Asymptotic F and t Tests in an Efficient GMM Setting*

Jungbin Hwang
Department of Economics,
University of Connecticut, USA

Yixiao Sun
Department of Economics,
University of California, San Diego, USA
& Economics and Management School,
Wuhan University, China

Abstract

This paper considers two-step efficient GMM estimation and inference where the weighting matrix and asymptotic variance matrix are based on the series long run variance estimator. We propose a simple and easy-to-implement modification to the trinity of test statistics in the two-step efficient GMM setting and show that the modified test statistics are all asymptotically F distributed under the so-called fixed-smoothing asymptotics. The modification is multiplicative and involves the J statistic for testing over-identifying restrictions. This leads to convenient asymptotic F tests whose critical values, i.e., the standard F critical values, are readily available from standard statistical tables and programming environments. For testing a single restriction with a one-sided alternative, an asymptotic t test theory using the standard t distribution as the reference distribution is also developed.

JEL Classification: C12, C32

Keywords: Efficient GMM, *F* distribution, Fixed-smoothing Asymptotics, Heteroskedasticity and Autocorrelation Robust, *t* distribution, Two-step GMM

*Email: jungbin.hwang@uconn.edu and yisun@ucsd.edu. Correspondence to: Yixiao Sun, Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508.

We thank the coeditor Han Hong, an associate editor, and three anonymous referees for helpful comments. We also thank David M. Kaplan for careful reading and constructive comments. Sun gratefully acknowledges partial research support from NSF under Grant No. SES-1530592.

1 Introduction

This paper considers the optimal two-step GMM estimator and the associated tests in a time series setting. In the presence of nonparametric temporal dependence, the optimal weighting matrix is the inverted long run variance (LRV) of the moment process. To implement the two-step GMM method, we often estimate the LRV using the nonparametric kernel or series method. Given the nonparametric nature of the LRV estimator, there is a high variation in the weighting matrix with consequent effects on the two-step point estimator and the associated tests. Recently Sun (2014b) employs the fixed-smoothing asymptotics and establishes a new asymptotic approximation that captures the estimation uncertainty in the LRV estimator. Under the fixed-smoothing asymptotics, the point estimator is asymptotically mixed normal, and the test statistics such as the Wald statistic converge to a nonstandard distribution. In the case of series LRV estimation, Sun (2014b) shows that the nonstandard limiting distribution can be approximated by a noncentral F distribution.

In this paper, we follow Sun (2014b) but focus on the series LRV estimator. We modify the usual test statistics, including the Wald statistic, the quasi LR statistic, and the LM statistic and show that the modified test statistics are all asymptotically standard F distributed. The standard F distribution is the exact limiting distribution. No additional approximation is needed. This is in contrast to Sun (2014b) where the noncentral F distribution is an approximation to the fixed-smoothing limiting distribution. The standard F distribution is more accessible than the noncentral F distribution, as standard F critical values are readily available from standard statistical tables.

The modification involves the usual J statistic for testing overidentifying restrictions. The modified test statistics are scaled versions of the original test statistics with the scaling factor depending on the J statistic. So the modification is very easy to implement. To understand the modification, in the supplementary appendix, we cast the two-step GMM estimation and inference into OLS estimation and inference in a classical normal linear regression (CNLR). We show that the modified Wald statistic in the GMM framework is exactly the usual Wald statistic constructed in the standard way in the CNLR framework.

Our proposed asymptotic F tests, which are based on the modified test statistics and use the standard F approximation, can be regarded as conditional tests conditioning on the J statistic. In contrast, the nonstandard tests of Sun (2014b) can be regarded as unconditional tests. As conditional tests, the F tests have the exact asymptotic level regardless of the value of the J statistic. This is an attractive property that the nonstandard tests lack. The conditioning argument is entirely analogous to that used in the linear regression model with stochastic regressors that are independent of the regression error.

Monte Carlo simulations show that our proposed asymptotic F tests are as accurate in size as the corresponding nonstandard tests of Sun (2014b). They are also as powerful as, and sometimes more powerful than, the latter tests. So there is no power loss in using the asymptotic F tests. Like the nonstandard tests of Sun (2014b), the asymptotic F tests are much more accurate in size than the usual chi-square tests without any power sacrifice. Given the convenience of the standard F approximation, we recommend the asymptotic F tests for practical use.

Our proposed F tests can still have some size distortion when temporal dependence is strong. The size distortion problem is not unique to the F tests but applies to any pointwise heteroskedasticity and autocorrelation robust (HAR) test, including both the fixed-smoothing asymptotic tests and the conventional asymptotic normal and chi-squared tests. Recently, Preinerstorfer and Pötscher (2016) have considered uniform inference and discovered some 'singularity' points in the data generating process under which any pointwise HAR test will suffer from severe size distortion. Our proposed F tests are no exception. A 'singularity' point that is especially relevant in economic applications is the near unity of an AR root in the underlying moment processes. See Preinerstorfer and Pötscher (2016, B(ii) in Sec 3.2.2), Müller (2014), and Sun (2014c) for discussions on the problems generated by the near-unit-root singularity. A standard remedy is to combine prewhitening with the fixed-smoothing asymptotics developed here. However, prewhitening is not effective in handling the singularities featured in Preinerstorfer and Pötscher (2016). See Preinerstorfer (2015) for more discussion.

When there is a single restriction to be tested and the alternative is one-sided, we construct the usual t-statistic and then modify it using the J statistic. We show that the modified t statistic is asymptotically t distributed. The resulting asymptotic t test is as easy to use as the asymptotic normal test. The theory of the asymptotic standard t test parallels that of asymptotic standard F tests.

This paper contributes to a growing body of literature on fixed-smoothing asymptotics. For kernel LRV estimators such as the Newey-West estimator (Newey and West, 1987), the fixed-smoothing asymptotics is the so-called fixed-b asymptotics first studied by Kiefer and Vogelsang

(2002a, 2002b, 2005) in the econometrics literature. Subsequent research includes Jansson (2004), Sun, Phillips and Jin (2008), Sun and Phillips (2009), and Gonçalves and Vogelsang (2011), among others. Papers that are most closely related to this paper are those that use the series LRV estimators. In this case, the fixed-smoothing asymptotics is the so-called fixed-K asymptotics. Some examples of these papers are Phillips (2005), Müller (2007), Sun (2011, 2013, 2014a&b), and Sun and Kim (2012).

In the case of series LRV estimation, the F and t limit theory has been established in Sun (2011) for trend regression, Sun (2013) for stationary moment processes, and Sun (2014c) for highly persistent moment processes. See also Sun and Kim (2012, 2015) for the J test, Wald test, and t test in the spatial setting. All these papers focus on the first-step GMM estimator or OLS estimator. This paper is the first to establish the F and t limit theory for the trinity of test statistics in a two-step efficient GMM framework. This is not trivial, as the asymptotic pivotality of these statistics under the fixed-smoothing asymptotics was not established until very recently in Sun (2014b).

The rest of the paper is organized as follows. Section 2 presents the basic setting and introduces the modified test statistics. Section 3 establishes the fixed-smoothing asymptotics of the modified test statistics and develops the asymptotic F and t tests. Section 4 investigates the asymptotic properties of the F and t tests under local alternative hypotheses. The next section reports simulation evidence. The last section concludes. Proofs are given in the appendix. An online supplementary appendix sheds further light on the asymptotic F and t tests by connecting them with the familiar F and t tests in a classical linear normal regression.

Notationally, $\mathcal{F}_{d_1,d_2}(\lambda)$ is a random variable that follows the noncentral F distribution with degrees of freedom d_1 and d_2 and noncentrality parameter λ, and $\mathcal{T}_d(\lambda)$ is a random variable that follows the noncentral t distribution with degrees of freedom d and noncentrality parameter λ. $\mathcal{F}_{d_1,d_2} := \mathcal{F}_{d_1,d_2}(0)$ and $\mathcal{T}_d := \mathcal{T}_d(0)$ are random variables following the standard F and t distributions. When there is no possibility of confusion, we sometimes identify a random variable with its distribution.

2 Two-step GMM Estimation and Testing

We consider the standard GMM setting with moment conditions

$$Ef\left(v_t,\theta_0\right)=0,\ t=1,2,\ldots,T, \tag{1}$$

where v_t is the vector of observations at time t, $\theta_0\in\Theta\subseteq\mathbb{R}^d$ is the parameter of interest, and $f\left(v_t,\theta\right)$ is the $m\times 1$ vector of moment conditions that are twice continuously differentiable. We assume that $Ef\left(v_t,\theta\right)=0$ if and only if $\theta=\theta_0$ so that θ_0 is point identified. The model may be overidentified with the degree of overidentification $q=m-d\geq 0$. We allow $\left\{f\left(v_t,\theta_0\right)\right\}$ to have autocorrelation of unknown forms.

Define

$$g_t\left(\theta\right)=\frac{1}{T}\sum_{j=1}^{t}f(v_j,\theta),$$

then the GMM estimator of θ_0 is given by

$$\hat{\theta}_{GMM}=\arg\min_{\theta\in\Theta}g_T\left(\theta\right)'W_T^{-1}g_T\left(\theta\right),$$

where W_T is a positive definite weighting matrix. The initial first-step GMM estimator can be obtained by choosing W_T to be a matrix $W_{o,T}$ that does not depend on any unknown parameter. This gives rise to

$$\tilde{\theta}_T=\arg\min_{\theta\in\Theta}g_T\left(\theta\right)'W_{o,T}^{-1}g_T\left(\theta\right).$$

Here $W_{o,T}$ may depend on the sample size T, but we assume that $W_{o,T}\overset{p}{\to}W_{o,\infty}$, a matrix that is positive definite almost surely.

With the first step estimator $\tilde{\theta}_T$, we can estimate the optimal weighting matrix W_T, which is the asymptotic variance matrix of $\sqrt{T}g_T\left(\theta_0\right)$. See Hansen (1982). Most, if not all, estimators of the asymptotic variance take the form

$$W_T(\tilde{\theta}_T)=\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}Q_K\left(\frac{t}{T},\frac{s}{T}\right)\left(f(v_t,\tilde{\theta}_T)-\frac{1}{T}\sum_{\tau=1}^{T}f(v_\tau,\tilde{\theta}_T)\right)\left(f(v_s,\tilde{\theta}_T)-\frac{1}{T}\sum_{\tau=1}^{T}f(v_\tau,\tilde{\theta}_T)\right)', \tag{2}$$

where $Q_K\left(r,s\right)$ is a symmetric weighting function that depends on the smoothing parameter K. In this paper, we focus on the series LRV estimator with

$$Q_K\left(r,s\right)=\frac{1}{K}\sum_{j=1}^{K}\Phi_j\left(r\right)\Phi_j\left(s\right), \tag{3}$$

where $\{\Phi_j(r)\}$ are orthonormal basis functions on $L^2[0,1]$ satisfying $\int_0^1 \Phi_j(r)\,dr = 0$. In the econometric literature, the series LRV estimator has been recently used, for example, in Phillips (2005), Müller (2007), and Sun (2011, 2013, 2014a&b).

Define the projection coefficient

$$\Lambda_j(\theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_j(\frac{t}{T}) \left[f(v_t, \theta_0) - \frac{1}{T} \sum_{\tau=1}^{T} f(v_\tau, \theta_0) \right] \text{ for } j = 1, 2, \ldots, K.$$

Then

$$W_T(\theta_0) = \frac{1}{K} \sum_{j=1}^{K} \Lambda_j(\theta_0) \Lambda_j(\theta_0)'. \tag{4}$$

In essence, each outer product $\Lambda_j(\theta_0)\Lambda_j(\theta_0)'$ is an approximately unbiased estimator of the LRV, and the series LRV estimator is a simple average of these estimators. Here K is the smoothing parameter underlying the series LRV estimator W_T. If K is even and $\{\Phi_j(r)\} = \{\sqrt{2}\sin(2\pi k r), \sqrt{2}\cos(2\pi k r), k = 1, 2, \ldots, K/2\}$, then the series LRV estimator is proportional to the spectral density estimator at the origin that takes a simple average of the first $K/2$ periodograms. The averaged periodogram estimator is a common spectral density estimator. In the traditional asymptotic framework, it can be shown that the averaged periodogram estimator is asymptotically equivalent to the kernel LRV estimator based on the Daniell kernel; see for example Phillips (2005). Sun (2013) provides more discussion on the relationship between the kernel LRV and series LRV estimators. To ensure that W_T is positive semidefinite, we assume that $K \geq m$ throughout the rest of the paper.

With the optimal weighting matrix estimator $W_T(\tilde{\theta}_T)$, the two-step GMM estimator is:

$$\hat{\theta}_T = \arg\min_{\theta \in \Theta} g_T(\theta)' W_T^{-1}(\tilde{\theta}_T) g_T(\theta).$$

Suppose that we want to perform hypothesis testing based on $\hat{\theta}_T$. Without loss of generality, we consider the linear null hypothesis $H_0 : R\theta_0 = r$ against the alternative $H_1 : R\theta_0 \neq r$ where R is a $p \times d$ matrix with full row rank. Nonlinear restrictions can be converted into linear ones by the Delta method with no consequence on the asymptotics we developed here. As in Sun (2014b), we consider the "trinity" of test statistics in the GMM setting. The first test statistic is the (normalized) Wald statistic given by

$$\mathbb{W}_T := \mathbb{W}_T(\hat{\theta}_T) = T(R\hat{\theta}_T - r)' \left\{ R \left[G_T(\hat{\theta}_T)' W_T^{-1}(\hat{\theta}_T) G_T(\hat{\theta}_T) \right]^{-1} R' \right\}^{-1} (R\hat{\theta}_T - r)/p, \tag{5}$$

where $G_T(\theta) = \frac{\partial g_T(\theta)}{\partial \theta'}$. When $p = 1$ and for one-sided alternative hypotheses, we can construct the t statistic:

$$\mathbb{T}_T := \mathbb{T}_T(\hat{\theta}_T) = \frac{\sqrt{T}(R\hat{\theta}_T - r)}{\{R[G_T(\hat{\theta}_T)'W_T^{-1}(\hat{\theta}_T)G_T(\hat{\theta}_T)]^{-1}R'\}^{1/2}}.$$

The second test statistic is the GMM criterion function statistic, which can be regarded as the LR analogue in the GMM setting. Let $\hat{\theta}_{T,R}$ be the restricted second-step GMM estimator:

$$\hat{\theta}_{T,R} = \arg\min_{\theta \in \Theta} g_T(\theta)' W_T^{-1}(\tilde{\theta}_T) g_T(\theta) \;\; s.t. \; R\theta = r.$$

The GMM criterion function statistic is

$$\mathbb{D}_T := \left[T g_T(\hat{\theta}_{T,R})' W_T^{-1}(\tilde{\theta}_T) g_T(\hat{\theta}_{T,R}) - T g_T(\hat{\theta}_T)' W_T^{-1}(\tilde{\theta}_T) g_T(\hat{\theta}_T) \right] / p,$$

which is often referred to as the quasi LR statistic.

The third test statistic is the GMM counterpart of the score or LM statistic. Let $\Delta_T(\theta) = G_T(\theta)' W_T^{-1}(\tilde{\theta}_T) g_T(\theta)$ be the gradient of the GMM criterion function. The score type test statistic is

$$\mathbb{S}_T = T \left[\Delta_T(\hat{\theta}_{T,R}) \right]' \left[G_T(\hat{\theta}_{T,R})' W_T^{-1}(\tilde{\theta}_T) G_T(\hat{\theta}_{T,R}) \right]^{-1} \Delta_T(\hat{\theta}_{T,R}) / p.$$

In the definitions of $\mathbb{D}_T$ and $\mathbb{S}_T$, $\tilde{\theta}_T$ can be replaced by $\hat{\theta}_T$ or any other $\sqrt{T}$ consistent estimator without affecting our asymptotic results.

To introduce the modified versions of the above three test statistics, we construct the standard J statistic for testing the over-identifying restrictions:

$$\mathbb{J}_T := \mathbb{J}_T(\hat{\theta}_T) = T g_T(\hat{\theta}_T)' W_T^{-1}(\hat{\theta}_T) g_T(\hat{\theta}_T).$$

The modified versions of $\mathbb{W}_T, \mathbb{D}_T$, and $\mathbb{S}_T$ are

$$\mathbb{W}_T^c := \frac{K - p - q + 1}{K} \frac{\mathbb{W}_T}{1 + \mathbb{J}_T / K},$$
$$\mathbb{D}_T^c := \frac{K - p - q + 1}{K} \frac{\mathbb{D}_T}{1 + \mathbb{J}_T / K},$$
$$\mathbb{S}_T^c := \frac{K - p - q + 1}{K} \frac{\mathbb{S}_T}{1 + \mathbb{J}_T / K}.$$

The multiplicative modifications are the same for all three statistics. The corresponding version of the t statistic is

$$\mathbb{T}_T^c := \sqrt{\frac{K - q}{K}} \frac{\mathbb{T}_T}{\sqrt{1 + \mathbb{J}_T / K}}.$$

Under the conventional asymptotic theory where K diverges to ∞ with the sample size T but $K/T \to 0$, all of $\mathbb{W}_T, \mathbb{D}_T$ and $\mathbb{S}_T$ and hence $\mathbb{W}_T^c, \mathbb{D}_T^c$ and $\mathbb{S}_T^c$ are asymptotically χ_p^2/p distributed. It is now well known that the chi-square approximation is not accurate in finite samples. This motivates the more accurate fixed-smoothing asymptotics under which K is held fixed as $T \to \infty$. We point out in passing that the fixed-K specification is an asymptotic device to help establish a more accurate approximation. We do not have to use a fixed K value in finite samples.

3 Asymptotics under the Null

Define

$$G_t(\theta) = \frac{\partial g_t(\theta)}{\partial \theta'} = \frac{1}{T}\sum_{j=1}^{t} \frac{\partial f(v_j, \theta)}{\partial \theta'} \text{ for } t \geq 1.$$

Let

$$u_t = f(v_t, \theta_0),\ \Phi_0(t) := 1, \text{ and } e_t \sim iidN(0, I_m).$$

We make the following assumptions on the basis functions, the GMM estimators, and the data generating process. These assumptions are the same as those in Sun (2014b) and are commonly used in the literature on the fixed-smoothing asymptotics.

Assumption 1 *For $j = 1, 2, \ldots, K$, the basis functions $\Phi_j(\cdot)$ are piecewise monotonic, continuously differentiable and orthonormal in $L^2[0,1]$ and $\int_0^1 \Phi_j(x)\,dx = 0$.*

Assumption 2 *As $T \to \infty$, $\hat{\theta}_T = \theta_0 + o_p(1)$, $\tilde{\theta}_T = \theta_0 + o_p(1)$ for an interior point $\theta_0 \in \Theta$ where $\Theta \subseteq \mathbb{R}^d$ is a parameter space of interest.*

Assumption 3 $\sum_{j=-\infty}^{\infty} \|\Gamma_j\| < \infty$ *where* $\Gamma_j = Eu_t u_{t-j}'$.

Assumption 4 *(a) $f(v_t, \theta)$ is twice continuously differentiable in θ for almost all v_t. (b) For any $\theta_T = \theta_0 + o_p(1)$, $G_{[rT]}(\theta_T) = rG + o_p(1)$ uniformly in r where $G = G(\theta_0)$ has rank d and $G(\theta) = E\partial f(v_t, \theta)/\partial\theta'$.*

Assumption 5 *(a) $T^{-1/2}\sum_{t=1}^{T} \Phi_j(t/T)\, u_t$ converges weakly to a continuous distribution, jointly over $j = 0, 1, \ldots, K$. (b) The following holds:*

$$P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \Phi_j\left(\frac{t}{T}\right) u_t \leq x \textit{ for } j = 0, 1, \ldots, K\right)$$
$$= P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \Phi_j\left(\frac{t}{T}\right) \Lambda e_t \leq x \textit{ for } j = 0, 1, \ldots, K\right) + o(1) \textit{ as } T \to \infty$$

where $x \in \mathbb{R}^m$ *and* Λ *is the matrix square root of* Ω*, i.e.,* $\Lambda\Lambda' = \Omega := \sum_{j=-\infty}^{\infty} \Gamma_j$*. (c)* Ω *is of full rank.*

Let

$$B_{p+q}(r) := \left(B_p(r)',\ B_q(r)'\right)',$$

where $B_p(r)$ and $B_q(r)$ are independent standard Brownian motion processes of dimensions p and q, respectively. Denote

$$C_{pp} = \int_0^1 \int_0^1 Q_K(r,s) dB_p(r) dB_p(s)' \tag{6}$$

and, for $q > 0$, let

$$C_{pq} = \int_0^1 \int_0^1 Q_K(r,s) dB_p(r) dB_q(s)', C_{qq} = \int_0^1 \int_0^1 Q_K(r,s) dB_q(r) dB_q(s)' \tag{7}$$

and $D_{pp} := C_{pp} - C_{pq} C_{qq}^{-1} C_{pq}'$.

Theorem 1 *Let Assumptions 1–5 hold and* $q > 0$*. Then, for a fixed* K*, the following weak convergence results hold jointly as* $T \to \infty$ *:*

(a) $\mathbb{W}_T, \mathbb{D}_T, \mathbb{S}_T \xrightarrow{d} \left[B_p(1) - C_{pq} C_{qq}^{-1} B_q(1)\right]' D_{pp}^{-1} \left[B_p(1) - C_{pq} C_{qq}^{-1} B_q(1)\right] / p :\overset{d}{=} \mathcal{F}_\infty,$

(b) $\mathbb{T}_T \xrightarrow{d} \left[B_p(1) - C_{pq} C_{qq}^{-1} B_q(1)\right] / \sqrt{D_{pp}} :\overset{d}{=} \mathcal{T}_\infty,$

(c) $\mathbb{J}_T \xrightarrow{d} B_q(1)' C_{qq}^{-1} B_q(1) :\overset{d}{=} \mathcal{J}_\infty,$

where $(B_p(1)', B_q(1)')'$ *is independent of* (C_{pq}, C_{qq}, D_{pp}) *and* D_{pp} *is independent of* (C_{pq}, C_{qq}) *.*

Remark 1 *If* $Q_K(\cdot,\cdot)$ *in (3) is replaced by a kernel function, then under some conditions on the kernel function, Theorem 1 also holds. A key advantage of using the series LRV estimator is that*

$$K \begin{bmatrix} C_{pp} & C_{pq} \\ C_{pq}' & C_{qq} \end{bmatrix} = \sum_{j=1}^{K} \left[\int_0^1 \Phi_j(r) dB_{p+q}(r)\right] \left[\int_0^1 \Phi_j(r) dB_{p+q}(r)\right]'$$

follows a standard Wishart distribution $\mathcal{W}_{p+q}(K, I_{p+q})$*. A well-known property of a Wishart random matrix is that* $D_{pp} = C_{pp} - C_{pq} C_{qq}^{-1} C_{pq}' \sim \mathcal{W}_p(K-q, I_p)/K$*. The fact that* D_{pp} *follows a (scaled) Wishart distribution and its independence of* (C_{pq}, C_{qq}) *are the two key properties of* D_{pp} *that drive our* F *and* t *limit theory. For kernel LRV estimation,* D_{pp} *will be neither Wishart nor independent of* (C_{pq}, C_{qq})*. So an exact* F *or* t *limit theory is not possible.*

Remark 2 *When $q=0$, we have $\mathbb{J}_T=0$, and the J-statistic correction factor $(1+\mathbb{J}_T/K)^{-1}$ reduces to unity. In this case, it is not difficult to show that*

$$\mathbb{W}_T, \mathbb{D}_T, \mathbb{S}_T \xrightarrow{d} \frac{1}{p} B_p(1)' C_{pp}^{-1} B_p(1) \overset{d}{=} \frac{K}{K-p+1} \mathcal{F}_{p,K-p+1},$$
$$\mathbb{T}_T \xrightarrow{d} B_p(1) / \sqrt{C_{pp}} \overset{d}{=} \mathcal{T}_K \text{ for } p=1.$$

These are identical to what are established in Sun (2013) for the Wald and t tests based on the first-step estimator. This is expected, as when $q=0$, the optimal weighting matrix becomes irrelevant, and the first-step estimator and two-step estimator become numerically identical.

Remark 3 *It follows from Theorem 1(c) that*

$$\frac{1}{K}\mathbb{J}_T(\hat{\theta}_T) \xrightarrow{d} \frac{q}{K-q+1}\mathcal{F}_{q,K-q+1} \overset{d}{=} \frac{\chi_q^2}{\chi_{K-q+1}^2} \tag{8}$$

for two independent chi-square random variables χ_q^2 and χ_{K-q+1}^2. See also Sun and Kim (2012). So, as K increases for a fixed q, $\mathbb{J}_T/K$ approaches zero and the modified GMM test statistics become close to the original GMM statistics. The multiplicative correction $(1+\mathbb{J}_T/K)^{-1}$ can be regarded as a finite sample correction under the conventional increasing-smoothing asymptotics. For the same reason, the other multiplicative correction $(K-p-q+1)/K$ can be regarded as a finite sample correction, as $(K-p-q+1)/K \to 1$ as $K \to \infty$. This correction factor can be motivated from the Bartlett correction. Sun (2013) provides detailed discussions.

On the basis of Theorems 1(a) and (b), we can construct asymptotically valid F_∞ and T_∞ tests. For example, for the Wald type test with the significance level α, the F_∞ test rejects the null if the unmodified Wald test statistic, $\mathbb{W}_T$, is greater than the critical value $\mathcal{F}_\infty^{1-\alpha}$, the $(1-\alpha)$ quantile of the limiting distribution $\mathcal{F}_\infty$. The reference distribution $\mathcal{F}_\infty$ is not standard, and the critical value $\mathcal{F}_\infty^{1-\alpha}$ has to be obtained by simulation.

To avoid having to simulate $\mathcal{F}_\infty$, we note that $\Delta := C_{pq}C_{qq}^{-1}B_q(1)$ is independent of $B_p(1)$ and D_{pp}. So, conditional on Δ,

$$\frac{K-p-q+1}{K}\mathcal{F}_\infty \overset{d}{=} \frac{K-p-q+1}{K}\frac{[B_p(1)-\Delta]' D_{pp}^{-1}[B_p(1)-\Delta]}{p} \overset{d}{=} \mathcal{F}_{p,K-p-q+1}(\|\Delta\|^2).$$

Unconditionally, $[(K-p-q+1)/K]\mathcal{F}_\infty$ follows a mixed noncentral F distribution, i.e., a noncentral F distribution with a random noncentrality parameter. Approximating $\|\Delta\|^2$ by $\delta^2 := E\|\Delta\|^2$, we obtain the noncentral F approximation $\mathcal{F}_{p,K-p-q+1}(\delta^2)$, which is the basis of the noncentral F test proposed in Sun (2014b). More specifically, the noncentral F test employs an unmodified test statistic and uses the critical value

$$\frac{K}{K-p-q+1}\mathcal{F}_{p,K-p-q+1}^{1-\alpha}\left(\delta^{2}\right)\text{ for }\delta^{2}=E\left\Vert \Delta\right\Vert ^{2}=\frac{pq}{K-q+1}$$

where $\mathcal{F}_{p,K-p-q+1}^{1-\alpha}\left(\delta^{2}\right)$ is the $1-\alpha$ quantile of the noncentral F distribution $\mathcal{F}_{p,K-p-q+1}\left(\delta^{2}\right)$. The noncentral F test is convenient, but the above critical value is an approximation to the exact nonstandard critical value $\mathcal{F}_{\infty}^{1-\alpha}$.

Using Theorem 1 and the continuous mapping theorem, we have:

$$\begin{aligned}\mathbb{W}_{T}^{c}(\hat{\theta}_{T}) &= \frac{K-p-q+1}{K}\frac{\mathbb{W}_{T}(\hat{\theta}_{T})}{1+\mathbb{J}_{T}(\hat{\theta}_{T})/K}\\ &\overset{d}{\to}\frac{K-p-q+1}{K}\frac{\mathcal{F}_{\infty}}{1+\mathcal{J}_{\infty}/K}=\frac{K-p-q+1}{pK}\xi_{p}^{\prime}D_{pp}^{-1}\xi_{p},\end{aligned}$$

where ξ_p is defined by

$$\xi_{p}:=\frac{B_{p}\left(1\right)-C_{pq}C_{qq}^{-1}B_{q}\left(1\right)}{\sqrt{1+\mathcal{J}_{\infty}/K}}=\frac{B_{p}\left(1\right)-C_{pq}C_{qq}^{-1}B_{q}\left(1\right)}{\sqrt{1+B_{q}(1)^{\prime}C_{qq}^{-1}B_{q}(1)/K}}. \tag{9}$$

Another key result that drives the F and t limit theory is that $\xi_p\sim N(0,I_p)$. To show this, we note that conditional on $B_q\left(\cdot\right):=\left\{B_q\left(r\right):r\in[0,1]\right\}$, $B_p\left(1\right)$ and C_{pq} are independent normals. They are conditionally independent because C_{pq} depends on $B_p\left(\cdot\right)$ only through $\int_0^1\Phi_k\left(r\right)dB_p\left(r\right)$ for $k=1,\ldots,K$ and each of these variates is conditionally independent of $B_p\left(1\right)$, as their conditional covariance is

$$cov\left[\int_{0}^{1}\Phi_{k}\left(r\right)dB_{p}\left(r\right),B_{p}\left(1\right)\right]=cov\left[\int_{0}^{1}\Phi_{k}\left(r\right)dB_{p}\left(r\right),\int_{0}^{1}dB_{p}\left(r\right)\right]=\int_{0}^{1}\Phi_{k}\left(r\right)dr=0.$$

Hence conditional on $B_q\left(\cdot\right)$, we have

$$B_{p}\left(1\right)-C_{pq}C_{qq}^{-1}B_{q}\left(1\right)\sim N\left(0,I_{p}+E\left[C_{pq}C_{qq}^{-1}B_{q}\left(1\right)B_{q}\left(1\right)^{\prime}C_{qq}^{-1}C_{qp}\right|\ B_{q}\left(\cdot\right)\right]\right). \tag{10}$$

In the appendix, we show that

$$E\left[C_{pq}C_{qq}^{-1}B_{q}\left(1\right)B_{q}\left(1\right)^{\prime}C_{qq}^{-1}C_{qp}\right|\ B_{q}\left(\cdot\right)\right]=I_{p}\cdot B_{q}\left(1\right)^{\prime}C_{qq}^{-1}B_{q}\left(1\right)/K. \tag{11}$$

So conditional on $B_q\left(\cdot\right)$, $\xi_p\sim N(0,I_p)$. Given that the conditional distribution $N(0,I_p)$ does not depend on the conditioning variable, ξ_p is independent of $B_q\left(\cdot\right)$, and $N(0,I_p)$ is also the unconditional distribution of ξ_p. Combining this with the independence of ξ_p from D_{pp}, we conclude that $p\mathcal{F}_\infty / (1+\mathcal{J}_\infty/K) = \xi_p' D_{pp}^{-1}\xi_p$ follows Hotelling's T^2 distribution. Using the relationship between the T^2 distribution and the standard F distribution, we obtain Part (a) of Theorem 2. Other parts can be similarly obtained. In particular, Parts (b) and (c) follow because, as shown by Sun (2014b), the asymptotic equivalence of $\mathbb{W}_T$, $\mathbb{D}_T$, and $\mathbb{S}_T$ continues to hold under the fixed-smoothing asymptotics.

Theorem 2 *Let Assumptions 1–5 hold. Then, for a fixed K as $T\to\infty$, we have:*

(a) $\mathbb{W}_T^c \overset{d}{\to} \mathcal{F}_{p,K-p-q+1}$;

(b) $\mathbb{D}_T^c \overset{d}{\to} \mathcal{F}_{p,K-p-q+1}$;

(c) $\mathbb{S}_T^c \overset{d}{\to} \mathcal{F}_{p,K-p-q+1}$;

(d) $\mathbb{T}_T^c \overset{d}{\to} \mathcal{T}_{K-q}$.

Remark 4 *$\xi_p \sim N(0,I_p)$ holds only for the case of series LRV estimation but not for the case of kernel LRV estimation. The conditioning argument we use to establish this result is different from that in Sun (2014b): the conditioning variable here is $B_q(\cdot)$, which can be reduced to $\mathcal{J}_\infty$, while the conditioning variable in Sun (2014b) is $\Delta = C_{pq}C_{qq}^{-1}B_q(1)$. Different conditioning strategies lead to different distributional approximations.*

Remark 5 *Since ξ_p is independent of $B_q(\cdot)$, ξ_p is independent of both $B_q(1)$ and C_{qq}. Note that D_{pp} is also independent of $B_q(1)$ and C_{qq}. So $\xi_p' D_{pp}^{-1}\xi_p$ is independent of $B_q(1)' C_{qq}^{-1} B_q(1)$. Now*

$$\begin{aligned}\mathcal{F}_\infty &\overset{d}{=} \frac{1}{p}\left(\xi_p' D_{pp}^{-1}\xi_p\right)\left[1+\frac{1}{K}B_q(1)'C_{qq}^{-1}B_q(1)\right] \\ &\overset{d}{=} \frac{K}{K-p-q+1}\cdot\mathcal{F}_{p,K-p-q+1}\cdot\left(1+\frac{1}{K}\mathcal{J}_\infty\right) \\ &\overset{d}{=} \frac{K}{K-p-q+1}\cdot\mathcal{F}_{p,K-p-q+1}\cdot\left(1+\frac{q}{K-q+1}\mathcal{F}_{q,K-q+1}\right)\end{aligned} \tag{12}$$

where $\mathcal{F}_{p,K-p-q+1}$ is independent of $\mathcal{F}_{q,K-q+1}$. This gives another characterization of the nonstandard limiting distribution developed by Sun (2014b). It can be used to simplify the simulation of the nonstandard distribution $\mathcal{F}_\infty$.

Theorem 2 allows us to perform asymptotically valid F and t tests. Consider the Wald type test as an example. The level-α F test rejects the null if the modified Wald statistic is larger than $\mathcal{F}_{p,K-p-q+1}^{1-\alpha}$, the $(1-\alpha)$ quantile of the F distribution $\mathcal{F}_{p,K-p-q+1}$. For easy reference and to contrast it with the nonstandard F_∞ test, we call the test the standard F test. It is the same as the test based on the original test statistic $\mathbb{W}_T$ but with the modified critical value

$$\left[1+\frac{1}{K}\mathbb{J}_T\right]\left[\frac{K}{K-p-q+1}\right]\mathcal{F}_{p,K-p-q+1}^{1-\alpha}. \tag{13}$$

Compared with the chi-square critical value $\chi_p^{1-\alpha}/p$ where $\chi_p^{1-\alpha}$ is the $(1-\alpha)$ quantile of the chi-squared distribution χ_p^2, the above critical value is larger for three reasons. First, $\mathcal{F}_{p,K-p-q+1}^{1-\alpha} > \chi_p^{1-\alpha}/p$ due to the random denominator in the F distribution. Second, $K/(K-p-q+1) > 1$ for $q > 0$ or $p > 1$. Third, $1+\mathbb{J}_T/K > 1$ almost surely for $q > 0$. A direct implication is that the chi-square critical value is too small, especially when q is large and K is relatively small. The small value of K can be empirically very relevant, as the moment process in economic applications often has high autocorrelation (e.g., Müller, 2014), which calls for a small value of K. Using the chi-square critical value can therefore lead to the finding of statistical significance that does not actually exist.

If we use the kernel LRV estimator, then we can choose an equivalent K value and use the critical value in (13). According to Sun and Kim (2012), the equivalent K value is given by the integer that is closest to

$$\frac{\left[\int_0^1 k_b(r,r)\,dr\right]^2}{\int_0^1\int_0^1 [k_b(r,s)]^2\,drds}, \tag{14}$$

where

$$k_b(t,\tau) = k(\frac{t-\tau}{b}) - \int_0^1 k(\frac{s-\tau}{b})ds - \int_0^1 k(\frac{t-s}{b})ds + \int_0^1\int_0^1 k(\frac{r-s}{b})drds,$$

$b = M/T$ for the truncation lag parameter M, and $k(\cdot)$ is the kernel function used in the LRV estimation. This procedure can be justified under the conventional asymptotics under which $b \to 0$, $bT \to \infty$ as $T \to \infty$, as in this case, the equivalent K value approaches ∞ and the critical value in (13) approaches the chi-squared critical value $\chi_p^{1-\alpha}/p$. In fact, as $b \to 0$, we can take

$$K = \left\lceil \frac{1}{b\int_{-\infty}^{\infty} k^2(x)dx} \right\rceil,$$

which provides a good approximation to (14). Here $\int_{-\infty}^{\infty} k^2(x)dx = 2/3, 0.54$, and 1 for the Bartlett, Parzen, and the quadratic spectral kernels, respectively. However, under the fixed-b asymptotics, the standard F distribution is not the exact limiting distribution. So, strictly speaking, we cannot justify this procedure under the fixed-b asymptotics. For this reason, one may argue that we should just simulate the nonstandard distribution and use the exact nonstandard critical value. However, the approximate critical value in (13) with an equivalent K is convenient to use and may be more appealing in applied research.

To compare the standard F test with the nonstandard F_∞ test, we note that, using (12):

$$\begin{aligned}
\alpha &= P\left(\mathcal{F}_\infty > \mathcal{F}_\infty^{1-\alpha}\right) \\
&= P\left[\frac{K}{K-p-q+1}\mathcal{F}_{p,K-p-q+1}\left(1+\frac{1}{K}\mathcal{J}_\infty\right) > \mathcal{F}_\infty^{1-\alpha}\right] \\
&= 1 - EG_{p,K-p-q+1}\left(\frac{\mathcal{F}_\infty^{1-\alpha}}{1+\mathcal{J}_\infty/K}\frac{K-p-q+1}{K}\right),
\end{aligned}$$

where $G_{p,K-p-q+1}(\cdot)$ denotes the CDF of $\mathcal{F}_{p,K-p-q+1}$. That is, the asymptotic level of the nonstandard F_∞ test is α when averaging over all realizations of $\mathcal{J}_\infty$. Conditional on $\mathcal{J}_\infty$, the asymptotic level is

$$1 - G_{p,K-p-q+1}\left(\frac{\mathcal{F}_\infty^{1-\alpha}}{1+\mathcal{J}_\infty/K}\frac{K-p-q+1}{K}\right),$$

which is strictly increasing in $\mathcal{J}_\infty$. So when the J statistic is large, which implies a large value of $\mathcal{J}_\infty$ in large samples, the nonstandard F_∞ test is expected to reject the null more often. In contrast, for the standard F test, the critical value in (13) is random and depends on the J statistic. It thus can be regarded as a conditional critical value. With a critical value that is tailored to the realized J statistic, the asymptotic conditional level of the standard F test is fixed at α regardless of the value of the J statistic.

One may wonder why the original GMM Wald statistic $\mathbb{W}_T$ ignores the J-statistic correction factor. The reason is that the underlying variance estimator is based on the conventional "sandwich" formula, which is derived under the conventional increasing-smoothing asymptotics where $K \to \infty$ as $T \to \infty$. Under this type of asymptotics, the asymptotic variance of the two-step GMM estimator is $\Omega_{11\cdot 2}$, and so $\hat{\Omega}_{11\cdot 2}$ is a natural estimator of the asymptotic variance. No J-statistic correction seems to be needed. However, under the fixed-smoothing asymptotics, we have:

$$\begin{aligned}
\sqrt{T}(\hat{\theta}_{T,GMM} - \theta_0) &= \begin{pmatrix} I_p, & -\hat{\beta} \end{pmatrix} \begin{pmatrix} \sum_{t=1}^T (y_{1t} - Ey_{1t})/\sqrt{T} \\ \sum_{t=1}^T y_{2t}/\sqrt{T} \end{pmatrix} \\
&\overset{d}{\to} (\Omega_{1\cdot 2})^{1/2}\left[B_p(1) - \tilde{\beta}_\infty B_q(1)\right].
\end{aligned}$$

where $\tilde{\beta}_\infty = C_{pq}C_{qq}^{-1}$.

Some simple calculations show that the conditional asymptotic variance[1] of $\hat{\theta}_{T,GMM}$ conditional on $B_q(\cdot)$ satisfies

$$\text{avar}(\hat{\theta}_{T,GMM}\ |B_q(\cdot)) = \Omega_{11\cdot 2}\left(1 + \frac{\mathcal{J}_\infty}{K}\right).$$

When we use the conventional "sandwich" formula for variance estimation, which attempts to estimate $\Omega_{11\cdot 2}$ only, we effectively ignore the term that involves $\mathcal{J}_\infty$. This will not cause any problem for asymptotic pivotal inference but will prevent us from developing the F limit theory. The modification we propose can be regarded as the multiplicative variance correction that takes into account the extra asymptotic variance term under the fixed-smoothing asymptotics. More specifically, instead of using $\hat{\Omega}_{11\cdot 2}$, we use $\hat{\Omega}_{11\cdot 2}(1+\mathbb{J}_T/K)$, a sample analogue of $\Omega_{11\cdot 2}(1+\mathcal{J}_\infty/K)$, as the asymptotic variance estimator. This gives rise to the modified statistic $\mathbb{W}_T^c$.

4 Asymptotics under Local Alternatives

We examine the asymptotic distributions of the GMM test statistics under the local alternatives

$$H_1 : R\theta_0 = r + \delta_0/\sqrt{T}, \tag{15}$$

where δ_0 is a $p \times 1$ non-zero vector that describes the local departure from the null hypothesis. Under H_1, we have a triangular array data generating process. For the sake of notational convenience, we suppress the dependence of the model parameter θ_0 on the sample size T.

Let $\chi_p^2(\lambda)$ be the noncentral chi-square distribution with noncentrality parameter λ. Under the conventional increasing-smoothing asymptotics, it is not difficult to show that under the local alternative hypotheses:

$$\mathbb{W}_T, \mathbb{D}_T, \mathbb{S}_T \xrightarrow{d} \chi_p^2(\|\delta\|^2)/p \text{ and } \mathbb{T}_T \xrightarrow{d} N(\delta, 1),$$

where $\delta = \Lambda_R^{-1}\delta_0$ is the scaled local deviation parameter, and the scaling matrix Λ_R is a matrix square root of the asymptotic variance of $R\hat{\theta}_T$, that is, $\Lambda_R\Lambda_R' = R(G'\Omega^{-1}G)^{-1}R'$. See, for example, Theorem 5.7 in Hall (2005) for detailed derivation.

As K increases, both the J-statistic correction and the multiplicative Bartlett correction disappear. So, under the conventional increasing-smoothing asymptotics, the limiting null distributions of the original GMM statistics and the modified statistics are identical. The limiting

[1] The conditional asymptotic variance refers to the variance of the asymptotic distribution of $\sqrt{T}(\hat{\theta}_{T,GMM} - \theta_0)$ conditional on some variate.

distributions under the local alternatives are also the same. However, under the fixed-smoothing asymptotics, they have different limiting distributions under the local alternatives, as they do under the null.

Let

$$\delta_J = \frac{\delta}{\sqrt{1+\mathcal{J}_\infty/K}} = \frac{\Lambda_R^{-1}\delta_0}{\sqrt{1+\mathcal{J}_\infty/K}},$$

$$\mathcal{F}_\infty(\|\delta_J\|^2) = \frac{K}{K-p-q+1}\cdot \mathcal{F}_{p,K-p-q+1}(\|\delta_J\|^2)\cdot\left[1+\frac{1}{K}\mathcal{J}_\infty\right],$$

and

$$\mathcal{T}_\infty(\delta_J) \overset{d}{=} \sqrt{\frac{K}{K-q}}\cdot \mathcal{T}_{K-q}(\delta_J)\cdot\sqrt{1+\frac{1}{K}\mathcal{J}_\infty},$$

where $\mathcal{F}_{p,K-p-q+1}(\|\delta_J\|^2)$ and $\mathcal{T}_{K-q}(\delta_J)$ are noncentral $\mathcal{F}_{p,K-p-q+1}$ and $\mathcal{T}_{K-q}$ distributions with random but independent noncentrality parameters $\|\delta_J\|^2$ and δ_J, respectively. In other words, $\mathcal{F}_{p,K-p-q+1}(\|\delta_J\|^2)$ and $\mathcal{T}_{K-q}(\delta_J)$ are mixed non-central F and t distributions. The following theorem establishes the asymptotic distributions of the original GMM test statistics and their modified versions under the fixed-smoothing asymptotics.

Theorem 3 *Let Assumptions 1–5 hold under the local alternatives in (15). Then, for a fixed K as $T\to\infty$, we have:*

(a) $\mathbb{W}_T, \mathbb{D}_T, \mathbb{S}_T \overset{d}{\to} \mathcal{F}_\infty(\|\delta_J\|^2)$;

(b) $\mathbb{T}_T \overset{d}{\to} \mathcal{T}_\infty(\delta_J)$;

(c) $\mathbb{W}_T^c, \mathbb{D}_T^c, \mathbb{S}_T^c \overset{d}{\to} \mathcal{F}_{p,K-p-q+1}(\|\delta_J\|^2)$;

(d) $\mathbb{T}_T^c \overset{d}{\to} \mathcal{T}_{K-q}(\delta_J)$.

It follows from Theorem 3 that the asymptotic local power function of the nonstandard F_∞ test is

$$\begin{aligned}\pi &= \pi(\|\delta\|^2; \alpha, K, p, q)\\ &= P\left[\frac{K}{K-p-q+1}\mathcal{F}_{p,K-p-q+1}(\|\delta_J\|^2)\left(1+\frac{1}{K}\mathcal{J}_\infty\right) > \mathcal{F}_\infty^{1-\alpha}\right] \\ &= P\left[\mathcal{F}_{p,K-p-q+1}(\|\delta_J\|^2) > \frac{\mathcal{F}_\infty^{1-\alpha}}{1+\mathcal{J}_\infty/K}\frac{K-p-q+1}{K}\right]\end{aligned} \quad (16)$$

and that the asymptotic local power function of the standard F test is

$$\begin{aligned}\pi^c &= \pi^c(\|\delta\|^2; \alpha, K, p, q) \\ &= P\left[\mathcal{F}_{p,K-p-q+1}(\|\delta_J\|^2) > \mathcal{F}_{p,K-p-q+1}^{1-\alpha}\right].\end{aligned} \quad (17)$$

So, conditional on $\mathcal{J}_\infty$, $\pi^c > \pi$ if and only if

$$\mathcal{F}_{p,K-p-q+1}^{1-\alpha} < \frac{\mathcal{F}_\infty^{1-\alpha}}{1+\mathcal{J}_\infty/K}\frac{K-p-q+1}{K}.$$

That is, conditional on $\mathcal{J}_\infty$, $\pi^c > \pi$ if and only if

$$\mathcal{J}_\infty < \frac{\mathcal{F}_\infty^{1-\alpha}}{\mathcal{F}_{p,K-p-q+1}^{1-\alpha}}(K-p-q+1) - K.$$

When the J statistic is small enough, we expect the standard F test to be (asymptotically and locally) more powerful than the nonstandard F_∞ test among the DGP's that yield more or less the same small value for the J statistic. We have shown before that the standard F test has the exact asymptotic level conditionally on the J statistic and unconditionally. In view of both size accuracy and power improvement, we recommend the standard F test when the J statistic is small. When the J statistic is large, the standard F test still has the exact asymptotic level, but the power advantage may not be present any longer. In this case, if we prefer a test with accurate asymptotic level both conditionally and unconditionally, then the standard F test is still our choice.

While the conditional power comparison is useful, we may also want to compare the unconditional power, which entails integrating the conditional power function with respect to the probability distribution of $\mathcal{J}_\infty$. In Tables 1–3, we report the two local asymptotic power functions $\pi(\|\delta\|^2; \alpha, K, p, q)$ and $\pi^c(\|\delta\|^2; \alpha, K, p, q)$ for $\alpha = 0.05$, $K = 8 \sim 24$, and $p, q = 1 \sim 3$ and selected values of the local deviation parameter $\|\delta\|$. Figures 1–3 plot the difference between the two power functions, $\pi^c(\cdot) - \pi(\cdot)$, for $p = 3$, $K = 8 \sim 24$, and $q = 1 \sim 3$. For most values of the local deviation parameter $\|\delta\|$ considered, the power of the standard F test dominates that of the nonstandard F_∞ test for $K = 8 \sim 12$. However, the power difference between the two tests is relatively small, mostly between -0.5 percentage point and 2 percentage points. The power advantage of the standard F test becomes more visible as the degree of overidentification q increases. For example, when (K, p, q) is equal to $(8, 3, 3)$, the power advantage can be as high as 5 percentage points while the power advantage is only around 2 percentage points when q is equal to 1. The power advantage of the standard F test decays and drops to zero when K increases and becomes larger than 24. This is consistent with our asymptotic theory.

5 Simulation Evidence

We follow Sun (2014b) and consider a linear model of the form:

$$y_t = x_t'\theta + \varepsilon_{y,t}$$

where $x_t := (1, x_{1,t}, x_{2,t}, x_{3,t})'$ is a vector of endogenous regressors. The unknown parameter vector is $\theta = (\gamma_0, \gamma_1, \ldots, \gamma_{d-1})' \in \mathbb{R}^d$. We have m instruments $z_{0,t}, z_{1,t}, \ldots, z_{m-1,t}$ with $z_{0,t} := 1$. The reduced-form equations for $x_{1,t}$, $x_{2,t}$ and $x_{3,t}$ are given by

$$x_{j,t} = z_{j,t} + \sum_{i=d}^{m-1} z_{i,t} + \varepsilon_{x_j,t} \text{ for } j = 1, \ldots, d-1.$$

We consider two different experiment designs: the autoregressive (AR) design and the centered moving average (CMA) design. In the AR design, each $z_{i,t}$ follows an AR(1) process of the form $z_{i,t} = \rho z_{i,t-1} + \sqrt{1-\rho^2} e_{z_i,t}$ where $e_{z_i,t} = \left(e_{zt}^i + e_{zt}^0\right)/\sqrt{2}$ and $e_t = [e_{zt}^0, e_{zt}^1, \ldots, e_{zt}^{m-1}]' \sim iidN(0, I_m)$. By construction, z_{it} has unit variance for all $i \geq 1$, and the correlation coefficient between the non-constant $z_{i,t}$ and $z_{j,t}$ for $i \neq j$ is 0.5. The DGP for $\varepsilon_t = (\varepsilon_{yt}, \varepsilon_{x_1 t}, \varepsilon_{x_2 t}, \varepsilon_{x_3 t})'$ is the same as that for $(z_{1,t}, \ldots, z_{m-1,t})$ except that there is a difference in the dimension. The two vector processes ε_t and $(z_{1,t}, \ldots, z_{m-1,t})$ are independent of each other. We take $\rho = \pm 0.95, \pm 0.85, \pm 0.50$, and 0.

In the CMA design, $\varepsilon_{y,t}$ is a scaled and centered moving average of an iid sequence $\varepsilon_{y,t} = \sum_{j=-L}^{L} e_{t+j}/\sqrt{2L+1}$ where $e_t \sim iidN(0,1)$ and L is the number of leads and lags in the average. The instruments are generated according to $z_{it} = [e_{t-L+i-1} - (2L+1)^{-1} \sum_{j=-L}^{L} e_{t+j}] \sqrt{(2L+1)/2L}$ for $i = 1, \ldots, m-1$. The error term in the reduced-form equation is given by $\varepsilon_{x_j,t} = \left(\varepsilon_{y,t} + e_{x_j,t}\right)/\sqrt{2}$ where $e_{x_j,t} \sim iidN(0,1)$ and is independent of the sequence $\{e_t\}$. We take $L = 3$, 6, and 9.

We consider $q = 0, 1, 2$ and $d = 4$ with the corresponding numbers of moment conditions $m = 4, 5, 6$. The null hypotheses of interest are

$$H_{01} : \gamma_1 = 0,$$

$$H_{02} : \gamma_1 = \gamma_2 = 0,$$

$$H_{03} : \gamma_1 = \gamma_2 = \gamma_3 = 0.$$

The numbers of joint hypotheses are $p = 1$, 2 and 3, respectively. We consider three different sample sizes $T = 100, 200, 500$ and two significance levels $\alpha = 5\%$ and $\alpha = 10\%$. We focus on the Wald type of test, but the simulation results are qualitatively similar for other types of tests.

We examine the empirical size of four different two-step tests. The first three tests are based on the same unmodified Wald test statistic, so they have the same size-adjusted power. The difference lies in the critical values used. We employ the following critical values: $\chi_p^{1-\alpha}/p$, $\frac{K}{K-p-q+1}\mathcal{F}_{p,K-p-q+1}^{1-\alpha}[pq/(K-q-1)]$, and $\mathcal{F}_{\infty}^{1-\alpha}$, leading to the χ^2 test, the NCF (noncentral F) test, and the nonstandard F_∞ test. The χ^2 test uses the conventional chi-square approximation. The NCF test uses the noncentral F approximation. The F_∞ test uses the nonstandard $\mathcal{F}_\infty$ approximation with simulated critical values. The fourth test is the standard F test proposed in this paper, which is based on the modified Wald statistic $\mathbb{W}_T^c$ and uses the standard F critical value $\mathcal{F}_{p,K-p-q+1}^{1-\alpha}$. Equivalently, our proposed test is based on the same unmodified Wald test statistic as the first three tests but uses the critical values given in (13). For each test, the initial first-step estimator is the IV estimator with weight matrix $W_{o,T} = Z'Z/T$ where Z is the matrix of instruments.

We use the basis functions $\Phi_{2j-1}(x) = \sqrt{2}\cos 2j\pi x,\ \Phi_{2j}(x) = \sqrt{2}\sin 2j\pi x,\ j = 1, \ldots, K/2$ and assume that K is even. In this case, the series LRV estimator can be computed using discrete Fourier transforms. We select K (and round it to an even number) based on the AMSE criterion implemented using the VAR(1) plug-in procedure in Phillips (2005), which is similar to the plug-in procedure of Andrews (1991). We compute the data-driven K on the basis of the initial first step estimator $\tilde{\theta}_T$ and use it in computing both $W_T(\tilde{\theta}_T)$ and $W_T(\hat{\theta}_T)$.

We also compare the size-adjusted power of the proposed standard F test with that of the nonstandard F_∞ test. The data is generated under the local alternative $H_1 : R\theta = c_0\ell_p/\sqrt{T}$ where c_0 is a scalar and ℓ_p is the p-vector of ones. The two tests use the same data-driven smoothing parameter K. To make the power comparison meaningful, we compute the power using the empirical finite sample critical values obtained from the null distribution. That is, we compare the size-adjusted power. It should be pointed out that size-adjustment is not feasible in practice.

Tables 4 and 5 report the finite sample size of the four tests for $T = 100$ and $\alpha = 5\%$. The number of simulation replications is 10,000. It is clear that all three tests, the standard F, nonstandard F_∞ , and noncentral F tests, perform very well when the processes are not strongly autocorrelated, say when ρ lies between $(-0.85, 0.85)$. Like the latter two tests, the standard F test is much more accurate in size than the conventional chi-square test, which can be highly size-distorted. These qualitative observations remain valid for other sample sizes and significance levels.

It should be pointed out that our fixed-smoothing approximation is established under a fixed data generating process where $|\rho|$ is strictly less than 1. The approximation is therefore a pointwise approximation with an approximation error that may not be uniformly small over all $\rho \in (-1, 1)$. In particular, the approximation error may not be small when $|\rho|$ is very close to 1, singular limiting scenarios pointed out by Preinerstorfer and Pötscher (2016)[2]. Table 4 indicates the size distortion of our standard F test can reach 23% when $|\rho| = 0.95$, $p = 3$ and $q = 1$. Interestingly, the fixed-smoothing asymptotic tests still outperform the conventional chi-square test whose size distortion can be as high as 65%.

Figures 4 and 5 report the size-adjusted power of the nonstandard F_∞ test and the standard F test for $\alpha = 5\%$ and $T = 100$. There is no real difference between the two power curves. In fact, the standard F test can be slightly more powerful in some scenarios, which is consistent with the local power analysis in Section 4. Note that the size-adjusted power of the nonstandard F_∞ test is the same as that of the conventional chi-square test, the standard F test is therefore as powerful as the conventional chi-square test.

Our finite sample simulation evidence lends strong support to the standard F test: it enjoys the same good size and power properties as the nonstandard F_∞ test, but it is easier to use, as the critical values are readily available from statistical tables, and no simulation or approximation is needed.

6 Conclusion

This paper has proposed a modification to the trinity of test statistics in an efficient two-step GMM framework. Each modified test statistic is a function of the original test statistic and the usual J statistic for testing overidentification. We show that the modified test statistics are all asymptotically F or t distributed. This leads to standard F and t tests that are based on

[2]We thank an anonymous referee for pointing this out. Note that the variance of the innovations in our DGP: $z_{i,t} = \rho z_{i,t-1} + \sqrt{1-\rho^2} e_{z_i,t}$ depends on ρ^2. As $\rho \to 1$ or -1, the DGP approaches the singularity points of Preinerstorfer and Pötscher (2016). It should be pointed out that for this DGP the limiting case under $\rho \to 1$ does not correspond to a unit-root process. To accommodate a unit root or near unit root process, we need to consider the DGP: $z_{i,t} = \rho z_{i,t-1} + e_{z_i,t}$. Under this alternative DGP, we find that the finite sample performance of the F tests relative to that of other tests are qualitatively similar. The difference is that '$\rho \to -1$' does not cause large size distortion for all tests.

the modified test statistics and use the standard F or t critical values. Simulation shows that the standard F tests have the same finite sample performance as the nonstandard tests recently proposed by Sun (2014b), but the standard F tests are much easier to use.

The paper complements Sun (2011, 2013, 2014a) which establish the F and t limit theory for the tests based on the first-step GMM estimation and the J test. When the series LRV estimator is used, the F and t limit theory appears to be applicable to all common tests in the first-step and two-step GMM settings.

There are a number of interesting extensions. The results of the paper can be easily extended to the continuous updating GMM (CU-GMM) framework. Recently, Zhang (2016) has shown that the Wald statistic based on the CU-GMM estimator has the same fixed-smoothing limit as what Sun (2014b) obtains in the two-step GMM framework. Given this, it is easy to see that our results hold without change if the CU-GMM estimator is used instead. Following the work of Bester, Conley, Hansen and Vogelsang (2016) and Sun and Kim (2015), we do not imagine much difficulty in extending our results to a spatial setting.

Table 1: Asymptotic local power of the nonstandard F_∞ and standard F tests ($p = 1$, $\alpha = 0.05$)

			$\|\delta\|^2$										
K	q	Tests	**0.5**	**1.0**	**2.0**	**3.0**	**5.0**	**7.0**	**10**	**15**	**21**	**25**	**30**
8	1	F_∞	0.091	0.123	0.207	0.286	0.435	0.557	0.715	0.864	0.950	0.978	0.990
		$F_{p,K-p-q+1}$	0.095	0.128	0.216	0.294	0.444	0.569	0.721	0.867	0.947	0.973	0.986
	2	F_∞	0.080	0.108	0.174	0.234	0.348	0.467	0.606	0.778	0.899	0.938	0.968
		$F_{p,K-p-q+1}$	0.084	0.115	0.189	0.252	0.374	0.494	0.634	0.792	0.895	0.929	0.959
	3	F_∞	0.074	0.096	0.140	0.184	0.277	0.377	0.494	0.654	0.806	0.859	0.908
		$F_{p,K-p-q+1}$	0.078	0.107	0.162	0.216	0.313	0.427	0.536	0.683	0.821	0.866	0.905
12	1	F_∞	0.094	0.141	0.234	0.328	0.490	0.630	0.788	0.922	0.977	0.992	0.996
		$F_{p,K-p-q+1}$	0.093	0.141	0.238	0.332	0.494	0.631	0.787	0.920	0.976	0.991	0.995
	2	F_∞	0.090	0.131	0.211	0.290	0.450	0.584	0.735	0.882	0.960	0.980	0.992
		$F_{p,K-p-q+1}$	0.095	0.133	0.214	0.301	0.460	0.597	0.741	0.883	0.958	0.977	0.990
	3	F_∞	0.080	0.121	0.188	0.257	0.394	0.524	0.671	0.837	0.935	0.963	0.983
		$F_{p,K-p-q+1}$	0.084	0.129	0.200	0.273	0.416	0.545	0.684	0.845	0.930	0.957	0.977
16	1	F_∞	0.098	0.151	0.249	0.347	0.524	0.672	0.821	0.937	0.985	0.996	0.998
		$F_{p,K-p-q+1}$	0.097	0.153	0.252	0.348	0.524	0.675	0.820	0.935	0.983	0.995	0.998
	2	F_∞	0.093	0.144	0.227	0.319	0.488	0.626	0.787	0.915	0.978	0.992	0.997
		$F_{p,K-p-q+1}$	0.095	0.146	0.233	0.325	0.492	0.628	0.794	0.913	0.974	0.991	0.997
	3	F_∞	0.091	0.134	0.216	0.294	0.452	0.591	0.743	0.895	0.965	0.983	0.995
		$F_{p,K-p-q+1}$	0.093	0.137	0.224	0.301	0.460	0.603	0.747	0.891	0.958	0.980	0.992
20	1	F_∞	0.100	0.152	0.256	0.357	0.537	0.685	0.830	0.950	0.988	0.997	1.000
		$F_{p,K-p-q+1}$	0.102	0.155	0.258	0.356	0.538	0.688	0.830	0.950	0.987	0.996	0.999
	2	F_∞	0.098	0.144	0.248	0.338	0.514	0.656	0.805	0.932	0.983	0.994	0.998
		$F_{p,K-p-q+1}$	0.098	0.145	0.252	0.342	0.518	0.658	0.808	0.931	0.980	0.993	0.998
	3	F_∞	0.095	0.136	0.229	0.317	0.481	0.633	0.775	0.921	0.977	0.989	0.997
		$F_{p,K-p-q+1}$	0.099	0.140	0.234	0.321	0.489	0.638	0.776	0.920	0.976	0.988	0.995
24	1	F_∞	0.110	0.151	0.257	0.368	0.556	0.692	0.840	0.950	0.989	0.997	0.999
		$F_{p,K-p-q+1}$	0.110	0.154	0.256	0.370	0.557	0.692	0.839	0.950	0.989	0.997	0.999
	2	F_∞	0.095	0.147	0.252	0.350	0.524	0.674	0.826	0.943	0.984	0.995	0.998
		$F_{p,K-p-q+1}$	0.095	0.150	0.255	0.352	0.527	0.675	0.825	0.944	0.983	0.994	0.998
	3	F_∞	0.092	0.145	0.244	0.344	0.513	0.648	0.806	0.932	0.982	0.993	0.998
		$F_{p,K-p-q+1}$	0.094	0.151	0.247	0.347	0.520	0.654	0.809	0.932	0.980	0.992	0.998

The table computes the asymptotic local power functions $\pi(\cdot)$ and $\pi^c(\cdot)$ of the nonstandard F_∞ test and the standard F test, respectively. The nonstandard F_∞ test employs the conventional statistic, i.e., $\mathbb{W}_T$, and the nonstandard critical value $\mathcal{F}_\infty^{1-\alpha}$. The standard F test employs the modified statistic, i.e., $\mathbb{W}_T^c$, and the standard F critical value $\mathcal{F}_{p,K-p-q+1}^{1-\alpha}$.

Table 2: Asymptotic local power of the nonstandard F_∞ and standard F tests ($p = 2$, $\alpha = 0.05$)

			$\|\delta\|^2$										
K	q	Tests	0.5	1.0	2.0	3.0	5.0	7.0	10	15	21	25	30
8	1	F_∞	0.075	0.093	0.139	0.188	0.285	0.372	0.512	0.682	0.836	0.890	0.936
		$F_{p,K-p-q+1}$	0.076	0.094	0.144	0.196	0.297	0.393	0.531	0.696	0.838	0.889	0.934
	2	F_∞	0.065	0.082	0.113	0.143	0.218	0.287	0.399	0.555	0.707	0.779	0.848
		$F_{p,K-p-q+1}$	0.066	0.086	0.121	0.157	0.242	0.320	0.436	0.588	0.729	0.799	0.856
	3	F_∞	0.059	0.071	0.093	0.115	0.167	0.214	0.290	0.403	0.541	0.603	0.679
		$F_{p,K-p-q+1}$	0.059	0.075	0.103	0.129	0.192	0.250	0.335	0.459	0.596	0.656	0.722
12	1	F_∞	0.074	0.105	0.163	0.226	0.346	0.480	0.631	0.816	0.926	0.965	0.983
		$F_{p,K-p-q+1}$	0.077	0.106	0.165	0.233	0.353	0.488	0.636	0.818	0.925	0.962	0.982
	2	F_∞	0.073	0.093	0.143	0.195	0.304	0.420	0.565	0.745	0.886	0.934	0.964
		$F_{p,K-p-q+1}$	0.074	0.097	0.151	0.208	0.317	0.434	0.583	0.757	0.886	0.931	0.961
	3	F_∞	0.066	0.090	0.135	0.168	0.262	0.361	0.492	0.674	0.823	0.877	0.939
		$F_{p,K-p-q+1}$	0.066	0.094	0.144	0.184	0.288	0.390	0.522	0.697	0.833	0.882	0.934
16	1	F_∞	0.083	0.108	0.172	0.248	0.388	0.514	0.680	0.854	0.950	0.976	0.992
		$F_{p,K-p-q+1}$	0.083	0.112	0.176	0.254	0.394	0.519	0.684	0.855	0.951	0.976	0.991
	2	F_∞	0.076	0.102	0.173	0.234	0.358	0.482	0.640	0.820	0.934	0.963	0.984
		$F_{p,K-p-q+1}$	0.076	0.103	0.179	0.239	0.369	0.491	0.650	0.820	0.931	0.960	0.984
	3	F_∞	0.076	0.101	0.152	0.203	0.321	0.434	0.594	0.783	0.903	0.943	0.976
		$F_{p,K-p-q+1}$	0.077	0.104	0.156	0.213	0.337	0.452	0.609	0.792	0.903	0.940	0.970
20	1	F_∞	0.085	0.109	0.186	0.253	0.406	0.549	0.711	0.883	0.966	0.985	0.995
		$F_{p,K-p-q+1}$	0.085	0.113	0.187	0.255	0.409	0.552	0.713	0.884	0.964	0.984	0.995
	2	F_∞	0.076	0.110	0.172	0.244	0.394	0.520	0.694	0.858	0.955	0.978	0.991
		$F_{p,K-p-q+1}$	0.077	0.111	0.179	0.248	0.399	0.529	0.697	0.858	0.952	0.977	0.989
	3	F_∞	0.079	0.097	0.169	0.229	0.363	0.500	0.656	0.821	0.934	0.969	0.986
		$F_{p,K-p-q+1}$	0.080	0.099	0.170	0.233	0.373	0.509	0.664	0.823	0.931	0.966	0.983
24	1	F_∞	0.082	0.123	0.197	0.273	0.425	0.570	0.733	0.898	0.972	0.988	0.996
		$F_{p,K-p-q+1}$	0.082	0.124	0.199	0.275	0.429	0.572	0.733	0.900	0.972	0.986	0.996
	2	F_∞	0.083	0.115	0.184	0.254	0.415	0.547	0.706	0.878	0.959	0.982	0.996
		$F_{p,K-p-q+1}$	0.083	0.117	0.188	0.259	0.418	0.551	0.711	0.876	0.959	0.983	0.995
	3	F_∞	0.079	0.115	0.175	0.258	0.384	0.514	0.680	0.857	0.955	0.977	0.991
		$F_{p,K-p-q+1}$	0.081	0.120	0.179	0.261	0.392	0.523	0.685	0.859	0.956	0.974	0.990

See footnotes to Table 1.

Table 3: Asymptotic local power of the nonstandard F_∞ and standard F tests ($p = 3$, $\alpha = 0.05$)

			$\|\delta\|^2$										
K	q	Tests	0.5	1.0	2.0	3.0	5.0	7.0	10	15	21	25	30
8	1	F_∞	0.061	0.075	0.106	0.134	0.195	0.270	0.365	0.506	0.665	0.747	0.814
		$F_{p,K-p-q+1}$	0.062	0.077	0.109	0.138	0.205	0.278	0.379	0.521	0.684	0.759	0.822
	2	F_∞	0.056	0.065	0.084	0.111	0.152	0.193	0.264	0.377	0.503	0.577	0.652
		$F_{p,K-p-q+1}$	0.058	0.066	0.090	0.116	0.165	0.212	0.292	0.410	0.534	0.609	0.678
	3	F_∞	0.057	0.060	0.067	0.086	0.115	0.146	0.185	0.244	0.323	0.373	0.432
		$F_{p,K-p-q+1}$	0.057	0.063	0.075	0.097	0.131	0.169	0.207	0.287	0.370	0.418	0.488
12	1	F_∞	0.069	0.086	0.132	0.181	0.284	0.365	0.514	0.707	0.843	0.907	0.947
		$F_{p,K-p-q+1}$	0.070	0.088	0.136	0.186	0.289	0.376	0.518	0.711	0.845	0.908	0.946
	2	F_∞	0.063	0.085	0.120	0.157	0.236	0.321	0.438	0.624	0.779	0.852	0.909
		$F_{p,K-p-q+1}$	0.066	0.089	0.128	0.167	0.251	0.341	0.462	0.641	0.790	0.859	0.912
	3	F_∞	0.062	0.076	0.099	0.132	0.201	0.259	0.364	0.526	0.691	0.760	0.836
		$F_{p,K-p-q+1}$	0.065	0.080	0.109	0.143	0.223	0.291	0.397	0.558	0.712	0.783	0.850
16	1	F_∞	0.066	0.090	0.146	0.193	0.312	0.422	0.583	0.775	0.899	0.947	0.975
		$F_{p,K-p-q+1}$	0.066	0.091	0.149	0.198	0.318	0.427	0.589	0.777	0.898	0.947	0.975
	2	F_∞	0.068	0.094	0.135	0.183	0.289	0.392	0.522	0.724	0.873	0.922	0.961
		$F_{p,K-p-q+1}$	0.069	0.095	0.138	0.188	0.297	0.404	0.530	0.735	0.874	0.922	0.958
	3	F_∞	0.065	0.080	0.124	0.169	0.250	0.348	0.489	0.668	0.829	0.893	0.944
		$F_{p,K-p-q+1}$	0.067	0.086	0.133	0.179	0.266	0.367	0.505	0.685	0.837	0.896	0.942
20	1	F_∞	0.071	0.099	0.153	0.217	0.326	0.463	0.628	0.815	0.927	0.963	0.988
		$F_{p,K-p-q+1}$	0.072	0.099	0.154	0.220	0.331	0.466	0.629	0.815	0.927	0.962	0.987
	2	F_∞	0.070	0.100	0.142	0.197	0.316	0.428	0.582	0.779	0.911	0.957	0.979
		$F_{p,K-p-q+1}$	0.072	0.101	0.148	0.205	0.324	0.437	0.591	0.785	0.913	0.956	0.978
	3	F_∞	0.069	0.089	0.138	0.183	0.292	0.391	0.545	0.740	0.881	0.937	0.970
		$F_{p,K-p-q+1}$	0.069	0.091	0.141	0.189	0.301	0.402	0.558	0.747	0.883	0.936	0.968
24	1	F_∞	0.079	0.102	0.155	0.233	0.359	0.488	0.636	0.828	0.945	0.973	0.989
		$F_{p,K-p-q+1}$	0.077	0.104	0.156	0.234	0.360	0.490	0.643	0.829	0.945	0.973	0.989
	2	F_∞	0.070	0.095	0.156	0.216	0.333	0.452	0.626	0.808	0.928	0.965	0.988
		$F_{p,K-p-q+1}$	0.071	0.097	0.158	0.218	0.342	0.460	0.628	0.813	0.926	0.966	0.987
	3	F_∞	0.071	0.097	0.142	0.203	0.316	0.441	0.590	0.787	0.914	0.952	0.980
		$F_{p,K-p-q+1}$	0.072	0.098	0.149	0.209	0.323	0.450	0.597	0.790	0.915	0.950	0.979

See footnotes to Table 1.

Table 4: Empirical size of the nominal 5% χ^2 test, noncentral F test, nonstandard F_∞ test and standard F test based on the series LRV estimator under the AR design with $T = 100$, number of joint hypotheses p, and number of overidentifying restrictions q

ρ	χ^2	NCF	F_∞	F	χ^2	NCF	F_∞	F	χ^2	NCF	F_∞	F
		$p=1,q=0$				$p=2,q=0$				$p=3,q=0$		
-0.95	0.184	0.135	0.137	0.135	0.330	0.194	0.191	0.194	0.499	0.261	0.265	0.261
-0.85	0.131	0.083	0.085	0.083	0.223	0.098	0.095	0.098	0.355	0.122	0.125	0.122
-0.5	0.088	0.064	0.064	0.064	0.122	0.066	0.066	0.066	0.169	0.075	0.076	0.075
0.5	0.086	0.061	0.061	0.061	0.137	0.067	0.068	0.067	0.219	0.080	0.081	0.080
0.85	0.145	0.091	0.093	0.091	0.243	0.113	0.110	0.113	0.378	0.140	0.144	0.140
0.9	0.166	0.114	0.117	0.114	0.284	0.143	0.139	0.143	0.429	0.179	0.183	0.179
0.95	0.195	0.140	0.143	0.140	0.349	0.203	0.199	0.203	0.508	0.257	0.260	0.257
		$p=1,q=1$				$p=2,q=1$				$p=3,q=1$		
-0.95	0.315	0.183	0.179	0.173	0.523	0.246	0.243	0.232	0.701	0.301	0.312	0.282
-0.85	0.194	0.089	0.086	0.088	0.346	0.109	0.107	0.107	0.497	0.122	0.128	0.119
-0.5	0.112	0.061	0.061	0.061	0.173	0.069	0.068	0.068	0.251	0.076	0.076	0.076
0.5	0.128	0.061	0.061	0.065	0.209	0.070	0.070	0.071	0.304	0.079	0.080	0.077
0.85	0.224	0.104	0.100	0.101	0.368	0.117	0.115	0.114	0.546	0.137	0.144	0.133
0.9	0.256	0.127	0.123	0.123	0.430	0.159	0.157	0.153	0.591	0.181	0.190	0.168
0.95	0.315	0.181	0.175	0.166	0.519	0.237	0.235	0.222	0.693	0.290	0.301	0.267
		$p=1,q=2$				$p=2,q=2$				$p=3,q=2$		
-0.95	0.398	0.182	0.179	0.160	0.650	0.239	0.224	0.202	0.815	0.274	0.270	0.231
-0.85	0.286	0.096	0.093	0.088	0.462	0.105	0.097	0.095	0.634	0.111	0.108	0.099
-0.5	0.148	0.059	0.060	0.058	0.241	0.066	0.063	0.065	0.355	0.079	0.078	0.075
0.5	0.169	0.055	0.057	0.055	0.287	0.066	0.063	0.067	0.411	0.074	0.072	0.073
0.85	0.289	0.097	0.095	0.086	0.484	0.113	0.104	0.100	0.672	0.119	0.116	0.102
0.9	0.333	0.126	0.123	0.107	0.544	0.150	0.140	0.131	0.718	0.165	0.161	0.136
0.95	0.390	0.178	0.174	0.148	0.631	0.236	0.222	0.199	0.799	0.259	0.255	0.205

The first three tests χ^2, NCF and F_∞ are based on the same unmodified Wald statistic but use different critical values. The χ^2 test uses the chi-squared critical value, the NCF test uses the noncentral F critical value, and the F_∞ test uses simulated nonstandard critical value. The standard F test is based on the modified Wald statistic and uses the standard F critical value.

Table 5: Empirical size of the nominal 5% χ^2 test, noncentral F test, nonstandard F_∞ test and standard F test based on the series LRV estimator under the centered MA design with $T = 100$, number of joint hypotheses p, and number of overidentifying restrictions q

L	χ^2	NCF	F_∞	F	χ^2	NCF	F_∞	F	χ^2	NCF	F_∞	F
	$p=1,q=0$				$p=2,q=0$				$p=3,q=0$			
3	0.017	0.007	0.007	0.007	0.089	0.030	0.029	0.030	0.201	0.047	0.048	0.047
6	0.030	0.017	0.017	0.017	0.068	0.023	0.022	0.023	0.134	0.028	0.029	0.028
9	0.048	0.029	0.030	0.029	0.079	0.027	0.026	0.027	0.142	0.032	0.033	0.032
	$p=1,q=1$				$p=2,q=1$				$p=3,q=1$			
3	0.102	0.033	0.031	0.036	0.229	0.057	0.056	0.059	0.299	0.047	0.050	0.049
6	0.106	0.039	0.037	0.046	0.169	0.031	0.031	0.036	0.275	0.034	0.037	0.039
9	0.108	0.035	0.034	0.039	0.159	0.032	0.031	0.034	0.259	0.033	0.036	0.036
	$p=1,q=2$				$p=2,q=2$				$p=3,q=2$			
3	0.180	0.046	0.046	0.042	0.286	0.050	0.046	0.042	0.387	0.043	0.043	0.035
6	0.164	0.039	0.037	0.045	0.265	0.036	0.032	0.040	0.425	0.036	0.036	0.039
9	0.165	0.040	0.039	0.043	0.265	0.032	0.029	0.034	0.402	0.032	0.031	0.034

See footnotes to Table 4

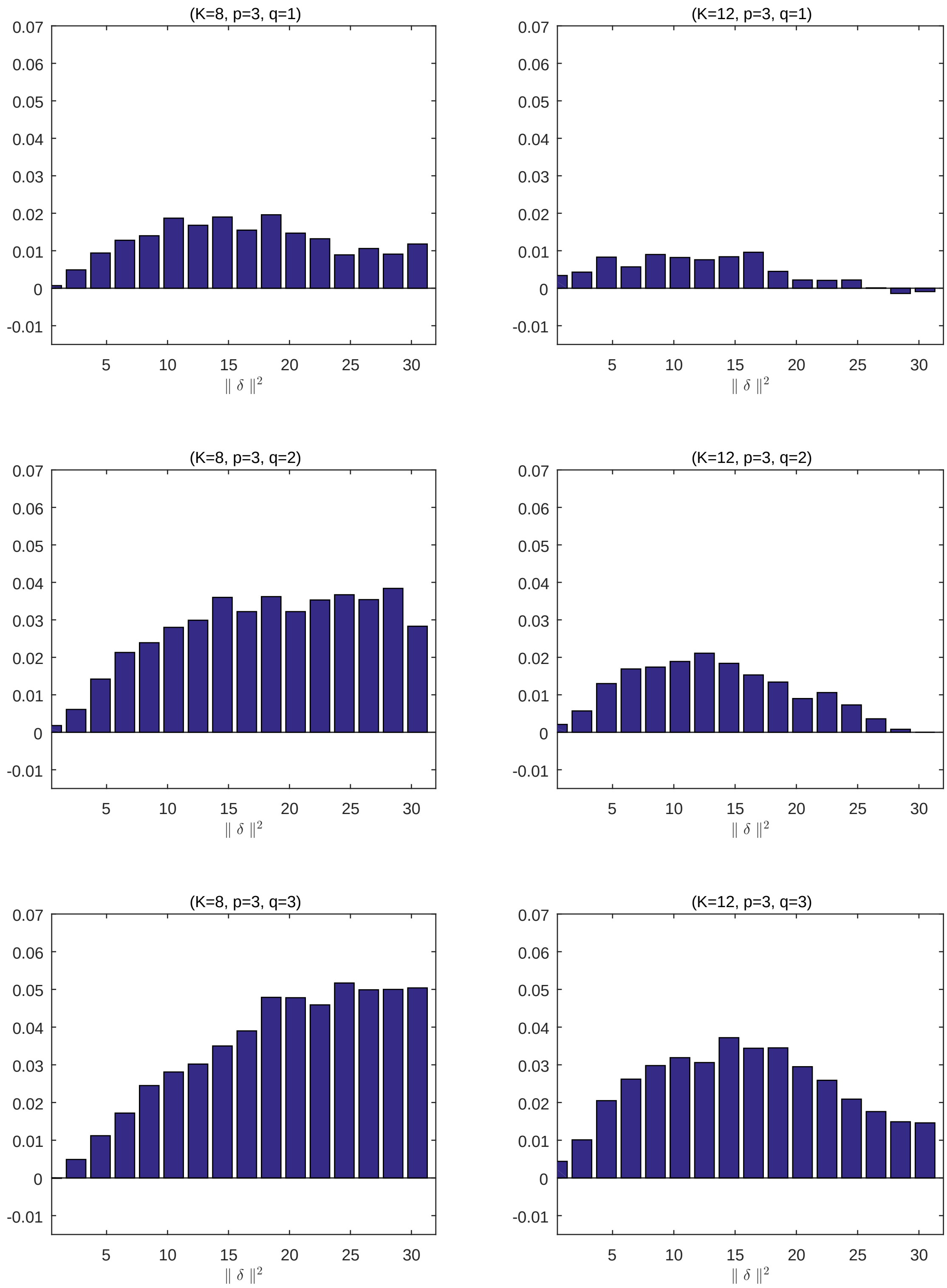

Figure 1: Difference of the asymptotic local power between the nonstandard F_∞ test and the standard F test $\pi^c(\cdot) - \pi(\cdot)$ for $K = 8, 12$, $q = 1, 2, 3$, and $p = 3$.

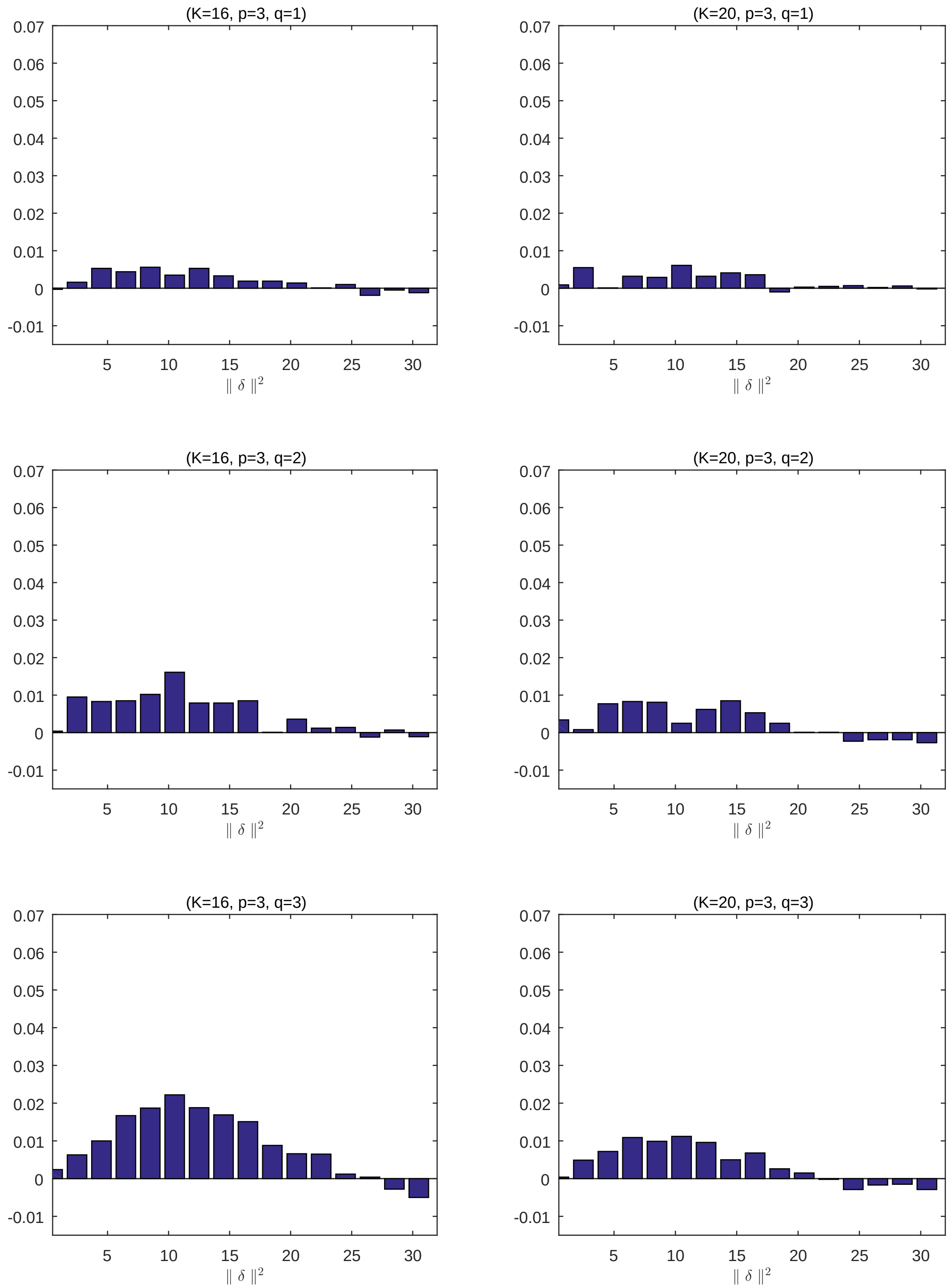

Figure 2: Difference of the asymptotic local power between the nonstandard F_∞ test and the standard F test $\pi^c(\cdot) - \pi(\cdot)$ for $K = 16, 20$, $q = 1, 2, 3$, and $p = 3$.

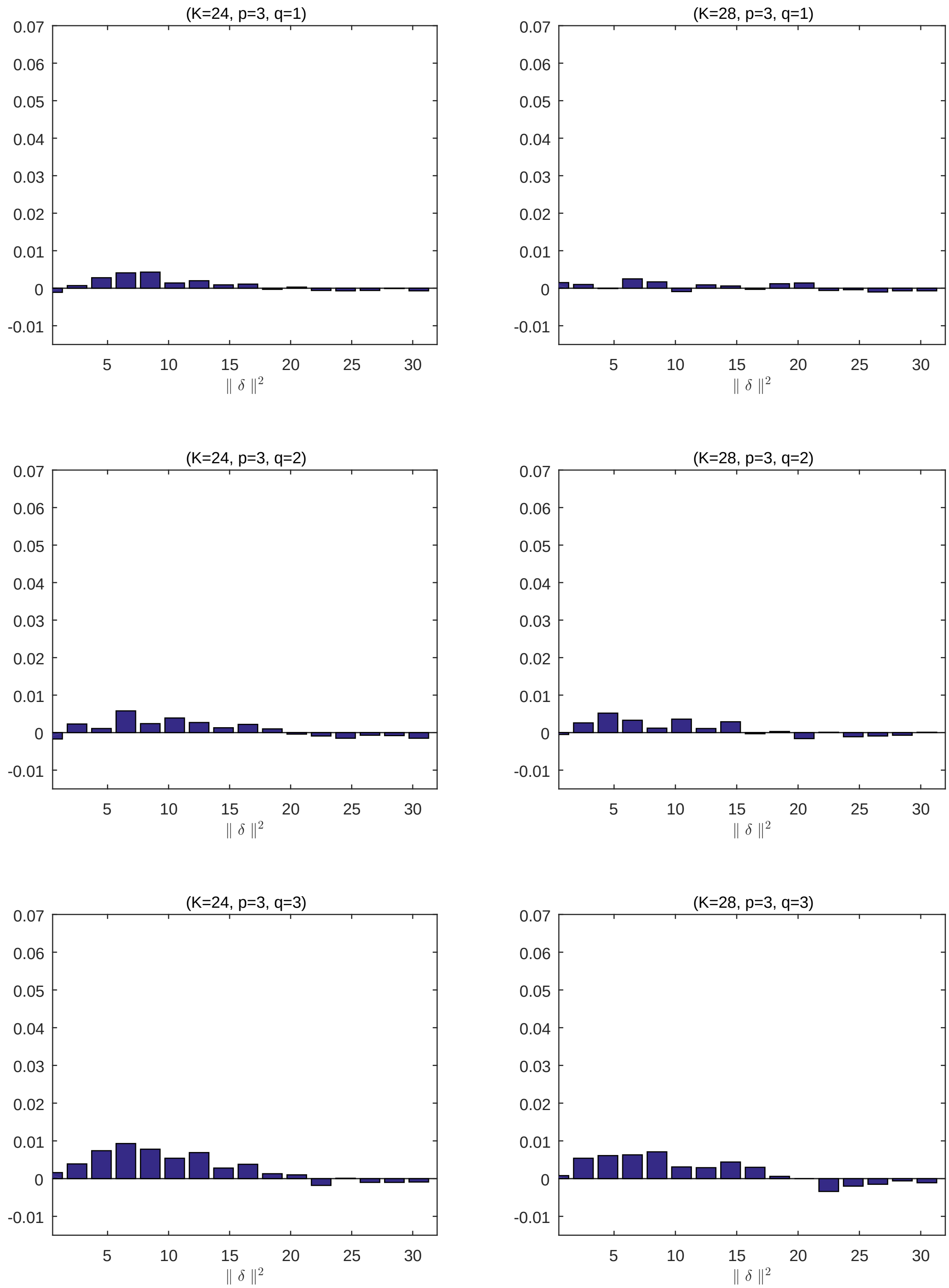

Figure 3: Difference of the asymptotic local power between the nonstandard F_∞ test and the standard F test $\pi^c(\cdot) - \pi(\cdot)$ for $K = 24, 28$, $q = 1, 2, 3$, and $p = 3$.

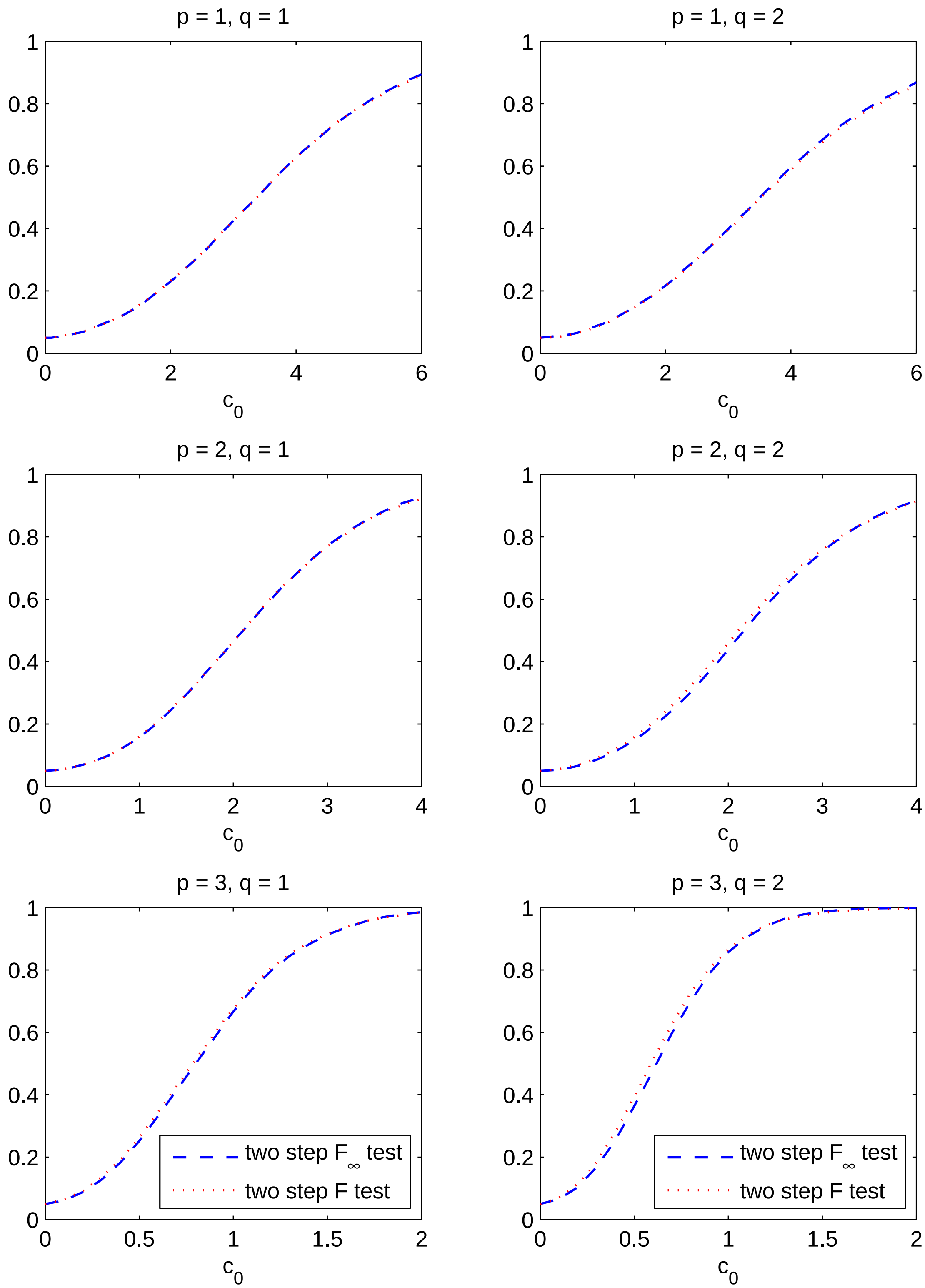

Figure 4: Size-adjusted power of two-step 5% F_∞ and F tests based on the series LRV estimator under the AR design with $\rho = 0.5$ and $T = 100$

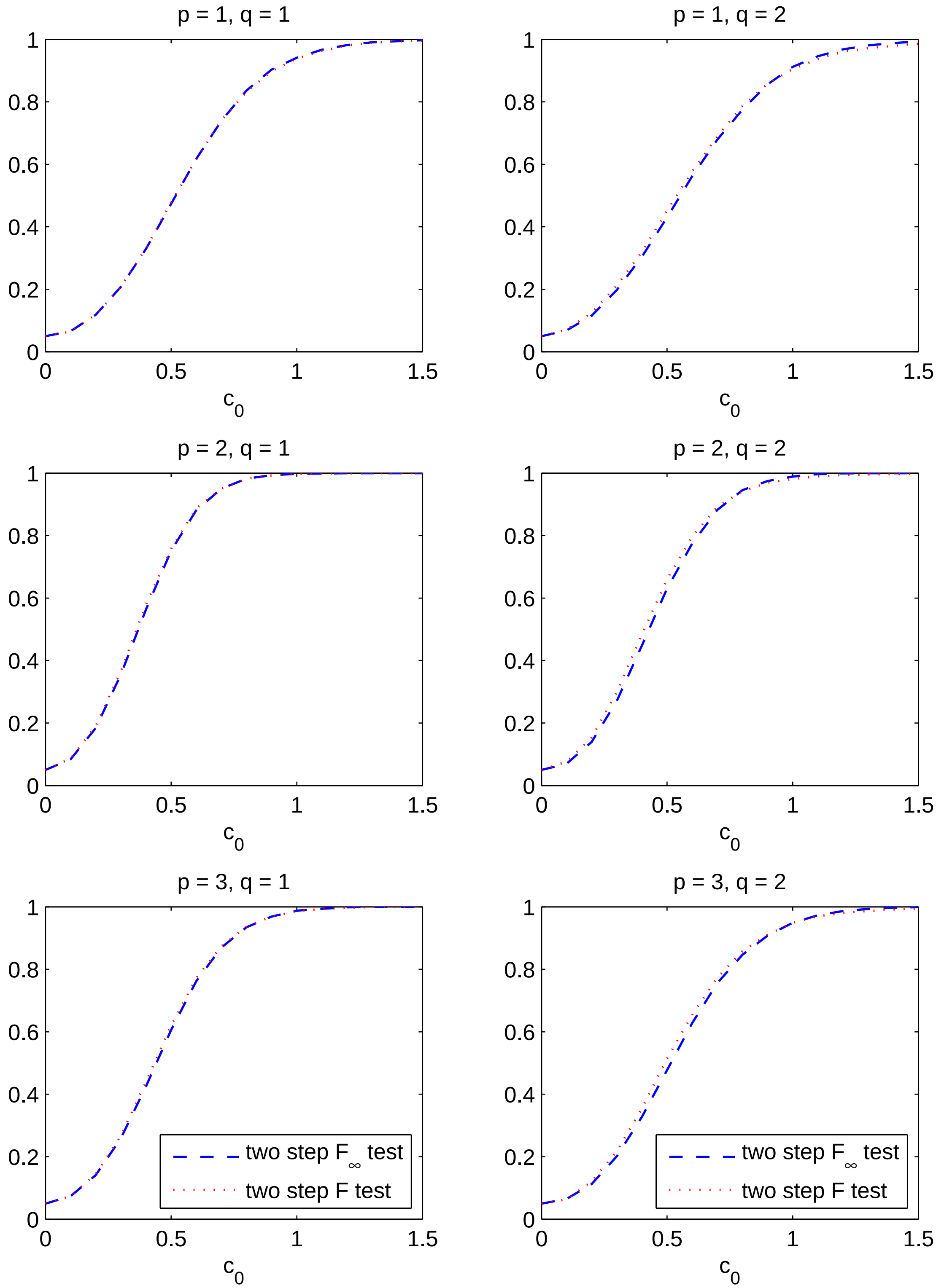

Figure 5: Size-adjusted power of two-step 5% F_∞ and F tests based on the series LRV estimator under the centered MA design with $L = 9$ and $T = 100$

7 Appendix of Proofs

Proof of Theorem 1. The marginal weak convergence results in (a) and (b) have been proved in Sun (2014b, Theorem 1), and the result in (c) has been proved in Sun and Kim (2012, Theorem 1 and equation (7)). It remains to show that the convergence results hold jointly. As a representative example, we prove that (a) and (c) hold jointly. We focus on $\mathbb{W}_T$, as Sun (2014b) has shown that $\mathbb{D}_T = \mathbb{W}_T + o_p(1)$ and $\mathbb{S}_T = \mathbb{W}_T + o_p(1)$ under the fixed-smoothing asymptotics.

Let

$$\tilde{W}_\infty = \int_0^1 \int_0^1 Q_K(r,s)\, dB_m(r)\, dB_m(s)'$$

and $G_\Lambda = \Lambda^{-1} G$, which is an $m \times d$ matrix, then it follows from Sun (2014b) and Sun and Kim (2012) that

$$W_T(\tilde{\theta}_T) \xrightarrow{d} \Lambda \tilde{W}_\infty \Lambda' \tag{18}$$

$$\begin{aligned} \mathbb{W}_T(\hat{\theta}_T) \xrightarrow{d} & \left\{ R[G_\Lambda' \tilde{W}_\infty^{-1} G_\Lambda]^{-1} G_\Lambda' \tilde{W}_\infty^{-1} B_m(1) \right\}' \left\{ R[G_\Lambda' \tilde{W}_\infty^{-1} G_\Lambda]^{-1} R' \right\}^{-1} \\ & \times \left\{ R[G_\Lambda' \tilde{W}_\infty^{-1} G_\Lambda]^{-1} G_\Lambda' \tilde{W}_\infty^{-1} B_m(1) \right\} /p := \mathcal{F}_\infty, \end{aligned} \tag{19}$$

$$\begin{aligned} \mathbb{J}_T(\hat{\theta}_T) \xrightarrow{d} & \left\{ B_m(1) - G_\Lambda [G_\Lambda' \tilde{W}_\infty^{-1} G_\Lambda]^{-1} G_\Lambda' \tilde{W}_\infty^{-1} B_m(1) \right\}' \tilde{W}_\infty^{-1} \\ \times & \left\{ B_m(1) - G_\Lambda [G_\Lambda' \tilde{W}_\infty^{-1} G_\Lambda]^{-1} G_\Lambda' \tilde{W}_\infty^{-1} B_m(1) \right\} := \mathcal{J}_\infty. \end{aligned} \tag{20}$$

In addition, a careful inspection shows that the above convergence results hold jointly. It remains to show that $(\mathcal{F}_\infty, \mathcal{J}_\infty)$ is equivalent in distribution to

$$\left(\left[B_p(1) - C_{pq} C_{qq}^{-1} B_q(1) \right]' D_{pp}^{-1} \left[B_p(1) - C_{pq} C_{qq}^{-1} B_q(1) \right] /p, \; B_q(1)' C_{qq}^{-1} B_q(1) \right).$$

Let $U_{m\times m} \Sigma_{m \times d} V_{d\times d}'$ be a singular value decomposition (SVD) of G_Λ. By definition, $U'U = UU' = I_m$, $VV' = V'V = I_d$ and

$$\Sigma = \begin{bmatrix} A_{d\times d} \\ O_{q \times d} \end{bmatrix},$$

where A is a diagonal matrix with singular values on the main diagonal and O is a matrix of zeros. Then we have:

$$\begin{aligned} & [G_\Lambda' \tilde{W}_\infty^{-1} G_\Lambda]^{-1} G_\Lambda' \tilde{W}_\infty^{-1} B_m(1) \\ & = [V\Sigma' U' \tilde{W}_\infty^{-1} U \Sigma V']^{-1} V \Sigma' U' \tilde{W}_\infty^{-1} B_m(1) \\ & = V[\Sigma' U' \tilde{W}_\infty^{-1} U \Sigma]^{-1} \Sigma' U' \tilde{W}_\infty^{-1} B_m(1) = V[\Sigma' U' \tilde{W}_\infty^{-1} U \Sigma]^{-1} \Sigma' [U' \tilde{W}_\infty^{-1} U] \left[U' B_m(1) \right], \end{aligned}$$

and

$$
\begin{aligned}
&B_m(1) - G_\Lambda[G'_\Lambda \tilde{W}_\infty^{-1} G_\Lambda]^{-1} G'_\Lambda \tilde{W}_\infty^{-1} B_m(1) \\
&= B_m(1) - U\Sigma V'[V\Sigma' U' \tilde{W}_\infty^{-1} U \Sigma V']^{-1} V \Sigma' U' \tilde{W}_\infty^{-1} B_m(1) \\
&= U \left\{ U' B_m(1) - \Sigma[\Sigma' U' \tilde{W}_\infty^{-1} U \Sigma]^{-1} \Sigma' [U' \tilde{W}_\infty^{-1} U] U' B_m(1) \right\}.
\end{aligned}
$$

Since $[U'\tilde{W}_\infty^{-1}U,\, U'B_m(1)]$ has the same joint distribution as $[\tilde{W}_\infty^{-1}, B_m(1)]$, we can write

$$
\begin{pmatrix} \mathcal{F}_\infty \\ \mathcal{J}_\infty \end{pmatrix} \overset{d}{=} \begin{pmatrix} \tilde{\mathcal{F}}_\infty \\ \tilde{\mathcal{J}}_\infty \end{pmatrix}
$$

where

$$
\begin{aligned}
\tilde{\mathcal{F}}_\infty = &\left\{ RV(\Sigma' \tilde{W}_\infty^{-1} \Sigma)^{-1} \Sigma' \tilde{W}_\infty^{-1} B_m(1) \right\}' \left\{ RV(\Sigma' \tilde{W}_\infty^{-1} \Sigma)^{-1} V' R' \right\}^{-1} \\
&\times RV(\Sigma' \tilde{W}_\infty^{-1} \Sigma)^{-1} \Sigma' \tilde{W}_\infty^{-1} B_m(1)/p,
\end{aligned}
$$

and

$$
\begin{aligned}
\tilde{\mathcal{J}}_\infty = &\left\{ \left[I_m - \Sigma(\Sigma' \tilde{W}_\infty^{-1} \Sigma)^{-1} \Sigma' \tilde{W}_\infty^{-1} \right]' B_m(1) \right\}' \tilde{W}_\infty^{-1} \\
&\times \left[I_m - \Sigma(\Sigma' \tilde{W}_\infty^{-1} \Sigma)^{-1} \Sigma' \tilde{W}_\infty^{-1} \right]' B_m(1).
\end{aligned}
$$

We proceed to simplify $\tilde{\mathcal{F}}_\infty$ and $\tilde{\mathcal{J}}_\infty$ starting with $\tilde{\mathcal{F}}_\infty$. We write

$$
\tilde{W}_\infty = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \text{ and } \tilde{W}_\infty^{-1} = \begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix}
$$

where C_{11} and C^{11} are $d \times d$ matrices, C_{22} and C^{22} are $q \times q$ matrices, and $C_{12} = C'_{21}$, $C^{12} = (C^{21})'$. By definition,

$$
C_{11} = \int_0^1 \int_0^1 Q_K(r,s) dB_d(r) dB_d(s)' = \begin{pmatrix} C_{pp} & C_{p,d-p} \\ C'_{p,d-p} & C_{d-p,d-p} \end{pmatrix} \tag{21}
$$

$$
C_{12} = \int_0^1 \int_0^1 Q_K(r,s) dB_d(r) dB_q(s)' = \begin{pmatrix} C_{pq} \\ C_{d-p,q} \end{pmatrix} \tag{22}
$$

$$
C_{22} = \int_0^1 \int_0^1 Q_K(r,s) dB_q(r) dB_q(s)' = C_{qq} \tag{23}
$$

where C_{pp}, C_{pq}, and C_{qq} are defined in (6) and (7), and $C_{d-p,d-p}$, $C_{p,d-p}$ and $C_{d-p,q}$ are similarly defined. It follows from the partitioned inverse formula that

$$
C^{11} = \left[C_{11} - C_{12} C_{22}^{-1} C_{21} \right]^{-1}, \; C^{12} = -C^{11} C_{12} C_{22}^{-1}.
$$

With the above partition of $\tilde{W}_{\infty}^{-1}$, we have

$$
\begin{aligned}
(\Sigma'\tilde{W}_{\infty}^{-1}\Sigma)^{-1} &= \left[\begin{pmatrix} A' & O' \end{pmatrix}\begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix}\begin{pmatrix} A \\ O \end{pmatrix}\right]^{-1} \\
&= (A'C^{11}A)^{-1} = A^{-1}\left(C^{11}\right)^{-1}\left(A'\right)^{-1},
\end{aligned}
$$

and so

$$
\begin{aligned}
& RV(\Sigma'\tilde{W}_{\infty}^{-1}\Sigma)^{-1}\Sigma'\tilde{W}_{\infty}^{-1} \\
&= RVA^{-1}\left(C^{11}\right)^{-1}\left(A'\right)^{-1}\begin{pmatrix} A' & O' \end{pmatrix}\begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix} \\
&= RVA^{-1}\left(C^{11}\right)^{-1}\left(A'\right)^{-1}A'\begin{pmatrix} C^{11} & C^{12} \end{pmatrix} \\
&= RVA^{-1}\left[\begin{array}{cc} I_d, & \left(C^{11}\right)^{-1}C^{12} \end{array}\right] \\
&= RVA^{-1}\left[\begin{array}{cc} I_d, & -C_{12}C_{22}^{-1} \end{array}\right] \qquad (24)
\end{aligned}
$$

and

$$
RV(\Sigma'\tilde{W}_{\infty}^{-1}\Sigma)^{-1}V'R' = RVA^{-1}\left(C^{11}\right)^{-1}\left(A'\right)^{-1}V'R'.
$$

As a result,

$$
\begin{aligned}
\tilde{\mathcal{F}}_{\infty} = {} & B_m(1)'\left[RVA^{-1}\begin{pmatrix} I_d, & -C_{12}C_{22}^{-1} \end{pmatrix}\right]'\left[RVA^{-1}\left(C^{11}\right)^{-1}\left(A'\right)^{-1}V'R'\right]^{-1} \\
& \times\left[RVA^{-1}\begin{pmatrix} I_d, & -C_{12}C_{22}^{-1} \end{pmatrix}\right]B_m(1)/p.
\end{aligned}
$$

Let $B_m(1) = [B_d(1)', B_q(1)']'$ and $\tilde{U}_{p\times p}\tilde{\Sigma}_{p\times d}\tilde{V}'_{d\times d}$ be a SVD of RVA^{-1}, where

$$
\tilde{\Sigma}_{p\times d} = \begin{pmatrix} \tilde{A}_{p\times p}, & \tilde{O}_{p\times(d-p)} \end{pmatrix} \qquad (25)
$$

$\tilde{A}$ is a diagonal matrix, and $\tilde{O}$ is a matrix of zeros. Then

$$
\begin{aligned}
\tilde{\mathcal{F}}_{\infty} &= \left\{\tilde{U}\tilde{\Sigma}\tilde{V}'\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]\right\}'\left[\tilde{U}\tilde{\Sigma}\tilde{V}'\left(C^{11}\right)^{-1}\tilde{V}\tilde{\Sigma}'\tilde{U}'\right]^{-1} \\
&\quad\times\tilde{U}\tilde{\Sigma}\tilde{V}'\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]/p \\
&= \left\{\tilde{\Sigma}\tilde{V}'\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]\right\}'\left[\tilde{\Sigma}\tilde{V}'\left(C^{11}\right)^{-1}\tilde{V}\tilde{\Sigma}'\right]^{-1} \\
&\quad\times\tilde{\Sigma}\tilde{V}'\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]/p \\
&= \left\{\tilde{\Sigma}\left[\tilde{V}'B_d(1) - \tilde{V}'C_{12}C_{22}^{-1}B_q(1)\right]\right\}'\left[\tilde{\Sigma}\tilde{V}'\left(C^{11}\right)^{-1}\tilde{V}\tilde{\Sigma}'\right]^{-1} \\
&\quad\times\tilde{\Sigma}\left[\tilde{V}'B_d(1) - \tilde{V}'C_{12}C_{22}^{-1}B_q(1)\right]/p.
\end{aligned}
$$

Using the same steps, we have

$$
\begin{aligned}
& I_m - \Sigma\left[\Sigma'\tilde{W}_\infty^{-1}\Sigma\right]^{-1}\Sigma'\tilde{W}_\infty^{-1} \\
&= I_m - \begin{pmatrix} A \\ O \end{pmatrix}\left[A'C^{11}A\right]^{-1}\begin{pmatrix} A' & O' \end{pmatrix}\begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix} \\
&= I_m - \begin{pmatrix} \left(C^{11}\right)^{-1} & O_{12} \\ O_{21} & O_{22} \end{pmatrix}\begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix} \\
&= I_m - \begin{pmatrix} I_d & \left(C^{11}\right)^{-1}C^{12} \\ O_{21} & O_{22} \end{pmatrix} = \begin{pmatrix} O_{11} & -\left(C^{11}\right)^{-1}C^{12} \\ O_{21} & I_q \end{pmatrix}
\end{aligned}
$$

where O_{ij} are matrices of zeros with the dimensions as C_{ij}. So

$$
\begin{aligned}
\tilde{J}_\infty &= \left[\begin{pmatrix} O_{11} & -\left(C^{11}\right)^{-1}C^{12} \\ O_{21} & I_q \end{pmatrix} B_m(1)\right]'\tilde{W}_\infty^{-1} \\
&\times \left[\begin{pmatrix} O_{11} & -\left(C^{11}\right)^{-1}C^{12} \\ O_{21} & I_q \end{pmatrix} B_m(1)\right] \\
&= \begin{pmatrix} -\left(C^{11}\right)^{-1}C^{12}B_q(1) \\ B_q(1) \end{pmatrix}'\begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix}\begin{pmatrix} -\left(C^{11}\right)^{-1}C^{12}B_q(1) \\ B_q(1) \end{pmatrix} \\
&= \begin{pmatrix} -\left(C^{11}\right)^{-1}C^{12}B_q(1) \\ B_q(1) \end{pmatrix}'\begin{pmatrix} O \\ \left[C^{22} - C^{21}\left(C^{11}\right)^{-1}C^{21}\right]B_q(1) \end{pmatrix} \\
&= B_q(1)'\left[C^{22} - C^{21}\left(C^{11}\right)^{-1}C^{21}\right]B_q(1) \\
&= B_q(1)'C_{qq}^{-1}B_q(1).
\end{aligned}
$$

In the last equality, we have used $[C^{22} - C^{21}\left(C^{11}\right)^{-1}C^{21}]^{-1} = C_{22} = C_{qq}$, which follows from the partitioned inverse formula.

Noting that the joint distribution of $\left[\tilde{V}'B_d(1), \tilde{V}'C_{12}, C_{22}, \ \tilde{V}'\left(C^{11}\right)^{-1}\tilde{V}\right]$ is invariant to the orthonormal matrix $\tilde{V}$, we have

$$
\begin{pmatrix} \tilde{\mathcal{F}}_\infty \\ \tilde{\mathcal{J}}_\infty \end{pmatrix} \overset{d}{=} \begin{pmatrix} \tilde{\mathcal{F}}_\infty^* \\ \tilde{\mathcal{J}}_\infty^* \end{pmatrix}
$$

where

$$\begin{aligned}\tilde{\mathcal{F}}_{\infty}^{*} &= \left\{\tilde{\Sigma}\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]\right\}'\left[\tilde{\Sigma}\left(C^{11}\right)^{-1}\tilde{\Sigma}'\right]^{-1} \\ &\times \tilde{\Sigma}\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]/p \\ &= \left\{\begin{pmatrix} \tilde{A}, & \tilde{O} \end{pmatrix}\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]\right\}' \\ &\times \left[\begin{pmatrix} \tilde{A}, & \tilde{O} \end{pmatrix}\left(C^{11}\right)^{-1}\begin{pmatrix} \tilde{A}, & \tilde{O} \end{pmatrix}'\right]^{-1} \\ &\times \left\{\begin{pmatrix} \tilde{A}, & \tilde{O} \end{pmatrix}\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]\right\}/p,\end{aligned}$$

and

$$\tilde{\mathcal{J}}_{\infty}^{*} = B_q(1)'C_{qq}^{-1}B_q(1).$$

Write

$$\left(C^{11}\right)^{-1} = C_{11} - C_{12}C_{22}^{-1}C_{21} = \begin{pmatrix} D_{pp} & D^{12} \\ D^{21} & D^{22} \end{pmatrix}$$

where $D_{pp} = C_{pp} - C_{pq}C_{qq}^{-1}C_{pq}'$ and D^{22} is a $(d-p)\times(d-p)$ matrix. Using the above partition of $\left(C^{11}\right)^{-1}$ and equations (21)–(23), we have

$$\tilde{\mathcal{F}}_{\infty}^{*} = \left[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right]' D_{pp}^{-1}\left[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right]/p.$$

So

$$\begin{pmatrix} \mathbb{W}_T(\hat{\theta}_T) \\ \mathbb{J}_T(\hat{\theta}_T) \end{pmatrix} \stackrel{d}{\rightarrow} \begin{pmatrix} \tilde{\mathcal{F}}_{\infty}^{*} \\ \tilde{\mathcal{J}}_{\infty}^{*} \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \tilde{\mathcal{F}}_{\infty} \\ \tilde{\mathcal{J}}_{\infty} \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \mathcal{F}_{\infty} \\ \mathcal{J}_{\infty} \end{pmatrix}.$$

This completes the proof of the theorem. ■

Proof of Theorem 2. We prove the case with $q > 0$ only. When $q = 0$, the results are the same as those established in Sun (2013) for the first-step GMM tests.

Part (a). We first prove (11). Let $B_p^{(i)}(r)$ be the i-th element of $B_p(r)$. Define

$$\begin{aligned} C_{p_{(i)},q} &= \int Q_K(r,s)\, dB_p^{(i)}(r)\, dB_q(r)' \in \mathbb{R}^{1\times q}, \\ C_{q,p_{(j)}} &= \int Q_K(r,s)\, dB_q(r)\, dB_p^{(j)}(r) \in \mathbb{R}^{q\times 1}, \end{aligned}$$

which are the i-th row of C_{pq} and the j-th column of C_{qp}, respectively. Then the (i,j)-th element of the left hand side of (11) can be written as

$$
\begin{aligned}
& E\left\{ C_{p(i),q}C_{qq}^{-1}B_q(1)B_q(1)'C_{qq}^{-1}C_{q,p(j)} \right| \left. B_q(\cdot)\right\} \\
&= E\left\{ \frac{1}{K}\sum_{\ell_1=1}^{K}\left(\int_0^1 \Phi_{\ell_1}(r)\,dB_p^{(i)}(r)\right)\right. \\
&\times \left(\int_0^1 \Phi_{\ell_1}(s)\,dB_q(s)'\right) C_{qq}^{-1}B_q(1)B_q(1)'C_{qq}^{-1}\frac{1}{K}\sum_{\ell_2=1}^{K}\left(\int_0^1 \Phi_{\ell_2}(\tilde r)\,dB_q(\tilde r)\right) \\
&\left.\left(\int_0^1 \Phi_{\ell_2}(\tilde s)\,dB_p^{(j)}(\tilde s)\right)\right| \left. B_q(\cdot)\right\} \\
&= E\left\{ \frac{1}{K^2}\sum_{\ell_1,\ell_2}\left(\int_0^1 \Phi_{\ell_1}(r)\,dB_p^{(i)}(r)\right)\right. \\
&\times \left(\int_0^1 \Phi_{\ell_1}(s)\,dB_q(s)'\right) C_{qq}^{-1}B_q(1)B_q(1)'C_{qq}^{-1}\left(\int_0^1 \Phi_{\ell_2}(\tilde r)\,dB_q(\tilde r)\right) \\
&\times \left.\left(\int_0^1 \Phi_{\ell_2}(\tilde s)\,dB_p^{(j)}(\tilde s)\right)\right| \left. B_q(\cdot)\right\}.
\end{aligned}
$$

Note that the second term in the above equation is a scalar that depends only on $B_q(\cdot)$. Taking this term out of the conditional expectation, we have

$$
\begin{aligned}
& E\left\{ C_{p(i),q}C_{qq}^{-1}B_q(1)B_q(1)'C_{qq}^{-1}C_{q,p(j)} \right| \left. B_q(\cdot)\right\} \\
&= \delta_{ij}\frac{1}{K^2}\sum_{\ell_1,\ell_2}\left(\int_0^1 \Phi_{\ell_1}(r)\Phi_{\ell_2}(r)\,dr\right) \\
&\times \left(\int_0^1 \Phi_{\ell_1}(s)\,dB_q(s)'\right) C_{qq}^{-1}B_q(1)B_q(1)'C_{qq}^{-1}\left(\int_0^1 \Phi_{\ell_2}(\tilde r)\,dB_q(\tilde r)\right) \\
&= \delta_{ij}\frac{1}{K^2}\sum_{\ell=1}^{K}\left(\int_0^1 \Phi_{\ell}(s)\,dB_q(s)'\right) C_{qq}^{-1}B_q(1)B_q(1)'C_{qq}^{-1}\left(\int_0^1 \Phi_{\ell}(\tilde r)\,dB_q(\tilde r)\right) \\
&= \frac{\delta_{ij}}{K}B_q(1)'C_{qq}^{-1}\frac{1}{K}\sum_{\ell=1}^{K}\left(\int_0^1 \Phi_{\ell}(s)\,dB_q(s)\right)\left(\int_0^1 \Phi_{\ell}(\tilde r)\,dB_q(\tilde r)'\right)C_{qq}^{-1}B_q(1) \\
&= \frac{\delta_{ij}}{K}B_q(1)'C_{qq}^{-1}B_q(1),
\end{aligned}
$$

where $\delta_{ij} = 1\{i = j\}$. Here we have used the independence of $B_p^{(i)}(\cdot)$ from $B_p^{(j)}(\cdot)$ when $i \neq j$ and $\int_0^1 \Phi_{\ell_1}(r)\Phi_{\ell_2}(r)\,dr = 0$ when $\ell_1 \neq \ell_2$. So, conditional on $B_q(\cdot)$,

$$
B_p(1) - C_{pq}C_{qq}^{-1}B_q(1) \sim N\left[0, I_p\left(1 + \frac{1}{K}B_q(1)'C_{qq}^{-1}B_q(1)\right)\right].
$$

That is, conditional on $B_q(\cdot)$,

$$
\frac{B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)}{\sqrt{1 + B_q(1)'C_{qq}^{-1}B_q(1)/K}} \sim N(0, I_p).
$$

But $N(0, I_p)$ does not depend on $B_q(\cdot)$, so

$$\xi_p := \frac{B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)}{\sqrt{1 + B_q(1)' C_{qq}^{-1} B_q(1)/K}} \sim N(0, I_p)$$

unconditionally. In addition, ξ_p is independent of D_{pp}. Using these results, we have

$$\begin{aligned}
&\frac{\mathcal{F}_\infty}{1 + B_q(1)' C_{qq}^{-1} B_q(1)/K} \overset{d}{=} \frac{\xi_p' D_{pp}^{-1} \xi_p}{p} \overset{d}{=} \frac{\chi_p^2/p}{\chi_{K-p-q+1}^2/K} \\
&\overset{d}{=} \frac{K}{(K-p-q+1)} \frac{\chi_p^2/p}{\chi_{K-p-q+1}^2/(K-p-q+1)} \\
&\overset{d}{=} \frac{K}{(K-p-q+1)} \mathcal{F}_{p,K-p-q+1}.
\end{aligned}$$

In view of Theorem 1, we have

$$\frac{K-p-q+1}{K} \frac{\mathbb{W}_T(\hat{\theta}_T)}{1 + \mathbb{J}_T(\hat{\theta}_T)/K} \overset{d}{\to} \mathcal{F}_{p,K-p-q+1},$$

completing the proof of Part (a).

Using the same argument, we can prove Part (d). Parts (b) and (c) hold because the asymptotic equivalence of $\mathbb{W}_T$, $\mathbb{D}_T$ and $\mathbb{S}_T$ still holds under the fixed-smoothing asymptotics. For more details, see Sun (2014b). ■

Proof of Theorem 3. It suffices to prove parts (a) and (c), as the proofs for parts (b) and (d) are similar. It is easy to show that

$$\begin{aligned}
\mathbb{W}_T, \mathbb{D}_T, \mathbb{S}_T \overset{d}{\to} & \left\{ R \left[G_\Lambda' \tilde{W}_\infty^{-1} G_\Lambda \right]^{-1} G_\Lambda' \tilde{W}_\infty^{-1} B_m(1) + \delta_0 \right\}' \left\{ R \left[G_\Lambda' \tilde{W}_\infty^{-1} G_\Lambda \right]^{-1} R' \right\}^{-1} \\
& \times \left\{ R \left[G_\Lambda' \tilde{W}_\infty^{-1} G_\Lambda \right]^{-1} G_\Lambda' \tilde{W}_\infty^{-1} B_m(1) + \delta_0 \right\} /p := \mathcal{F}_{\infty,\delta_0}.
\end{aligned}$$

Repeating the same arguments in the proof of Theorem 1 while keeping tracking of the term containing δ_0, we have

$$\mathcal{F}_{\infty,\delta_0} \overset{d}{=} \left[B_p(1) + \delta^* - C_{pq}C_{qq}^{-1} B_q(1) \right]' D_{pp}^{-1} \left[B_p(1) + \delta^* - C_{pq}C_{qq}^{-1} B_q(1) \right] /p \tag{26}$$

where

$$\delta^* := \tilde{A}^{-1} \tilde{U}' \delta_0.$$

and $\tilde{A}$ and $\tilde{U}$ are defined in (25).

By the rotation invariance of the multivariate standard normal distribution, we have, for any orthonormal matrices $H_{p\times p}$:

$$
\begin{aligned}
\mathcal{F}_{\infty,\delta_0} &\overset{d}{=} \left[HB_p(1) + H\delta^* - HC_{pq}C_{qq}^{-1}B_q(1)\right]' HD_{pp}^{-1}H' \\
&\times \left[HB_p(1) + H\delta^* - HC_{pq}C_{qq}^{-1}B_q(1)\right]/p \\
&\overset{d}{=} \left[B_p(1) + H\delta^* - C_{pq}C_{qq}^{-1}B_q(1)\right]' D_{pp}^{-1}\left[B_p(1) + H\delta^* - C_{pq}C_{qq}^{-1}B_q(1)\right]/p.
\end{aligned}
$$

In particular, upon choosing $H = (\delta^*/\left\|\delta^*\right\|, \tilde{H})'$ for some $\tilde{H}$ such that $\tilde{H}'\delta^* = 0_{(p-1)\times 1}$, we obtain

$$
\mathcal{F}_{\infty,\delta_0} \overset{d}{=} \left[B_p(1) + \left\|\delta^*\right\| e_p - C_{pq}C_{qq}^{-1}B_q(1)\right]' D_{pp}^{-1}\left[B_p(1) + \left\|\delta^*\right\| e_p - C_{pq}C_{qq}^{-1}B_q(1)\right]/p,
$$

where $e_p := (1, 0, \ldots, 0)' \in \mathbb{R}^p$. That is, the distribution of $\mathcal{F}_{\infty,\delta_0}$ depends on δ^* only through $\left\|\delta^*\right\|$.

Noting $\tilde{U}\tilde{A}\tilde{A}'\tilde{U}' = \tilde{U}\tilde{\Sigma}\tilde{V}'(\tilde{U}\tilde{\Sigma}\tilde{V}')' = RVA^{-1}(RVA^{-1})'$, we have

$$
\begin{aligned}
\left\|\delta^*\right\|^2 &= \delta_0'\tilde{U}(\tilde{A}^{-1})'\tilde{A}^{-1}\tilde{U}'\delta_0 = \delta_0'(\tilde{U}\tilde{A}\tilde{A}'\tilde{U}')^{-1}\delta_0 \\
&= \delta_0'[\tilde{U}\tilde{\Sigma}\tilde{V}'\tilde{V}\tilde{\Sigma}'\tilde{U}']^{-1}\delta_0 = \delta_0'\left[RVA^{-1}(A^{-1})'V'R'\right]^{-1}\delta_0 \\
&= \delta_0'\{R\left(V\Sigma U'U\Sigma'V'\right)^{-1}R'\}^{-1}\delta_0 = \delta_0'\{R\left(G'\Omega^{-1}G\right)^{-1}R'\}^{-1}\delta_0 \\
&= \delta_0'\left(\Lambda_R\Lambda_R'\right)^{-1}\delta_0 = \left\|\delta\right\|^2.
\end{aligned}
$$

Therefore,

$$
\mathcal{F}_{\infty,\delta_0} \overset{d}{=} \left[B_p(1) + \delta - C_{pq}C_{qq}^{-1}B_q(1)\right]' D_{pp}^{-1}\left[B_p(1) + \delta - C_{pq}C_{qq}^{-1}B_q(1)\right]/p. \tag{27}
$$

Following the proof for Theorem 2, we now have

$$
\begin{aligned}
\mathcal{F}_{\infty,\delta_0} &\overset{d}{=} \frac{1}{p}\left(1+\frac{\mathcal{J}_\infty}{K}\right)\left(\xi_p + \frac{\delta}{\sqrt{1+\mathcal{J}_\infty/K}}\right)' D_{pp}^{-1}\left(\xi_p + \frac{\delta}{\sqrt{1+\mathcal{J}_\infty/K}}\right) \\
&= \frac{K}{K-p-q+1}\left(1+\frac{\mathcal{J}_\infty}{K}\right)\mathcal{F}_{p,K-p-q+1}\left(\frac{\left\|\delta\right\|^2}{1+\mathcal{J}_\infty/K}\right) := \mathcal{F}_\infty(\left\|\delta_J\right\|^2),
\end{aligned}
$$

where $\mathcal{J}_\infty, \xi_p$ and D_{pp} are mutually independent from each other. This completes the proof of part (a).

Part (c) follows directly from the joint convergence of $(\mathbb{W}_T, \mathbb{J}_T)'$ to $(\mathcal{F}_\infty(\left\|\delta_J\right\|^2), \mathcal{J}_\infty)'$ and the continuous mapping theorem. ■

References

[1] Andrews, D. W. K., 1991, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. Econometrica 59, 817–858.

[2] Bester, C. A., T. G. Conley, C. B. Hansen, and T. J. Vogelsang, 2016, Fixed-b Asymptotics for Spatially Dependent Robust Nonparametric Covariance Matrix Estimators. Econometric Theory 32(1), 154–186.

[3] Gonçalves, S., 2011. The Moving Blocks Bootstrap for Panel Linear Regression Models with Individual Fixed Effects. Econometric Theory 27(5), 1048–1082.

[4] Gonçalves, S. and T. Vogelsang, 2011. Block Bootstrap HAC Robust Tests: The Sophistication of the Naive Bootstrap. Econometric Theory 27(4), 745–791.

[5] Hall, A. R., 2005, Generalized Method of Moments. Oxford University Press. New York.

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

[7] Hwang, J. and Y. Sun, 2015, Should We Go One Step Further? An Accurate Comparison of One-Step and Two-Step Procedures in a Generalized Method of Moments Framework. Working paper, Department of Economics, UC San Diego.

[8] Jansson, M. 2004, On the Error of Rejection Probability in Simple Autocorrelation Robust Tests. Econometrica 72, 937–946.

[9] Kiefer, N. M. and T. J. Vogelsang, 2002a, Heteroskedasticity-autocorrelation Robust Testing Using Bandwidth Equal to Sample Size. Econometric Theory 18, 1350–1366.

[10] Kiefer, N. M. and T. J. Vogelsang, 2002b, Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation. Econometrica 70, 2093–2095.

[11] Kiefer, N. M. and T. J. Vogelsang, 2005, A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests. Econometric Theory 21, 1130–1164.

[12] Müller, U. K., 2007, A Theory of Robust Long-Run Variance Estimation. Journal of Econometrics 141, 1331–1352.

[13] Müller, U. K., 2014, HAC Corrections for Strongly Autocorrelated Time Series. Journal of Business & Economic Statistics, 32(3), 311–322

[14] Newey, W. K. and K. D. West, 1987, A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55, 703–708.

[15] Phillips, P. C. B., 2005, HAC Estimation by Automated Regression. Econometric Theory 21, 116–142.

[16] Preinerstorfer, D., 2015, Finite Sample Properties of Tests based on Prewhitened Nonparametric Covariance Estimators. Working paper, Department of Statistics and OR, University of Vienna.

[17] Preinerstorfer, D. and B. M. Pötscher, 2016, On Size and Power of Heteroskedasticity and Autocorrelation Robust Tests. Econometric Theory, 32(2), 261–358.

[18] Sun, Y., 2011, Robust Trend Inference with Series Variance Estimator and Testing-optimal Smoothing Parameter. Journal of Econometrics 164(2), 345–366.

[19] Sun, Y., 2013, A Heteroskedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator. Econometrics Journal 16, 1–26.

[20] Sun, Y., 2014a, Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroscedasticity and Autocorrelation Robust Inference. Journal of Econometrics 178(3), 659–677.

[21] Sun, Y., 2014b, Fixed-smoothing Asymptotics in a Two-step GMM Framework. Econometrica 82(6), 2327–2370.

[22] Sun, Y., 2014c, Fixed-smoothing Asymptotics and Asymptotic F and t Tests in the Presence of Strong Autocorrelation. Advances in Econometrics 33, 23–63.

[23] Sun, Y., P. C. B. Phillips and S. Jin, 2008, Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing. Econometrica 76(1), 175–194.

[24] Sun, Y., P. C. B. Phillips and S. Jin, 2011, Power Maximization and Size Control in Heteroscedasticity and Autocorrelation Robust Tests with Exponentiated Kernels. Econometric Theory 27(6), 1320–1368.

[25] Sun, Y. and P. C. B. Phillips, 2009, Bandwidth Choice for Interval Estimation in GMM Regression. Working paper, Department of Economics, UC San Diego.

[26] Sun, Y. and M. S. Kim, 2012, Simple and Powerful GMM Over-identification Tests with Accurate Size. Journal of Econometrics 166(2), 267–281.

[27] Sun, Y. and M. S. Kim, 2015, Asymptotic F Test in a GMM Framework with Cross Sectional Dependence. Review of Economics and Statistics 97(1), 210–223.

[28] Zhang, X, 2016, Fixed-smoothing Asymptotics in the Generalized Empirical Likelihood Estimation Framework. Journal of Econometrics 193(1), 123–146.

Online Supplementary Appendix

Asymptotic F and t Tests in an Efficient GMM Setting

by Jungbin Hwang and Yixiao Sun

In this supplementary appendix, we present some further intuition on the J-statistic modification. We show that our asymptotic F and t tests in the efficient GMM setting are related to the exact F and t tests in a classical normal linear regression (CNLR) setting.

For illustration purposes, we consider the following location model, which is perhaps the simplest model in an overidentified GMM setting:

$$y_{1t} = \theta_0 + u_{1t},\ y_{1t} \in \mathbb{R}^p,$$

$$y_{2t} = u_{2t},\ y_{2t} \in \mathbb{R}^q,$$

where θ_0 is the parameter of interest, and $u_t = (u_{1t}', u_{2t}')' \in \mathbb{R}^{p+q}$ is a mean zero stationary process that can exhibit autocorrelation of unknown forms. The long run variance of u_t is

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix},$$

which has been partitioned conformably with the two blocks of equations. As simple as it is, the location model captures all the essentials in a GMM setting. In fact, a general GMM model can be reduced to the above location model in an asymptotic sense. The location model is an ideal framework to present the basic ideas and intuition, as it abstracts away the unnecessary details and complications. For more discussion, see Hwang and Sun (2015).

At the mechanical level, the parameter θ_0 can be estimated using the GMM. The moment conditions are

$$E\begin{pmatrix} y_{1t} - \theta_0 \\ y_{2t} \end{pmatrix} = 0,$$

and the GMM estimator of θ_0 is $\hat{\theta}_{GMM} = \arg\min_{\theta\in\Theta} g_T(\theta)' W_T^{-1} g_T(\theta)$ with

$$g_T(\theta) = \begin{pmatrix} T^{-1}\sum_{t=1}^T y_{1t} - \theta \\ T^{-1}\sum_{t=1}^T y_{2t} \end{pmatrix}.$$

If we take $W_{o,T} = I_{p+q}$, we obtain the initial GMM estimator $\tilde{\theta}_T = \bar{y}_1 := T^{-1}\sum_{t=1}^T y_{1t}$, which is the OLS estimator based on the first block of equations. If we take W_T to be the long run variance estimator

$$\hat{\Omega} = \begin{pmatrix} \hat{\Omega}_{11} & \hat{\Omega}_{12} \\ \hat{\Omega}_{21} & \hat{\Omega}_{22} \end{pmatrix} = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q_K\left(\frac{t}{T}, \frac{s}{T}\right)(y_t - \bar{y})(y_s - \bar{y})', \tag{S.1}$$

where $y_t = (y_{1t}', y_{2t}')'$ and $\bar{y} = T^{-1}\sum_{t=1}^{T} y_t$, then we obtain the efficient two-step GMM estimator: $\hat{\theta}_T = \bar{y}_1 - \hat{\beta}\bar{y}_2$ with

$$\hat{\beta} = \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1},$$

which is an estimator of the long run regression coefficient $\beta_0 = \Omega_{12}\Omega_{22}^{-1}$. Compared with the initial estimator $\tilde{\theta}_T$, which ignores the second block of equations, the two-step estimator $\hat{\theta}_T$ aims to exploit the additional information embodied in the second block.

As a special case of the GMM setting, the location model permits the asymptotic F and t tests as described in the previous section. Some calculation shows that the J statistic is given by

$$\mathbb{J}_T = T g_T(\hat{\theta}_T)'\hat{\Omega}^{-1} g_T(\hat{\theta}_T) = (\sqrt{T}\bar{y}_2)'\hat{\Omega}_{22}^{-1}(\sqrt{T}\bar{y}_2). \tag{S.2}$$

This is quite intuitive. The overidentifying moment conditions are the second block of moment conditions. To test for overidentification is to test whether the second block of time series, i.e., $\{y_{2t}\}$, has mean zero. The J statistic above is exactly the usual Wald statistic for such a test.

To demystify the asymptotic F and t tests, we cast the GMM estimator as an OLS estimator in a linear regression model. Let

$$\begin{aligned} \omega_i(y_1) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_i\left(\frac{t}{T}\right) y_{1t},\ \omega_i(y_2) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_i\left(\frac{t}{T}\right) y_{2t} \\ \omega_i(u_1) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_i\left(\frac{t}{T}\right) u_{1t},\ \omega_i(u_2) = \omega_i(y_2), \\ x_i &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_i\left(\frac{t}{T}\right) \text{ for } i = 0, 1, \ldots, K. \end{aligned}$$

These transforms are analogous to the Fourier transforms and are designed to capture the long run behavior of the underlying processes. Then

$$\begin{aligned} \omega_i(y_1) &= \theta_0 x_i + \omega_i(u_1) \\ \omega_i(y_2) &= \omega_i(u_2) \end{aligned}$$

for $i = 0, 1, \ldots, K$. This can be regarded as a system of cross-sectional regressions with dependent variables $\omega_i(y_1)$ and $\omega_i(y_2)$ and sample size $K + 1$.

To obtain an efficient estimator of θ_0, we use $\omega_i(u_2)$ to predict and hence reduce the error term in the first block of equations. This is equivalent to adding $\omega_i(y_2)$ to the first block of equations, leading to the regression model of the form:

$$\omega_i(y_1) = \theta_0 x_i + \beta_0 \omega_i(y_2) + \omega_i(\varepsilon),$$

where as before $\beta_0 = \Omega_{12}\Omega_{22}^{-1} \in \mathbb{R}^{p\times q}$, $\varepsilon = u_1 - \beta_0 u_2$, and $\omega_i(\varepsilon) = \omega_i(u_1) - \beta_0\omega_i(u_2)$ is the new error term. Under some assumptions on the basis functions and the error process, we have

$$\left(\begin{array}{c} \omega_i(u_1) \\ \omega_i(u_2) \end{array}\right) \overset{d}{\to} iidN(0,\Omega).$$

Hence the error term $\omega_i(\varepsilon)$ is asymptotically normal. More specifically, $\omega_i(\varepsilon)$ is asymptotically iid $N(0,\Omega_{11\cdot 2})$ where

$$\Omega_{11\cdot 2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21} \in \mathbb{R}^{p\times p}.$$

In addition, $\omega_i(\varepsilon)$ is asymptotically independent of $\omega_i(y_2)$.

The above model is close to a CNLR model with fixed regressors. However, there are three differences. First, the normality of the error term and its independence from the regressors hold only asymptotically. To remove this difference and for simplicity, we assume that normality holds exactly from now on, i.e., $\omega_i(\varepsilon) \sim iid\ N(0,\Omega_{11\cdot 2})$ and that $\omega_i(\varepsilon)$ is independent of $\omega_i(y_2)$. Finite sample results obtained under these assumptions then hold asymptotically without these assumptions. Second, when $p > 1$, we have a system of regressions while there is typically only one regression in a CNLR model. Of course, we can focus on the case of $p = 1$ to gain some insights but we will consider a general p. Third, $\omega_i(y_2)$ is random rather than fixed. This is innocuous, as we can follow the standard practice and use the conditioning argument.

To simplify the presentation, we assume that $T^{-1}\sum_{t=1}^{T}\Phi_k(t/T) = 0$ for $k = 1, 2, \ldots, K$[3]. This assumption holds for $\Phi_k(t/T) = \sqrt{2}\sin(2\pi kt/T)$, $\sqrt{2}\cos(2\pi kt/T)$, which are the basis functions in common use for the series LRV estimation. In this case, $x_0 = \sqrt{T}$ and $x_k = 0$ for $k = 1, 2, \ldots, K$. The OLS estimators of θ_0 and β_0 are then given by

$$(\hat{\theta}_{T,OLS}, \hat{\beta}_{T,OLS}) = \arg\min_{\theta,\beta}\left\{\left\|\omega_0(y_1) - \theta x_0 - \beta\omega_0(y_2)\right\|^2 + \sum_{k=1}^{K}\left\|\omega_k(y_1) - \beta\omega_k(y_2)\right\|^2\right\}.$$

[3]If, instead of $T^{-1}\sum_{t=1}^{T}\Phi_k(t/T) = 0$, we have $\int_0^1 \Phi_k(r)dr = 0$, then the results in this appendix still hold approximately in large samples.

Observe that, no matter what value β takes, we can choose θ to make the first term zero. For this reason, the OLS estimator of β_0 is determined solely by minimizing the second term. Let

$$S_{22} = \sum_{i=1}^{K} \omega_i(y_2)\,\omega_i(y_2)' \text{ and } S_{12} = \sum_{i=1}^{K} \omega_i(y_1)\,\omega_i(y_2)',$$

then $\hat{\beta}_T, OLS = S_12S_22^-1$ and

$$\hat{\theta}_{T,OLS} = x_0^{-1}\left[\omega_0(y_1) - \hat{\beta}_{T,OLS}\omega_0(y_2)\right] = \bar{y}_1 - \hat{\beta}_{T,OLS}\bar{y}_2. \tag{S.3}$$

But it is easy to see that $S_12S_22^-1 = \hat{\Omega}_12\hat{\Omega}_22^-1 = \hat{\beta}$. Therefore, $\hat{\theta}_T, OLS = \bar{y}_1 - \hat{\beta}\bar{y}_2$, and $\hat{\theta}_T, OLS$ is numerically identical to $\hat{\theta}_T, GMM := \hat{\theta}_T$. Here we have added the subscript 'GMM' to $\hat{\theta}_T$ to signify its origin.

Given the numerical equivalence, it is interesting to see how $\hat{\theta}_{T,GMM}$ is scaled for hypothesis testing in the CNLR framework. Let

$$\omega_1 = \begin{pmatrix} \omega_0'(y_1) \\ \omega_1'(y_1) \\ \cdots \\ \omega_K'(y_1) \end{pmatrix}_{(K+1)\times p}, \omega_2 = \begin{pmatrix} \omega_0'(y_2) \\ \omega_1'(y_2) \\ \cdots \\ \omega_K'(y_2) \end{pmatrix}_{(K+1)\times q},$$

$$\omega_\varepsilon = \begin{pmatrix} \omega_0'(\varepsilon) \\ \omega_1'(\varepsilon) \\ \cdots \\ \omega_K'(\varepsilon) \end{pmatrix}_{(K+1)\times p}, \text{ and } X = \begin{pmatrix} x_0 \\ x_1 \\ \cdots \\ x_K \end{pmatrix}_{(K+1)\times 1}.$$

Then

$$\omega_1 = X\theta_0' + \omega_2\beta_0' + \omega_\varepsilon, \tag{S.4}$$

and the OLS estimator of θ_0 is given by

$$\hat{\theta}_{T,OLS} = \left(\omega_1' M_2 X\right)\left(X' M_2 X\right)^{-1},$$

where $M_2 = I_{K+1} - \omega_2\left(\omega_2'\omega_2\right)^{-1}\omega_2'$. This is just an alternative and standard way to represent the OLS estimator in a linear regression. It is numerically identical to what is given in (S.3).

Conditional on ω_2,

$$\hat{\theta}_{T,OLS} - \theta_0 = \left(\omega_\varepsilon' M_2 X\right)\left(X' M_2 X\right)^{-1} \sim N\left[0,\ \left(X' M_2 X\right)^{-2} var\left(\omega_\varepsilon' M_2 X | \omega_2\right)\right].$$

Note that for any $a = (a_0, a_1, \ldots, a_K)' \in \mathbb{R}^{K+1}$ the variance of $\omega_\varepsilon' a = \sum_{j=0}^{K} a_i \omega_j(\varepsilon)$ is $\Omega_{11\cdot 2}(a'a)$, and so $var(\omega_\varepsilon' M_2 X | \omega_2)$ is $\Omega_{11\cdot 2} X' M_2 X$. Therefore, conditional on ω_2,

$$\hat{\theta}_{T,OLS} - \theta_0 \sim N\left[0, \Omega_{11\cdot 2}\left(X' M_2 X\right)^{-1}\right].$$

This implies that $\hat{\theta}_{T,OLS} - \theta_0$ is mixed normal unconditionally. We have therefore reproduced the (asymptotic) mixed normality of the two-step GMM estimator, which has been recently established by Sun (2014) in a general GMM setting.

Now suppose that we follow the mechanics in the CNLR framework to conduct inference. Conditional on $(X' M_2 X)^{-1}$, the variance of $\hat{\theta}_{T,OLS}$ is $\Omega_{11\cdot 2}(X' M_2 X)^{-1}$. Following a routine in the CNLR framework, we can estimate the conditional variance by $\tilde{\Omega}_{11\cdot 2}(X' M_2 X)^{-1}$ where

$$\tilde{\Omega}_{11\cdot 2} = \frac{1}{K-q}(\omega_1 - X\hat{\theta}_{T,OLS}' - \omega_2 \hat{\beta}_{T,OLS}')'(\omega_1 - X\hat{\theta}_{T,OLS}' - \omega_2 \hat{\beta}_{T,OLS}').$$

Here we have used $1/(K-q) = 1/(K+1-q-1)$ instead of $1/(K+1)$ as the scaling function. This is the usual degree-of-freedom correction in a standard linear system of regressions with sample size $K+1$ and number of (vector) regressors $q+1$. Constructing the Wald statistic for testing $H_0 : \theta_0 = r$ in the same way as what we would do in a CNLR framework, we obtain the (normalized) Wald statistic

$$\mathbb{W}_{CNLR} = \sqrt{T}(\hat{\theta}_{T,OLS} - r)' \left[\tilde{\Omega}_{11\cdot 2}\left(\frac{X' M_2 X}{T}\right)^{-1}\right]^{-1} \sqrt{T}(\hat{\theta}_{T,OLS} - r)/p.$$

This is a finite sample construction for any sample size $K+1$ and thus involves no asymptotic approximation.

To formally compare $\mathbb{W}_{CNLR}$ with the unmodified GMM Wald statistic reproduced below:

$$\mathbb{W}_T := \mathbb{W}_T(\hat{\theta}_T) = T(R\hat{\theta}_T - r)' \left\{R\left[G_T(\hat{\theta}_T)' W_T^{-1}(\hat{\theta}_T) G_T(\hat{\theta}_T)\right]^{-1} R'\right\}^{-1} (R\hat{\theta}_T - r)/p, \quad \text{(S.5)}$$

we note that for the location model $G_T(\hat{\theta}_T) = (I_p, O_{p\times q})'$. Using this and plugging $W_T = \hat{\Omega}$ and $R = I_p$ into (S.5), we obtain

$$\mathbb{W}_T = \sqrt{T}(\hat{\theta}_{T,GMM} - r)'[\hat{\Omega}_{11\cdot 2}]^{-1}\sqrt{T}(\hat{\theta}_{T,GMM} - r)/p, \quad \text{(S.6)}$$

where $\hat{\Omega}_{11\cdot 2} = \hat{\Omega}_{11} - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}$ and $\hat{\Omega}_{ij}$ are given in (S.1). It is not hard to show that $(K-q)\tilde{\Omega}_{11\cdot 2} = K\hat{\Omega}_{11\cdot 2}$; see (S.7) below. The substantive difference between $\mathbb{W}_{CNLR}$ and $\mathbb{W}_T$ is that $\mathbb{W}_{CNLR}$ has the additional factor $(X' M_2 X/T)^{-1}$ in the variance estimator. The presence of this factor is necessary for $\mathbb{W}_{CNLR}$ to follow the standard F distribution in the CNLR framework. This suggests that such a factor is also indispensable for $\mathbb{W}_T$ to be asymptotically standard F distributed in the GMM framework.

What is the variance correction factor $X'M_2X/T$? In the proof of Proposition S.1 below, it is shown that

$$\left(X'M_2X/T\right)^{-1} = 1 + T\bar{y}_2' S_{22}^{-1}\bar{y}_2;$$

see equation (S.8). But $T\bar{y}_2' \left(S_{22}/K\right)^{-1}\bar{y}_2$ is exactly the J statistic given in (S.2). So

$$\left(X'M_2X/T\right)^{-1} = 1 + \mathbb{J}_T/K.$$

The variance correction factor, which appears naturally in the CNLR framework, suggests the J-statistic correction in the GMM framework.

The following proposition establishes the connection between $\mathbb{W}_{CNLR}$ and $\mathbb{W}_T^c$ rigorously.

Proposition S.1 *Assume that the basis functions $\Phi_j(\cdot)$ are piecewise monotonic, continuously differentiable and orthonormal in $L^2[0,1]$ and satisfy $T^{-1}\sum_{t=1}^T \Phi_k(t/T) = 0$ for $k = 1, 2, \ldots, K$. Then*

$$\mathbb{W}_{CNLR} = \frac{K-q}{K-p-q+1}\mathbb{W}_T^c.$$

When $p = 1$, the proposition shows that the Wald statistic constructed in the standard way in a CNLR is numerically identical to the modified Wald statistic we propose in the GMM setting. While the modification can be motivated on the ground of obtaining a convenient standard F limiting distribution, it is a built-in feature of the standard Wald statistic in a linear regression. The modification may appear to be mysterious at first sight, but it becomes natural from the regression perspective.

Proof of Proposition S.1. Observing that

$$\omega_0(y_1) - \hat{\theta}_{T,OLS}x_0 - \hat{\beta}_{T,OLS}\omega_0(y_2) = 0$$

and using the definition of $\tilde{\Omega}_{11\cdot 2}$, we have

$$\begin{aligned}
\tilde{\Omega}_{11\cdot 2} &= \frac{1}{K-q}\sum_{k=1}^{K}\left[\omega_k(y_1) - S_{12}S_{22}^{-1}\omega_k(y_2)\right]\left[\omega_k(y_1) - S_{12}S_{22}^{-1}\omega_k(y_2)\right]' \\
&= \frac{1}{K-q}\left(S_{11} - S_{21}'S_{22}^{-1}S_{21}\right) = \frac{K}{K-q}\left(\hat{\Omega}_{11} - \hat{\Omega}_{21}'\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}\right) \\
&= \frac{K}{K-q}\hat{\Omega}_{11\cdot 2},
\end{aligned} \tag{S.7}$$

where we have used $S_{ij} = K\hat{\Omega}_{ij}$.

Next, let $e_{K+1} = (1, 0, \ldots, 0)'$ be the first unit vector in $\mathbb{R}^{K+1}$, then

$$\frac{X'M_2X}{T} = e'_{K+1}M_2e_{K+1} = 1 - e'_{K+1}\omega_2\left(\omega_2'\omega_2\right)^{-1}\omega_2'e_{K+1}.$$

Using the Sherman-Morrison formula, we have

$$\begin{aligned}
& e'_{K+1}\omega_2\left(\omega_2'\omega_2\right)^{-1}\omega_2'e_{K+1} \\
&= \omega_0\left(y_2\right)'\left[\omega_0\left(y_2\right)\omega_0'\left(y_2\right) + \sum_{i=1}^{K}\omega_i\left(y_2\right)\omega_i\left(y_2\right)'\right]^{-1}\omega_0\left(y_2\right) \\
&= T\bar{y}_2'\left(T\bar{y}_2\bar{y}_2' + S_{22}\right)^{-1}\bar{y}_2 = T\bar{y}_2'\left[S_{22}^{-1} - \frac{S_{22}^{-1}\left(T\bar{y}_2\bar{y}_2'\right)S_{22}^{-1}}{1 + T\bar{y}_2'S_{22}^{-1}\bar{y}_2}\right]\bar{y}_2 \\
&= T\bar{y}_2'S_{22}^{-1}\bar{y}_2 - \frac{\left(T\bar{y}_2'S_{22}^{-1}\bar{y}_2\right)^2}{1 + T\bar{y}_2'S_{22}^{-1}\bar{y}_2} = \frac{T\bar{y}_2'S_{22}^{-1}\bar{y}_2}{1 + T\bar{y}_2'S_{22}^{-1}\bar{y}_2}.
\end{aligned}$$

Therefore,

$$\frac{X'M_2X}{T} = 1 - \frac{T\bar{y}_2'S_{22}^{-1}\bar{y}_2}{1 + T\bar{y}_2'S_{22}^{-1}\bar{y}_2} = \frac{1}{1 + T\bar{y}_2'S_{22}^{-1}\bar{y}_2} = \frac{1}{1 + \mathbb{J}_T/K}. \tag{S.8}$$

Combining the above with (S.7), we have

$$\begin{aligned}
\mathbb{W}_{CNLR} &= \sqrt{T}(\hat{\theta}_{T,OLS} - r)'\left[\tilde{\Omega}_{11\cdot 2}\left(\frac{X'M_2X}{T}\right)^{-1}\right]^{-1}\sqrt{T}(\hat{\theta}_{T,OLS} - r)/p \\
&= \frac{K-q}{K}\sqrt{T}(\hat{\theta}_{T,GMM} - r)'\hat{\Omega}_{11\cdot 2}^{-1}\sqrt{T}(\hat{\theta}_{T,GMM} - r)\frac{1}{p}\cdot\frac{1}{1 + \mathbb{J}_T/K} \\
&= \frac{K-q}{K}\frac{\mathbb{W}_T(\hat{\theta}_{T,GMM})}{1 + \mathbb{J}_T/K} \\
&= \frac{K-q}{K-p-q+1}\left[\frac{K-p-q+1}{K}\frac{\mathbb{W}_T(\hat{\theta}_{T,GMM})}{1 + \mathbb{J}_T/K}\right] \\
&= \frac{K-q}{K-p-q+1}\mathbb{W}_T^c(\hat{\theta}_{T,GMM})
\end{aligned}$$

as desired. ■

References

[1] Hwang, J. and Y. Sun, 2015, Should We Go One Step Further? An Accurate Comparison of One-Step and Two-Step Procedures in a Generalized Method of Moments Framework. Working paper, Department of Economics, UC San Diego.

[2] Sun, Y., 2014, Fixed-smoothing Asymptotics in a Two-step GMM Framework. Econometrica 82(6), 2327–2370.