}_D = \tilde{D}_T$ will be the same as those based on $\left(g_T, \Delta g_T, \hat{\gamma}\right)$ with estimate $\hat{\mu}_D = D_T$. To prove the result we will focus on tests based on $\left(\tilde{g}_T, \Delta\tilde{g}_T, \tilde{\gamma}\right)$.

As established in the main text, $\left(\tilde{g}_T, \Delta\tilde{g}_T, \tilde{\gamma}\right)$ converges in distribution to $(g, \Delta g, \gamma)$ in a Gaussian shift model with $\mu = E\left[Z_t Z_t\right]' c$ and $b = b^*$. Note that in linear IV we have

$$\mathcal{M}_D\left(\mu_D, \gamma\right) = \left\{\left(I - \Sigma_{\beta g} \cdot b\right)^{-1} \mu_D \cdot b : b \in \mathbb{R}\right\}.$$

Hence, for any sequence $\left(\mu_{D,n}, \gamma_n\right)$ with $\mu_{D,n} \to \mu$, $||\mu|| > 0$, and $\Sigma_{\beta g}\left(\gamma_n\right) \to 0$ we can see that for any $C > 0$

$$d_H\left(\mathcal{M}_D\left(\mu_{D,n}, \gamma_n\right) \cap B_C, \left\{\mu_D \cdot b : b \in \mathbb{R}^p\right\} \cap B_C\right) \to 0,$$

so by Theorem 5 we have that $a_{PI}\left(\mu_{D,n}, \gamma_n\right) \to 0$. Note, however, that under our assumptions $\left(\hat{\mu}_D, \hat{\gamma}\right) \to_p (\mu, \gamma)$ with $||\mu|| > 0$ and $\Sigma_{\beta g}(\gamma) = \Sigma_D(\gamma) = 0$. Thus, the Continuous Mapping Theorem yields that $a_{PI}\left(\hat{\mu}_D, \hat{\gamma}\right) \to_p 0$.

## Proof of Corollary 3

Note that

$$\mathcal{M}\left(\hat{\gamma}\right) = \left\{\hat{\Omega}_\eta^{-\frac{1}{2}}\left(f(\theta) - f\left(\theta_0\right)\right) : \theta \in \Theta\right\} = r_T^{-\frac{1}{2}}\left\{\left(r_T^{-1}\hat{\Omega}_\eta\right)^{-\frac{1}{2}}\left(f(\theta) - f\left(\theta_0\right)\right) : \theta \in \Theta\right\}.$$

For any sequence $r_T^{-1}\Omega_{\eta,T} \to \Omega_{\eta,0}$ and $B_C = \{m \in \mathbb{R}^p : ||m|| \leq C\}$ for $C > 0$ we have that

$$\lim_{T\to\infty} d_H\left(\left\{r_T^{-\frac{1}{2}}\left(r_T^{-1}\Omega_{\eta,T}\right)^{-\frac{1}{2}}\left(f(\theta) - f(\theta_0)\right) : \theta \in \Theta\right\} \cap B_C, \left\{r_T^{-\frac{1}{2}}\Omega_{\eta,0}^{-\frac{1}{2}}\left(f(\theta) - f(\theta_0)\right) : \theta \in \Theta\right\} \cap B_C\right) = 0.$$

From the definition of differentiability, we know that

$$\lim_{\theta\to\theta_0} \frac{f(\theta) - f(\theta_0) - \frac{\partial}{\partial\theta'}f(\theta_0)(\theta - \theta_0)}{||\theta - \theta_0||} = 0.$$

Thus, for any sequence $\delta_T \to 0$,

$$\lim_{\delta_T\to 0} \sup_{||\theta-\theta_0||\leq\delta_T} \frac{1}{\delta_T}\left(f(\theta) - f(\theta_0) - \frac{\partial}{\partial\theta'}f(\theta_0)(\theta - \theta_0)\right) = 0.$$

Moreover, by our identifiability assumption on $\theta_0$ we know that for any constant $K > 0$,

$$\lim_{T\to\infty} \sup_{\theta : r_T^{-\frac{1}{2}}||\Omega_{\eta,0}^{-\frac{1}{2}}f(\theta) - \Omega_{\eta,0}^{-\frac{1}{2}}f(\theta_0)|| \leq K} ||\theta - \theta_0|| = 0.$$

Combined with the previous equation, this implies that

$$\lim_{T\to\infty} \sup_{\theta : r_T^{-\frac{1}{2}}||\Omega_{\eta,0}^{-\frac{1}{2}}f(\theta) - \Omega_{\eta,0}^{-\frac{1}{2}}f(\theta_0)|| \leq K} r_T^{-\frac{1}{2}}\left\|\Omega_{\eta,0}^{-\frac{1}{2}}\left(f(\theta) - f(\theta_0)\right) - \Omega_{\eta,0}^{-\frac{1}{2}}\frac{\partial}{\partial\theta'}f(\theta_0)(\theta - \theta_0)\right\| = 0$$

which in turn shows that for any $C > 0$, provided $\theta$ belongs to the interior of $\Theta$

$$d_H\left(r_T^{-\frac{1}{2}}\left\{\Omega_{\eta,0}^{-\frac{1}{2}}\left(f(\theta) - f(\theta_0)\right) : \theta \in \Theta\right\} \cap B_C, r_T^{-\frac{1}{2}}\left\{\Omega_{\eta,0}^{-\frac{1}{2}}\frac{\partial}{\partial\theta'}f(\theta_0) \cdot b : b \in \mathbb{R}^p\right\} \cap B_C\right) \to 0.$$

Thus, we see that for any $r_T^{-1}\Omega_{\eta,T} \to \Omega_{\eta,0}$ the convergence required by Theorem 5 holds, so for the corresponding sequence $\{\gamma_T\}_{T=1}^{\infty}$ we have that $a_{MMR}(\gamma_T) \to 0$. Hence, by the Continuous Mapping Theorem we have that under our assumptions $a_{MMR}(\hat{\gamma}) \to_p 0$.

One can show that sequences of local alternatives of the form $\theta_T = \theta_0 + r_T^{\frac{1}{2}}b^*$ yield Gaussian Shift limit problems in this model. The fact that $a_{MMR}(\hat{\gamma}) \to_p 0$ implies, by Proposition 4, that the MMR test is asymptotically efficient against such sequences, and hence that the MMR test is asymptotically efficient under strong and semi-strong identification, as we wanted to prove.

# Additional References


Donald W.K. Andrews and Patrik Guggenberger. Asymptotic size of kleibergen's lm and conditional lr tests for moment condition models. Unpublished Manuscript, 2014.

D.W.K. Andrews, X. Cheng, and P. Guggenberger. Generic results for establishing the asymptotic size of confidence sets and tests. Cowles Foundation working paper., 2011.

M. Jansson and M.J. Moreira. Optimal inference in a model with nearly integrated regressors. *Econometrica*, 74:681–714, 2006.