Asymptotic F and t Tests in an Efficient GMM Setting

Jungbin Hwang and Yixiao Sun*
Department of Economics,
University of California, San Diego

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 that use standard F critical values. Simulation shows that, in terms of both size and power, the asymptotic F tests perform as well as the nonstandard tests proposed recently by Sun (2014b) in finite samples. But the F tests are more appealing as the critical values are readily available from standard statistical tables. Compared to the conventional chi-square tests, the F tests are as powerful, but are much more accurate in size.

JEL Classification: C12, C32

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

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 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

*Email: j6hwang@ucsd.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. Sun gratefully acknowledges partial research support from NSF under Grant No. SES-1530592.

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, 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. 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 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.

The paper contributes to a growing body of literature on the 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 the fixed-b asymptotics first studied by Kiefer and Vogelsang (2002a, 2002b, 2005) in the econometrics literature. Subsequent research includes Jansson (2004), Sun, Phillips, Jin (2008), Sun and Phillips (2009), Gonçlaves and Vogelsang (2011) and 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 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 and the Wald 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 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 casts the GMM estimator as an OLS estimator in a regression setting and shows that the modified Wald statistic is the usual Wald statistic in a CNLR model. The next section reports simulation evidence. The last section concludes. Proofs are given in the appendix.

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 $\{f\left(v_t,\theta_0\right)\}$ 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 $W_{o,T}\xrightarrow{p}W_{o,\infty}$, a matrix that is positive definite almost surely.

With the first step estimator $\tilde{\theta}_T$, we can construct 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 following form

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

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

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

where $\{\Phi_j\left(r\right)\}$ are orthonormal basis functions on $L^2[0,1]$ satisfying $\int_0^1\Phi_j\left(r\right)dr=0$. 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

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

Then

$$W_T(\theta_0) = \frac{1}{K}\sum_{j=1}^{K} \Lambda_j(\theta_0)\Lambda_j'(\theta_0). \tag{3}$$

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 $\Phi_j(r) = \sqrt{2}\sin(2\pi jr)$ or $\sqrt{2}\cos(2\pi jr)$, then the series LRV estimator is proportional to the spectral density estimator at the origin that takes a simple average of the first K 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. 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

$$\mathbb{W}_T := \mathbb{W}_T(\hat{\theta}_T) = T(R\hat{\theta}_T - r)' \left\{ R\left[G_T(\hat{\theta}_T)' W_T^{-1}(\hat{\theta}_T) G_T(\hat{\theta}_T)\right]^{-1} R' \right\}^{-1} (R\hat{\theta}_T - r)/p, \tag{4}$$

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:

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

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 given by

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

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 given by

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

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

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

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

$$\mathbb{W}_T^c := \mathbb{W}_T^c(\hat{\theta}_T) = \frac{K-p-q+1}{K} \frac{\mathbb{W}_T(\hat{\theta}_T)}{1+\frac{1}{K}J_T(\hat{\theta}_T)},$$
$$\mathbb{D}_T^c := \mathbb{D}_T^c(\hat{\theta}_T) = \frac{K-p-q+1}{K} \frac{\mathbb{D}_T(\hat{\theta}_T)}{1+\frac{1}{K}J_T(\hat{\theta}_T)},$$
$$\mathbb{S}_T^c := \mathbb{S}_T^c(\hat{\theta}_T) = \frac{K-p-q+1}{K} \frac{\mathbb{S}_T(\hat{\theta}_T)}{1+\frac{1}{K}J_T(\hat{\theta}_T)}.$$

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

$$t_T^c(\hat{\theta}_T) = \frac{K-q}{K} \frac{t_T(\hat{\theta}_T)}{1+\frac{1}{K}J_T(\hat{\theta}_T)}.$$

Under the conventional asymptotic theory where K diverges to ∞ with the sample size T but $K/T \to 0$, both correction factors $K-p-q+1/K$ and $(1+J_T(\hat{\theta}_T)/K)^{-1}$ approach unity. So they do not matter in large samples and can thus be regarded as finite sample corrections. Under this type of asymptotics, $\mathbb{W}_T$, $\mathbb{D}_T$ and $\mathbb{S}_T$ and hence $\mathbb{W}_T^c$, $\mathbb{D}_T^c$ and $\mathbb{S}_T^c$ are all 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 The Asymptotic F and t Tests

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 $u_t = f(v_t, \theta_0)$ and $\Phi_0(t) \equiv 1$, $e_t \sim iidN(0, I_m)$. 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 *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 $\theta_0 \in \Theta$, a compact parameter space.*

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

Assumption 4 *(a) $f(v_t, \theta)$ is twice continuously differentiable in θ for almost all v_t. (b) For any $\theta_T = \theta_0 + o_p(1)$, $plim_{T\to\infty} G_{[rT]}(\theta_T) = rG$ 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 $j = 0, 1, \ldots, J$ for every finite J.*

(b) The following holds:

$$
\begin{aligned}
&P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)u_t \le x \text{ for } j = 0, 1, \ldots, J\right) \\
&= P\left(\Lambda\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)e_t \le x \text{ for } j = 0, 1, \ldots, J\right) + o(1) \text{ as } T \to \infty
\end{aligned}
$$

for every finite J 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

$$
\begin{aligned}
C_{pp} &= \int_0^1\int_0^1 Q_K(r,s)dB_p(r)dB_p(s)',\ C_{pq} = \int_0^1\int_0^1 Q_K(r,s)dB_p(r)dB_q(s)' \qquad (5)\\
C_{qq} &= \int_0^1\int_0^1 Q_K(r,s)dB_q(r)dB_q(s)',\ D_{pp} = C_{pp} - C_{pq}C_{qq}^{-1}C_{pq}'.
\end{aligned}
$$

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

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

(b) $t_T(\hat{\theta}_T) \xrightarrow{d} \left[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right]/\sqrt{D_{pp}} :\overset{d}{=} t_\infty$,

(c) $J_T(\hat{\theta}_T) \xrightarrow{d} B_q'(1)C_{qq}^{-1}\left[B_q(1)\right] :\overset{d}{=} J_\infty$,

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

The weak convergence of the marginal distributions in Theorem 1(a,b) and 1(c) has been established in Sun (2014b) and Sun and Kim (2012), respectively. It suffices to show that the weak convergence holds jointly. A proof is given in the appendix.

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

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

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 Wishart distribution and its independence of (C_{pq}, C_{qq}) are the two key properties of D_{pp} that drive our F limit theory. For kernel LRV estimation, D_{pp} will not be Wishart and will not be independent of (C_{pq}, C_{qq}). So an exact F limit theory is not possible.

Remark 2 *Note that* $\Delta = C_{pq}C_{qq}^{-1}B_q(1)$ *is independent of* $B_p(1)$ *and* D_{pp}*, the limiting distribution* F_∞ *in Theorem 1(a) conditional on* Δ *satisfies*

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

which is a noncentral F distribution with noncentrality parameter $\|\Delta\|^2$. *Unconditionally,* $\frac{K-p-q+1}{K}F_\infty$ *follows a mixed noncentral F distribution, i.e., a noncentral F distribution with a random noncentrality parameter. The noncentral F test proposed in Sun (2014b) is based on the noncentral F approximation to the mixed F distribution.*

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

$$\frac{K-q+1}{Kq}J_T(\hat{\theta}_T) \xrightarrow{d} F_{q,K-q+1}, \tag{6}$$

where $F_{q,K-q+1}$ *is the standard F distribution with degrees of freedom* q *and* $K-q+1$*. This is a result first established in Sun and Kim (2012).*

Using Theorem 1, we have

$$\begin{aligned}\mathbb{W}_T^c(\hat{\theta}_T) &= \frac{K-p-q+1}{K}\frac{\mathbb{W}_T(\hat{\theta}_T)}{1+\frac{1}{K}J_T(\hat{\theta}_T)} \\ &\xrightarrow{d} \frac{K-p-q+1}{K}\frac{F_\infty}{1+\frac{1}{K}J_\infty} = \frac{K-p-q+1}{K}\xi_p' D_{pp}^{-1}\xi_p\end{aligned}$$

where

$$\xi_p := \frac{B_p(1)-C_{pq}C_{qq}^{-1}B_q(1)}{\sqrt{1+\frac{1}{K}J_\infty}} = \frac{B_p(1)-C_{pq}C_{qq}^{-1}B_q(1)}{\sqrt{1+\frac{1}{K}B_q'(1)C_{qq}^{-1}B_q(1)}}.$$

Another key result that drives the F limit theory is that $\xi_p \sim N(0, I_p)$. This holds for the case of series LRV estimation but not for the kernel LRV estimation. The result is proved in the proof of Theorem 2 using the conditioning argument with J_∞ as the conditioning variable. This is in contrast with Sun (2014b) which uses Δ or $\|\Delta\|^2$ as the conditioning variable. Given that $\xi_p \sim N(0, I_p)$ and that ξ_p is independent of D_{pp}, $F_\infty\left(1+K^{-1}J_\infty\right)^{-1} = \xi_p' D_{pp}^{-1}\xi_p$ 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) $\mathbb{W}_T^c(\hat{\theta}_T) \xrightarrow{d} F_{p,K-p-q+1}$;

(b) $\mathbb{D}_T^c(\hat{\theta}_T) \xrightarrow{d} F_{p,K-p-q+1}$;

(c) $\mathbb{S}_T^c(\hat{\theta}_T) \xrightarrow{d} F_{p,K-p-q+1}$;

(d) $t_T^c(\hat{\theta}_T) \xrightarrow{d} t_{K-q}$.

Remark 4 *When $q = 0$, we have $J_T(\hat{\theta}_T) = 0$ and the multiplicative correction degenerates. In this case, we have*

$$\frac{K-p+1}{K}\mathbb{W}_T(\hat{\theta}_T) \xrightarrow{d} F_{p,K-p+1}.$$

This is identical to a result obtained in Sun (2013) for the Wald test 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 5 *It follows from (6) that*

$$\frac{1}{K}J_T(\hat{\theta}_T) \xrightarrow{d} \frac{q}{K-q+1}F(q, K-q+1) \overset{d}{=} \frac{\chi_q^2}{\chi_{K-q+1}^2}$$

for two independent chi-square random variables χ_q^2 and χ_{K-q+1}^2. So, as K increases for a fixed q, $J_T(\hat{\theta}_T)/K$ approaches zero and the modified Wald statistic becomes close to the original Wald statistic. The multiplicative correction $1 + J_T(\hat{\theta}_T)/K$ 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 under the conventional increasing-smoothing asymptotics, as $(K-p-q+1)/K \to 1$ as $K \to \infty$. This correction factor can be motivated from the Bartlett correction. See Sun (2013) for more discussion.

Remark 6 *Let $F_{p,K-p-q+1}^{\alpha}$ be the $(1-\alpha)$ quantile of the F distribution $F_{p,K-p-q+1}$. According to Theorem 2, the critical value for the original test statistic $\mathbb{W}_T(\hat{\theta}_T)$ can be taken to be*

$$\left[1 + \frac{1}{K}J_T(\hat{\theta}_T)\right]\left[\frac{K}{K-p-q+1}\right]F_{p,K-p-q+1}^{\alpha}. \tag{7}$$

Compare with the chi-square critical value χ_p^{α}/p where χ_p^{α} is the $(1-\alpha)$ quantile of the chi-squared distribution χ_p^2, the above critical value is larger for three reasons. First, $F_{p,K-p-q+1}^{\alpha} > \chi_p^{\alpha}/p$ due to the random denominator in the F distribution. Second, $K/(K-p-q+1) > 1$ for $q > 1$ or $p > 1$. Third, $1 + J_T(\hat{\theta}_T)/K > 1$ almost surely. A direct implication is that the chi-square critical values are 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.

Remark 7 *If we use the kernel LRV estimator, then we can choose an equivalent K value and use the critical value in (7). According to Sun and Kim (2012), the equivalent K value is given by the integer that is closest to*

$$\frac{\left[\int_0^1 k_b(r,r)\,dr\right]^2}{\int_0^1\int_0^1 [k_b(r,s)]^2\,drds}, \tag{8}$$

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 (7) approaches the chi-squared critical value χ_p^α/p. In fact, as $b \to 0$, we can take

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

which provides a good approximation to (8). 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 (7) with an equivalent K is convenient to use and may be more appealing in applied research.

Remark 8 *In the proof of the theorem, we show that conditional on $B_q(\cdot)$, $\xi_p \sim N(0, I_p)$. Since the conditional distribution does not depend on $B_q(\cdot)$, we can conclude that ξ_p is independent of $B_q(\cdot)$. As a result, ξ_p is independent of $B_q(1)$ 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*

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

where $\mathcal{F}_{p,K-p-q+1} \sim F_{p,K-p-q+1}$, $\mathcal{J}_\infty \sim J_\infty$, $\mathcal{F}_{p,K-q+1} \sim F_{p,K-q+1}$ and $\mathcal{F}_{p,K-p-q+1}$ is independent of $\mathcal{J}_\infty$ and $\mathcal{F}_{p,K-q+1}$. 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 F_∞.

Remark 9 *Let CV^α be the nonstandard critical value for $\left[(K-p-q+1)/K\right]\mathbb{W}_T(\hat{\theta}_T)$ as proposed in Sun (2014b). Using the characterization in the previous remark, we have*

$$\begin{aligned}
&\lim_{T\to\infty} P\left(\frac{K-p-q+1}{K}\mathbb{W}_T(\hat{\theta}_T) > CV^\alpha\right) \\
&= P\left[\mathcal{F}_{p,K-p-q+1}\left(1 + \frac{1}{K}\mathcal{J}_\infty\right) > CV^\alpha\right] = P\left[\mathcal{F}_{p,K-p-q+1} > \frac{CV^\alpha}{1 + \frac{1}{K}\mathcal{J}_\infty}\right] \\
&= 1 - EF_{p,K-p-q+1}\left(\frac{CV^\alpha}{1 + \frac{1}{K}\mathcal{J}_\infty}\right) = \alpha.
\end{aligned}$$

That is, the asymptotic level of the nonstandard test is α when averaging over all realizations of $\mathcal{J}_\infty$. Conditional on $\mathcal{J}_\infty$, the asymptotic level is

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

which is strictly increasing in $\mathcal{J}_\infty$. So when the J statistic is large, which implies a large $\mathcal{J}_\infty$ in large samples, the nonstandard Wald test is expected to reject the null more often. In contrast, the critical value in (7) is based on the conditional distribution of $\left[(K-p-q+1)/K\right]\mathbb{W}_T(\hat{\theta}_T)$ conditional on $J_T(\hat{\theta}_T)$. With the conditional critical value, the asymptotic conditional level of the test is fixed at α regardless of the value of $J_T(\hat{\theta}_T)$.

4 Understanding the Asymptotic F and t Tests

The asymptotic F and t tests may appear mysterious at first sight. To shed some light on the two tests, we consider the location model, which is perhaps the simplest model in an overidentified GMM setting:

$$\begin{aligned} y_{1t} &= \theta_0 + u_{1t},\ y_{1t} \in \mathbb{R}^p, \\ y_{2t} &= u_{2t},\ y_{2t} \in \mathbb{R}^q, \end{aligned} \tag{9}$$

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

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

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

$$\hat{\Omega} = \begin{pmatrix} \hat{\Omega}_{11} & \hat{\Omega}_{12} \\ \hat{\Omega}_{21} & \hat{\Omega}_{22} \end{pmatrix} = \frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T Q_K\left(\frac{t}{T}, \frac{s}{T}\right)(y_t - \bar{y})(y_s - \bar{y}), \tag{10}$$

where $y_t = (y_{1t}', y_{2t}')'$, 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 explore the additional information embodied in the second block. As a special case of the GMM setting, the location model permits the asymptotic F tests and t test as described in the previous section.

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 Assumptions 1–5 for the location model, of which Assumptions 2 and 4 hold trivially, we have

$$\begin{pmatrix} \omega_i(u_1) \\ \omega_i(u_2) \end{pmatrix} \xrightarrow{d} iidN(0, \Omega).$$

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 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 $\omega_i(y_2)$. The 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.

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

$$\omega_1 = X\theta_0' + \omega_2\beta_0' + \omega_\varepsilon.$$

Based on this, we obtain the OLS estimator of θ_0' below:

$$\hat{\theta}_{T,OLS}' = \left(X'M_2X\right)^{-1}\left(X'M_2\omega_1\right),$$

where $M_2 = I_{K+1} - \omega_2\left(\omega_2'\omega_2\right)^{-1}\omega_2'$. Conditional on ω_2, we have

$$(\hat{\theta}_{T,OLS}' - \theta_0') \sim N\left[0, \Omega_{11\cdot 2}\left(X'M_2X\right)^{-1}\right].$$

Hence it is mixed normal unconditionally. This result is analogous to the asymptotic mixed normality of the two-step GMM estimator. In fact we can show that $\hat{\theta}_{T,OLS}$ and the two-step GMM estimator $\hat{\theta}_{T,GMM} \equiv \hat{\theta}_T$ are numerically identical under a slightly stronger condition on the basis functions. Here we add the subscript 'GMM' to $\hat{\theta}_T$ to signify its origin.

Proposition 3 *Let Assumption 1 hold with $\int_0^1 \Phi_k(r)\,dr = 0$ replaced by $T^{-1}\sum_{t=1}^T \Phi_k(t/T) = 0$ for $k = 1, 2, \ldots, K$, then $\hat{\theta}_{T,OLS} = \hat{\theta}_{T,GMM}$. If $\int_0^1 \Phi_k(r)\,dr = 0$ but not $T^{-1}\sum_{t=1}^T \Phi_k(t/T) = 0$ for $k = 1, 2, \ldots, K$, then under Assumptions 1–5, we have $\sqrt{T}(\hat{\theta}_{T,OLS} - \hat{\theta}_{T,GMM}) = o_p(1)$ for a fixed K as $T \to \infty$.*

While the asymptotic equivalence between $\hat{\theta}_{T,OLS}$ and $\hat{\theta}_{T,GMM}$ is well expected, it is nontrivial to show that they are numerically identical under the assumption that $T^{-1}\sum_{t=1}^T \Phi_k(t/T) = 0$. 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.

The conditional distribution of $(\hat{\theta}_{T,OLS}' - \theta_0')$ conditional on ω_2 depends on ω_2 only through $(X'M_2X)^{-1}$. It then follows that the conditional distribution of $(\hat{\theta}_{T,OLS}' - \theta_0')$ conditional on $(X'M_2X)^{-1}$ is also $N[0, \Omega_{11\cdot 2}(X'M_2X)^{-1}]$. In the proof of the proposition, it is shown that $(X'M_2X)^{-1} = (1 + T\bar{y}_2'\hat{\Omega}_{22}^{-1}\bar{y}_2/K)/T$. Therefore, we can take $T\bar{y}_2'\hat{\Omega}_{22}^{-1}\bar{y}_2$ as the conditioning variable. But $T\bar{y}_2'\hat{\Omega}_{22}^{-1}\bar{y}_2$ is exactly the J statistic in the overidentified location model. So the minimal conditioning variable in the CNLR coincides with the conditioning variable we use in the GMM framework.

Now suppose that we follow the mechanics in the CNLR framework to conduct inference. Conditional on $(X'M_2X)^{-1}$, the variance of $\hat{\theta}_{T,OLS}$ is $\Omega_{11\cdot 2}(X'M_2X)^{-1}$. Following a routine in the CNLR framework, we can estimate the conditional variance by $\tilde{\Omega}_{11\cdot 2}(X'M_2X)^{-1}$ where

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

and $\hat{\beta}'_{T,OLS}$ is the OLS estimator of β_0. 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 regression model. 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

$$\mathbb{W}_{CNLR} = \sqrt{T}\left(\hat{\theta}_{T,OLS} - r\right)' \left[\tilde{\Omega}_{11\cdot 2}\left(\frac{X'M_2X}{T}\right)^{-1}\right]^{-1} \sqrt{T}\left(\hat{\theta}_{T,OLS} - r\right)/p.$$

We can also construct other type statistics such as the LR, LM and t statistics but we focus on the Wald statistic here.

To formally compare $\mathbb{W}_{CNLR}$ with the unmodified GMM Wald statistic as given in (4), we note that for the location model $G_T(\hat{\theta}_T) = (I_p, O_{p\times q})'$. Using this and plugging $W_T(\hat{\theta}_{T,GMM}) = \hat{\Omega}$ and $R = I_p$ into (4), we obtain

$$\mathbb{W}_T = \sqrt{T}(\hat{\theta}_{T,GMM} - r)' \left[\hat{\Omega}_{11\cdot 2}\right]^{-1} \sqrt{T}(\hat{\theta}_{T,GMM} - r)/p, \tag{11}$$

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 (10). A formal comparison of $\mathbb{W}_{CNLR}$ with $\mathbb{W}_T$ reveals that $\mathbb{W}_{CNLR}$ has the additional factor $(X'M_2X/T)^{-1}$ in the variance estimator that the GMM Wald statistic $\mathbb{W}_T$ ignores. The reason that $\mathbb{W}_T$ ignores this factor 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, $(X'M_2X/T)^{-1} \to^p 1$ and so the factor is negligible in large samples. Under the fixed-smoothing asymptotics, it follows from Hwang and Sun (2015) that

$$\begin{aligned}\sqrt{T}\left(\hat{\theta}_{T,GMM} - \theta_0\right) &= \begin{pmatrix} I_p, & -\hat{\beta} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^T (y_{1t} - Ey_{1t}) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^T y_{2t} \end{pmatrix} \\ &\to^d \begin{pmatrix} I_p, & -\beta_\infty \end{pmatrix} \Lambda B_{p+q}(1),\end{aligned}$$

where

$$\beta_\infty = \Omega_{11\cdot 2}^{1/2}\tilde{\beta}_\infty\Omega_{22}^{-1/2} + \Omega_{12}\Omega_{22}^{-1} \text{ and } \tilde{\beta}_\infty = C_{pq}C_{qq}^{-1}. \tag{12}$$

Some simple calculations show that the asymptotic variance of $\hat{\theta}_{T,GMM}$ conditional on $\tilde{\beta}_\infty$ satisfies:

$$\text{avar}(\hat{\theta}_{T,GMM}) = \Omega_{11\cdot 2}^{1/2}\left(I_p + \tilde{\beta}_\infty\tilde{\beta}'_\infty\right)(\Omega_{11\cdot 2}^{1/2})' = \Omega_{11\cdot 2} + \Omega_{11\cdot 2}^{1/2}\tilde{\beta}_\infty\tilde{\beta}'_\infty(\Omega_{11\cdot 2}^{1/2})'.$$

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 $\tilde{\beta}_\infty\tilde{\beta}'_\infty$. This will not cause any problem for asymptotic pivotal inference but will prevent us from developing an 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 $\hat{\Omega}_{11\cdot 2}(1 + \hat{J}_T/K)$ as the asymptotic variance estimator.

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

Proposition 4 *Let Assumption 1 hold with $\int_0^1 \Phi_k(r)\,dr = 0$ replaced by $T^{-1}\sum_{t=1}^T \Phi_k(t/T) = 0$ for $i = 1, 2, \ldots, K$. Then*

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

In particular, $\mathbb{W}_{CNLR} = \mathbb{W}_T^c$ when $p = 1$. If $\int_0^1 \Phi_k(r)\,dr = 0$ but not $T^{-1}\sum_{t=1}^T \Phi_k(t/T) = 0$, then under Assumptions 1–5, we have $\mathbb{W}_{CNLR} = \frac{K-q}{K-p-q+1}\mathbb{W}_T^c + o_p(1)$ for a fixed K as $T \to \infty$.

Remark 10 *When $p = 1$, the proposition shows that the Wald statistic constructed in the standard way 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.*

Remark 11 *When $p > 1$, $\mathbb{W}_{CNLR}$ does not follow an F distribution but a rescaled version does:*

$$\frac{K - p - q + 1}{K - q} \mathbb{W}_{CNLR} \sim F_{p,K-p-q+1}.$$

This follows from Theorem 2 and Proposition 4. Of course this can be proved directly in the CNLR setting but there is no need to do so, as the limit theory established in the GMM setting is directly applicable to the CNLR model.

Remark 12 *Looking at the GMM problem from the regression perspective motivates us to use the modified Wald statistic even if there is no serial dependence. In this case, we can take $K = T$ and the modified Wald statistic becomes*

$$\mathbb{W}_T^c := \frac{T - p - q + 1}{T} \frac{\mathbb{W}_T}{1 + \frac{1}{T} J_T},$$

where $\mathbb{W}_T$ and J_T are the standard Wald and J statistics in the GMM framework with iid data. In addition, we use $F_{p,T-p-q+1}$ instead of χ_p^2/p as the reference distribution. From an asymptotic point of view, $\mathbb{W}_T^c = \mathbb{W}_T + o_p(1)$ and $F_{p,T-p-q+1}^{\alpha} = \chi_p^{\alpha}/p + o(1)$ as $T \to \infty$. So the modified Wald test based on the F approximation can be justified in the same manner as the conventional chi-square test. However, in finite samples, the new test can be more accurate in size.

5 Simulation Evidence

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

$$y_t = x_{0,t}\gamma_0 + x_{1,t}\gamma_1 + x_{2,t}\gamma_2 + x_{3,t}\gamma_3 + \varepsilon_{y,t},$$

where $x_{0,t} \equiv 1$ and $x_{1,t}$, $x_{2,t}$ and $x_{3,t}$ are scalar endogenous regressors. The unknown parameter vector is $\theta = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)' \in \mathbb{R}^d$ for $d = 4$. We have m instruments $z_{0,t}, z_{1,t}, \ldots, z_{m-1,t}$ with $z_{0,t} \equiv 1$. The reduced-form equations for $x_{1,t}$, $x_{2,t}$ and $x_{3,t}$ are given by

$$x_{j,t} = z_{j,t} + \sum_{i=d}^{m-1} z_{i,t} + \varepsilon_{x_j,t} \text{ for } j = 1, 2, 3.$$

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 for $i \geq 1$, 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 $\varepsilon_t = (\varepsilon_{yt}, \varepsilon_{x_1t}, \varepsilon_{x_2t}, \varepsilon_{x_3t})'$ 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 from each other. We take $\rho=-0.5, 0.0, 0.5$, 0.8 and 0.9.

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 the corresponding numbers of moment conditions are $m=4,5,6$. The null hypotheses of interest are

$$
\begin{aligned}
&H_{01}:\gamma_1=0,\\
&H_{02}:\gamma_1=\gamma_2=0,\\
&H_{03}:\gamma_1=\gamma_2=\gamma_3=0.
\end{aligned}
$$

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 type 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: $\chi_p^{1-\alpha}/p$, $\frac{K}{K-p-q+1}\mathcal{F}_{p,K-p-q+1}^{1-\alpha}\left(\delta^2\right)$ with $\delta^2=pq/(K-q-1)$, 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 F_∞ approximation with simulated critical values. The NCF test and the F_∞ test are developed in Sun (2014b). The fourth test is the 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 the unmodified Wald test statistic as the first three tests but uses the critical values given in (7). For easy reference, we now refer to our test as the standard F test, which should not be confused with the standard F test in a CNLR model. For each test, the initial first-step estimator is the IV estimator with weight matrix $W_o=Z'Z/T$ where Z is the matrix of instruments.

We use the following 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 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 1 and 2 report the finite sample size of the four tests for $T=100$ and $\alpha=5\%$. The number of simulation replications is 10000. It is clear that the standard F test has as accurate size as the nonstandard F_∞ test and noncentral F test. 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.

Figures 1 and 2 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. 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 simulation evidence lends a 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 distributed. This leads to standard F tests that are based on the modified test statistics and use the standard F 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) and Sun and Kim (2012) which establish the F 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 limit theory appears to be applicable to all common tests in the first-step and two-step GMM settings. The results of the paper can be easily extended to the continuous updating GMM (CU-GMM) framework. Recently, Zhang (2015) 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 result holds without change if the CU-GMM estimator is used instead. Following the work of Bester, Conley, Hansen and Vogelsang (2015) and Sun and Kim (2015), we also do not imagine much difficulty in extending our results to the spatial setting.

Table 1: 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

	χ^2	NCF	F_∞	F	χ^2	NCF	F_∞	F	χ^2	NCF	F_∞	F
ρ		$p=1, q=0$				$p=2, q=0$				$p=3, q=0$		
-0.8	0.114	0.072	0.073	0.072	0.197	0.087	0.084	0.087	0.310	0.109	0.111	0.109
-0.5	0.081	0.060	0.059	0.060	0.117	0.066	0.066	0.066	0.174	0.077	0.078	0.077
0.0	0.063	0.051	0.050	0.051	0.083	0.052	0.053	0.052	0.112	0.060	0.062	0.060
0.5	0.094	0.063	0.063	0.063	0.142	0.065	0.065	0.065	0.222	0.077	0.078	0.077
0.8	0.134	0.086	0.088	0.086	0.229	0.100	0.097	0.100	0.355	0.119	0.122	0.119
0.9	0.166	0.117	0.120	0.117	0.290	0.150	0.146	0.150	0.437	0.181	0.184	0.181
		$p=1, q=1$				$p=2, q=1$				$p=3, q=1$		
-0.8	0.186	0.081	0.077	0.079	0.307	0.088	0.087	0.086	0.457	0.107	0.113	0.105
-0.5	0.113	0.065	0.065	0.064	0.175	0.069	0.068	0.067	0.247	0.079	0.080	0.078
0.0	0.081	0.053	0.052	0.052	0.113	0.057	0.056	0.056	0.155	0.060	0.060	0.060
0.5	0.128	0.064	0.063	0.063	0.204	0.073	0.072	0.071	0.308	0.079	0.080	0.078
0.8	0.196	0.089	0.086	0.087	0.331	0.101	0.099	0.100	0.489	0.112	0.118	0.112
0.9	0.252	0.126	0.123	0.122	0.420	0.155	0.153	0.147	0.589	0.183	0.192	0.172
		$p=1, q=1$				$p=1, q=2$				$p=1, q=2$		
-0.8	0.260	0.080	0.079	0.077	0.425	0.090	0.083	0.084	0.602	0.100	0.097	0.090
-0.5	0.154	0.061	0.062	0.061	0.244	0.065	0.061	0.065	0.351	0.074	0.073	0.072
0.0	0.104	0.055	0.055	0.055	0.148	0.062	0.058	0.060	0.211	0.065	0.063	0.065
0.5	0.171	0.065	0.066	0.064	0.279	0.065	0.061	0.062	0.415	0.073	0.072	0.070
0.8	0.268	0.085	0.082	0.080	0.449	0.090	0.082	0.085	0.623	0.100	0.097	0.088
0.9	0.332	0.124	0.121	0.109	0.529	0.148	0.137	0.132	0.712	0.160	0.157	0.139

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 2: 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

	χ^2	NCF	F_∞	F	χ^2	NCF	F_∞	F	χ^2	NCF	F_∞	F
L		$p=1, q=0$				$p=2, q=0$				$p=3, q=0$		
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 1

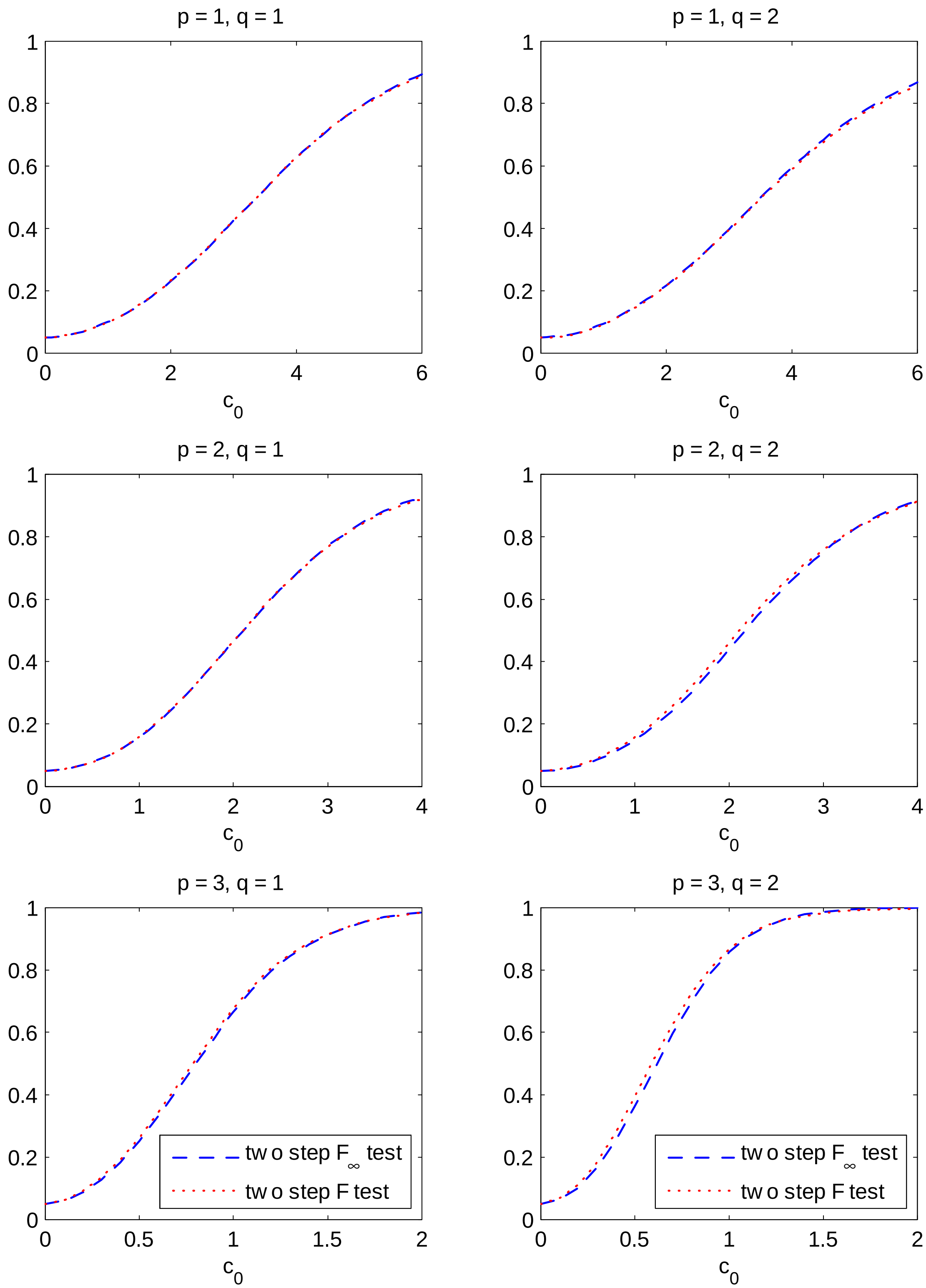

Figure 1: Size-adjusted power of two-step 5% F_∞ and F tests based on the series LRV estimator under the AR design with $\rho = 0.5$ and $T = 100$

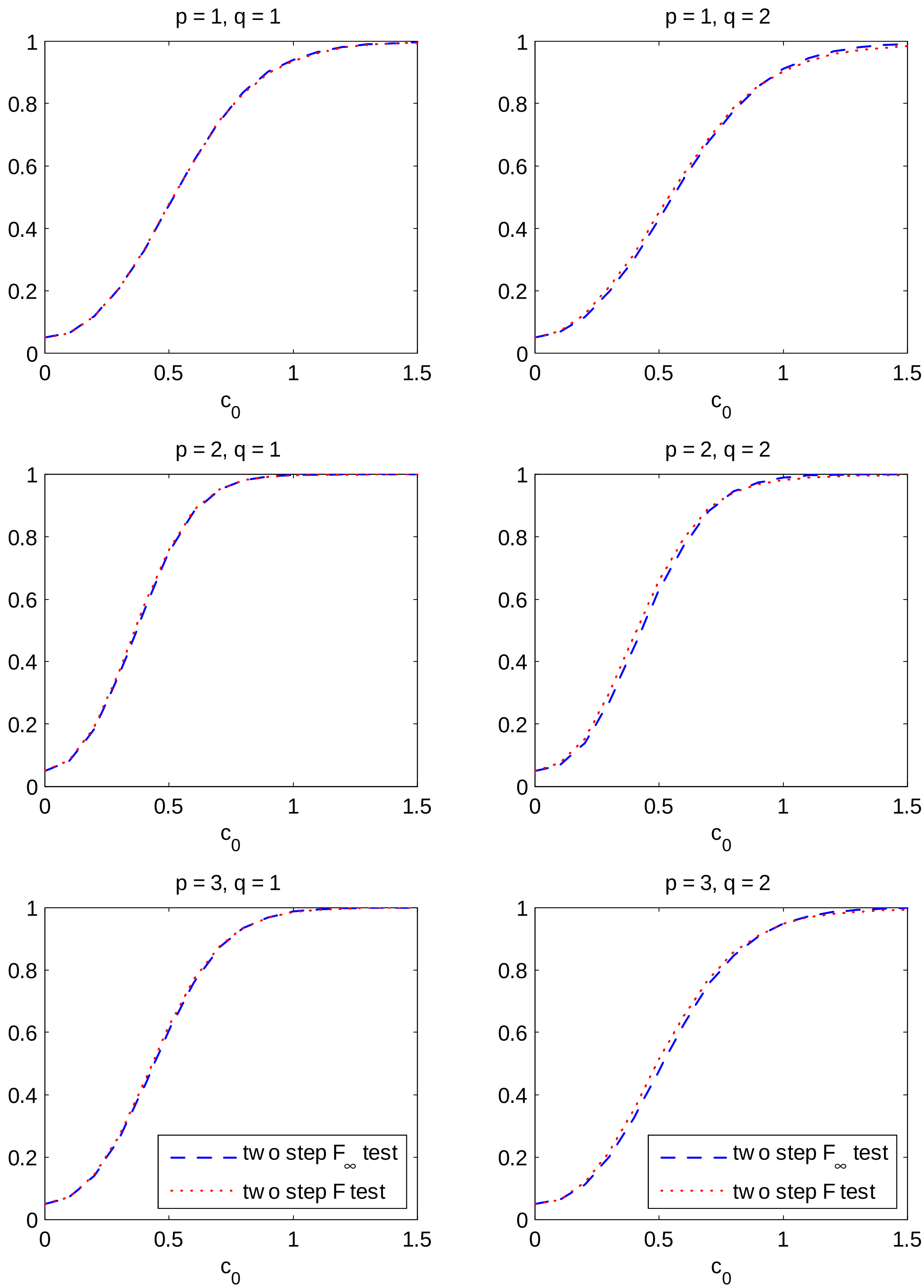

Figure 2: Size-adjusted power of two-step 5% F_∞ and F tests based on the series LRV estimator under the centered MA design with $L = 9$ and $T = 100$

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.

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

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

$$\begin{aligned}J_T(\hat{\theta}_T) &\xrightarrow{d} \left\{B_m(1) - G_\Lambda\left[G_\Lambda'\tilde{W}_\infty^{-1}G_\Lambda\right]^{-1}G_\Lambda'\tilde{W}_\infty^{-1}B_m(1)\right\}'\tilde{W}_\infty^{-1}\\ &\times\left\{B_m(1) - G_\Lambda\left[G_\Lambda'\tilde{W}_\infty^{-1}G_\Lambda\right]^{-1}G_\Lambda'\tilde{W}_\infty^{-1}B_m(1)\right\} := J_\infty.\end{aligned} \tag{13}$$

In addition, careful inspection shows that the above convergence results hold jointly. It remain to show that (F_∞, J_∞) 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:

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

and

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

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

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

where

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

and

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

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

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

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,

$$C_{11} = \int_0^1\int_0^1 Q_K(r,s)dB_d(r)dB_d(s)' = \begin{pmatrix} C_{pp} & C_{p,d-p} \\ C_{p,d-p}' & C_{d-p,d-p}\end{pmatrix} \tag{14}$$

$$C_{12} = \int_0^1\int_0^1 Q_K(r,s)dB_d(r)dB_q(s)' = \begin{pmatrix} C_{pq} \\ C_{d-p,q}\end{pmatrix} \tag{15}$$

$$C_{22} = \int_0^1\int_0^1 Q_K(r,s)dB_q(r)dB_q(s)' = C_{qq} \tag{16}$$

where C_{pp}, C_{pq}, and C_{qq} are defined in (5), 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

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

and so

$$\begin{aligned}&RV\left[\Sigma'\tilde{W}_\infty^{-1}\Sigma\right]^{-1}\Sigma'\tilde{W}_\infty^{-1} \\ &= RVA^{-1}\left(C^{11}\right)^{-1}\left(A'\right)^{-1}\begin{pmatrix} A' & O'\end{pmatrix}\begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22}\end{pmatrix} \\ &= RVA^{-1}\left(C^{11}\right)^{-1}\left(A'\right)^{-1}A'\begin{pmatrix} C^{11} & C^{12}\end{pmatrix} \\ &= RVA^{-1}\begin{pmatrix} I_d, & \left(C^{11}\right)^{-1}C^{12}\end{pmatrix},\end{aligned} \tag{17}$$

and

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

As a result,

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

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

$$\tilde{\Sigma}_{p\times d} = \left(\begin{array}{cc} \tilde{A}_{p\times p}, & \tilde{O}_{p\times(d-p)}\end{array}\right).$$

Then

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

Using the same steps, we have

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

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 $\left[C^{22} - C^{21}\left(C^{11}\right)^{-1} C^{21}\right]^{-1} = C_{22} = C_{qq}$, 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

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

where

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

and

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

Writing

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

where $D_{pp} = C_{pp} - C_{pq} C_{qq}^{-1} C_{pq}'$ and D^{22} is a $(d-p) \times (d-p)$ matrix and using equations (14)–(16), we have

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

So

$$
\begin{pmatrix} \mathbb{W}_T(\hat{\theta}_T) \\ J_T(\hat{\theta}_T) \end{pmatrix} \overset{d}{\to} \begin{pmatrix} F_{\infty} \\ J_{\infty} \end{pmatrix} \overset{d}{=} \begin{pmatrix} \tilde{F}_{\infty}^* \\ \tilde{J}_{\infty}^* \end{pmatrix}.
$$

The theorem then follows if we let $(F_{\infty}, J_{\infty})' = (F_{\infty}^*, J_{\infty}^*)'$, which is innocuous for the weak convergence result. ■

Proof of Theorem 2. Part (a). Conditional on $B_q(\cdot) := \{B_q(r) : r \in [0,1]\}$, both $B_p(1)$ and C_{pq} are normal. Hence conditional on $B_q(\cdot)$, we have

$$B_p(1) - C_{pq}C_{qq}^{-1}B_q(1) \sim N\left(0, I_p + E\left[C_{pq}C_{qq}^{-1}B_q(1)B_q(1)' C_{qq}^{-1}C_{qp}|B_q(\cdot)\right]\right). \tag{18}$$

Let $B_p^{(i)}(r)$ be the i-th element of $B_p(r)$. Define

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

which are the i-th row of C_{pq} and j-th column of C_{qp}, respectively. Then the (i,j)-th element of the conditional variance in (18) can be written as

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

where $\delta_{ij} = 1\{i=j\}$. 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 $B_q(\cdot)$,

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

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

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

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

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

In view of Theorem 1, we have

$$\frac{K-p-q+1}{K} \frac{\mathbb{W}_T(\hat{\theta}_T)}{1 + \frac{q}{K}J_T(\hat{\theta}_T)} \xrightarrow{d} F_{p,K-p-q+1},$$

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 $\mathbb{W}_T(\hat{\theta}_T)$, $\mathbb{D}_T(\hat{\theta}_T)$ and $\mathbb{S}_T(\hat{\theta}_T)$ still holds under the fixed-smoothing asymptotics. For more details, see Sun (2014b). ■

Proof of Proposition 3. If $T^{-1}\sum_{t=1}^{T}\Phi_k(t/T) = 0$ for $k = 1, 2, \ldots, K$, then $X = \sqrt{T}e_{K+1}$ where $e_{K+1} = (1, 0, ..., 0)'$ is the first unit vector in $\mathbb{R}^{K+1}$. So

$$\begin{aligned}
&\sqrt{T}\left(\hat{\theta}_{T,OLS}' - \theta_0'\right) \\
&= \left[\frac{X'}{\sqrt{T}}M_2\frac{X}{\sqrt{T}}\right]^{-1}\frac{X'}{\sqrt{T}}M_2\omega_\varepsilon = \left[e_{K+1}'M_2e_{K+1}\right]^{-1}e_{K+1}'M_2\omega_\varepsilon \\
&= \left[1 - e_{K+1}'\omega_2\left(\omega_2'\omega_2\right)^{-1}\omega_2'e_{K+1}\right]^{-1}\left[e_{K+1}'\omega_{u_1} - e_{K+1}'\omega_2\left(\omega_2'\omega_2\right)^{-1}\omega_2'\omega_{u_1}\right]
\end{aligned}$$

where ω_{u_1} is defined in the same way as ω_ε is defined. Let

$$S_{22} = \sum_{i=1}^{K}\omega_i(y_2)\,\omega_i(y_2)' \text{ and } S_{21} = \sum_{i=1}^{K}\omega_i(y_2)\,\omega_i(u_1)'.$$

Using the Sherman-Morrison formula, we have

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

and

$$
\begin{aligned}
& e_{K+1}'\omega_2\left(\omega_2'\omega_2\right)^{-1}\omega_2'\omega_{u_1} \\
&= \omega_0\left(y_2\right)'\left[\omega_0\left(y_2\right)\omega_0'\left(y_2\right)+\sum_{i=1}^{K}\omega_i\left(y_2\right)\omega_i\left(y_2\right)'\right]^{-1}\left[\omega_0\left(y_2\right)\omega_0'\left(u_1\right)+\sum_{i=1}^{K}\omega_i\left(y_2\right)\omega_i'\left(u_1\right)\right] \\
&= \sqrt{T}\bar{y}_2'\left(T\bar{y}_2\bar{y}_2'+S_{22}\right)^{-1}\left[T\bar{y}_2\bar{u}_1'+S_{21}\right] = \sqrt{T}\bar{y}_2'\left[S_{22}^{-1}-\frac{S_{22}^{-1}\left(T\bar{y}_2\bar{y}_2'\right)S_{22}^{-1}}{1+T\bar{y}_2'S_{22}^{-1}\bar{y}_2}\right]\left[T\bar{y}_2\bar{u}_1'+S_{21}\right] \\
&= \left(T\bar{y}_2'S_{22}^{-1}\bar{y}_2\right)\sqrt{T}\bar{u}_1'+\sqrt{T}\bar{y}_2'S_{22}^{-1}S_{21} \\
&\quad -\frac{\left(T\bar{y}_2'S_{22}^{-1}\bar{y}_2\right)\times\left(T\bar{y}_2'S_{22}^{-1}\bar{y}_2\right)\sqrt{T}\bar{u}_1'}{1+T\bar{y}_2'S_{22}^{-1}\bar{y}_2}-\frac{\left(T\bar{y}_2'S_{22}^{-1}\bar{y}_2\right)\times\left(\sqrt{T}\bar{y}_2'S_{22}^{-1}\right)}{1+T\bar{y}_2'S_{22}^{-1}\bar{y}_2}S_{21} \\
&= \frac{\left(T\bar{y}_2'S_{22}^{-1}\bar{y}_2\right)\sqrt{T}\bar{u}_1'+\sqrt{T}\bar{y}_2'S_{22}^{-1}S_{21}}{1+T\bar{y}_2'S_{22}^{-1}\bar{y}_2}.
\end{aligned}
$$

Hence

$$
e_{K+1}'M_2e_{K+1}=1-e_{K+1}'\omega_2\left(\omega_2'\omega_2\right)^{-1}\omega_2'e_{K+1}=1-\frac{T\bar{y}_2'S_{22}^{-1}\bar{y}_2}{1+T\bar{y}_2'S_{22}^{-1}\bar{y}_2}=\frac{1}{1+T\bar{y}_2'S_{22}^{-1}\bar{y}_2},
$$

$$
\begin{aligned}
e_{K+1}'M_2\omega_\varepsilon &= e_{K+1}'\omega_{u_1}-e_{K+1}'\omega_2\left(\omega_2'\omega_2\right)^{-1}\omega_2'\omega_{u_1} \\
&= \sqrt{T}\bar{u}_1'-\frac{\left(T\bar{y}_2'S_{22}^{-1}\bar{y}_2\right)\sqrt{T}\bar{u}_1'+\sqrt{T}\bar{y}_2'S_{22}^{-1}S_{21}}{1+T\bar{y}_2'S_{22}^{-1}\bar{y}_2}=\frac{\sqrt{T}\bar{u}_1'-\sqrt{T}\bar{y}_2'S_{22}^{-1}S_{21}}{1+T\bar{y}_2'S_{22}^{-1}\bar{y}_2}.
\end{aligned}
$$

It then follows that

$$
\sqrt{T}\left(\hat{\theta}_{T,OLS}'-\theta_0'\right)=\sqrt{T}\left(\bar{u}_1'-\bar{y}_2'S_{22}^{-1}S_{21}\right)=\sqrt{T}\left(\bar{u}_1'-\bar{u}_2'S_{22}^{-1}S_{21}\right).
$$

It is easy to see that $S_{22}^{-1}S_{21}=\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}$. So

$$
\hat{\theta}_{T,OLS}-\theta_0=\bar{u}_1-\hat{\beta}\bar{u}_2=\hat{\theta}_{T,GMM}-\theta_0.
$$

This implies that $\hat{\theta}_{T,OLS}=\hat{\theta}_{T,GMM}$, as desired.

If $\int_0^1 \Phi_k(r)\,dr = 0$ but not $T^{-1}\sum_{t=1}^T \Phi_k(t/T) = 0$, then we have $X = \sqrt{T}e_{K+1} + O\left(1/\sqrt{T}\right)$. Using this and the assumptions in the proposition, we have

$$\sqrt{T}\left(\hat{\theta}'_{T,OLS} - \theta'_0\right) = \left[e'_{K+1}M_2e_{K+1}\right]^{-1} e'_{K+1}M_2\omega_\varepsilon + o_p(1).$$

Following the same argument as above, we have $\sqrt{T}(\hat{\theta}_{T,OLS} - \theta_0) = \sqrt{T}(\hat{\theta}_{T,GMM} - \theta_0) + o_p(1)$, which implies that $\sqrt{T}(\hat{\theta}_{T,OLS} - \hat{\theta}_{T,GMM}) = o_p(1)$. ■

Proof of Proposition 4. We first give a representation of $\mathbb{W}_{CNLR}$. We focus on the case that $T^{-1}\sum_{t=1}^T \Phi_k(t/T) = 0$ for $k = 1, 2, \ldots, K$, as the other case follows from the similar arguments. Using $(\hat{\theta}_{T,OLS} - r)' = [X'M_2X]^{-1} X'M_2\omega_\varepsilon$, we have

$$\begin{aligned}
\mathbb{W}_{CNLR} &= \left(\hat{\theta}_{T,OLS} - r\right)' \left\{\tilde{\Omega}_{11\cdot 2}\left(X'M_2X\right)^{-1}\right\}^{-1} \left(\hat{\theta}_{T,OLS} - r\right)/p \\
&= \left[X'M_2X\right]^{-1} X'M_2\omega_\varepsilon \left\{\tilde{\Omega}_{11\cdot 2}\left(X'M_2X\right)^{-1}\right\}^{-1} \omega'_\varepsilon M_2X\left[X'M_2X\right]^{-1} \\
&= \frac{X'M_2\omega_\varepsilon \times \tilde{\Omega}_{11\cdot 2}^{-1} \times \omega'_\varepsilon M_2X}{X'M_2X}\frac{1}{p},
\end{aligned}$$

using the fact that $X'M_2X$ is a scalar.

In the proof of Proposition 3, we have shown that

$$\begin{aligned}
X'M_2\omega_\varepsilon &= \sqrt{T}e'_{K+1}M_2\omega_\varepsilon = \frac{T\left(\bar{u}'_1 - \bar{y}'_2S_{22}^{-1}S_{21}\right)}{1 + T\bar{y}'_2S_{22}^{-1}\bar{y}_2} \text{ and} \\
X'M_2X &= Te'_{K+1}M_2e_{K+1} = \frac{T}{1 + T\bar{y}'_2S_{22}^{-1}\bar{y}_2}.
\end{aligned}$$

Hence

$$\mathbb{W}_{CNLR} = \frac{\sqrt{T}\left(\bar{u}'_1 - \bar{y}'_2S_{22}^{-1}S_{21}\right)}{\sqrt{1 + T\bar{y}'_2S_{22}^{-1}\bar{y}_2}} \times \tilde{\Omega}_{11\cdot 2}^{-1} \times \frac{\sqrt{T}\left(\bar{u}_1 - S_{12}S_{22}^{-1}\bar{y}_2\right)}{\sqrt{1 + T\bar{y}'_2S_{22}^{-1}\bar{y}_2}}\frac{1}{p}.$$

To simplify $\tilde{\Omega}_{11\cdot 2}^{-1}$, we note that

$$\hat{\beta}'_{T,OLS} = \left[\omega'_2M_X\omega_2\right]^{-1}\omega'_2M_X\omega_1$$

where

$$M_X = I_{K+1} - X(X'X)^{-1}X' = I_{K+1} - XX' = \begin{pmatrix} 0 & 0 \\ 0 & I_K \end{pmatrix}.$$

So $\hat{\beta}'_{OLS} = S_{22}^{-1}S_{21}$. Plugging this and $\hat{\theta}_{OLS}$ into the estimated residuals yields

$$\begin{aligned}
&\omega_1 - X\hat{\theta}'_{T,OLS} - \omega_2\hat{\beta}'_{T,OLS} \\
&= \omega_\varepsilon - X\left(\hat{\theta}'_{T,OLS} - \theta'_0\right) - \omega_2\left(\hat{\beta}'_{T,OLS} - \beta'_0\right) \\
&= \omega_\varepsilon - X\left(\bar{u}'_1 - \bar{u}'_2S_{22}^{-1}S_{21}\right) - \omega_2S_{22}^{-1}S_{21} + \omega_2\beta'_0 \\
&= \begin{pmatrix} \sqrt{T}\left(\bar{\varepsilon}' - \bar{u}'_1 + \bar{u}'_2S_{22}^{-1}S_{21} - \bar{u}'_2S_{22}^{-1}S_{21} + \bar{u}_2\beta'_0\right) \\ \omega'_1(\varepsilon) - \omega'_2(u_2)S_{22}^{-1}S_{21} + \omega'_2(u_2)\beta'_0 \\ \cdots \\ \omega'_K(\varepsilon) - \omega'_K(u_2)S_{22}^{-1}S_{21} + \omega'_K(u_2)\beta'_0 \end{pmatrix} = \begin{pmatrix} 0 \\ \omega'_1(u_1) - \omega'_2(u_2)S_{22}^{-1}S_{21} \\ \cdots \\ \omega'_K(u_1) - \omega'_2(u_2)S_{22}^{-1}S_{21} \end{pmatrix}.
\end{aligned}$$

Therefore

$$\tilde{\Omega}_{11\cdot 2} = \frac{1}{K-q}\sum_{i=1}^{K}\left[\omega_i\left(u_1\right) - S_{12}S_{22}^{-1}\omega_i\left(u_2\right)\right]\left[\omega_i\left(u_1\right) - S_{12}S_{22}^{-1}\omega_i\left(u_2\right)\right]'$$
$$= \frac{1}{K-q}\left(S_{11} - S_{12}S_{22}^{-1}S_{21}\right).$$

Using this and noting that $S_{ij} = K\hat{\Omega}_{ij}$, we have

$$\tilde{\Omega}_{11\cdot 2} = \frac{K}{K-q}\left(\hat{\Omega}_{11} - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}\right),$$

and so

$$\mathbb{W}_{CNLR} = \frac{K-q}{K}\frac{\sqrt{T}\left(\bar{u}_1 - S_{12}S_{22}^{-1}\bar{y}_2\right)'}{\sqrt{1+T\bar{y}_2'S_{22}^{-1}\bar{y}_2}} \times \left(\hat{\Omega}_{11} - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}\right)^{-1} \times \frac{\sqrt{T}\left(\bar{u}_1 - S_{12}S_{22}^{-1}\bar{y}_2\right)}{\sqrt{1+T\bar{y}_2'S_{22}^{-1}\bar{y}_{22}}}\frac{1}{p}$$
$$= \frac{K-q}{K}\frac{\sqrt{T}\left(\bar{u}_1 - \hat{\beta}\bar{u}_2\right)'\left(\hat{\Omega}_{11} - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}\right)^{-1}\sqrt{T}\left(\bar{u}_1 - \hat{\beta}\bar{u}_2\right)}{1+\frac{1}{K}(\sqrt{T}\bar{u}_2)'\hat{\Omega}_{22}^{-1}(\sqrt{T}\bar{u}_2)}\frac{1}{p},$$

where we have used $S_{12}S_{22}^{-1} = \hat{\beta}_{T,OLS} = \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} = \hat{\beta}$.

Next, we give a representation of $\mathbb{W}_T^c(\hat{\theta}_T)$ when $R = I_p$. For the location model, $G_T(\hat{\theta}_T)' = (I_p, O_{p\times q})$. We have

$$\mathbb{W}_T := \sqrt{T}(\hat{\theta}_T - r)'\left(\hat{\Omega}_{11} - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}\right)^{-1}\sqrt{T}(\hat{\theta}_T - r)/p.$$

Combining this with

$$J_T = (\sqrt{T}\bar{u}_2)'\hat{\Omega}_{22}^{-1}(\sqrt{T}\bar{u}_2),$$

we have

$$\mathbb{W}_T^c(\hat{\theta}_T) = \frac{K-p-q+1}{K}\frac{\sqrt{T}\left(\bar{u}_1 - \hat{\beta}\bar{u}_2\right)'\left(\hat{\Omega}_{11} - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}\right)^{-1}\sqrt{T}\left(\bar{u}_1 - \hat{\beta}\bar{u}_2\right)}{1+\frac{1}{K}(\sqrt{T}\bar{u}_2)'\hat{\Omega}_{22}^{-1}(\sqrt{T}\bar{u}_2)}\frac{1}{p}.$$

So

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

In particular, $\mathbb{W}_{CNLR} = \mathbb{W}_T^c(\hat{\theta}_T)$ when $p = 1$. ■

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 (2015): "Fixed-b Asymptotics for Spatially Dependent Robust Nonparametric Covariance Matrix Estimators." *Econometric Theory,* forthcoming.

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

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

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

[6] 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.

[7] Jansson, M. (2004): "On the Error of Rejection Probability in Simple Autocorrelation Robust Tests." *Econometrica* 72, 937–946.

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

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

[10] Kiefer, N. M. and T. J. Vogelsang(2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests." *Econometric Theory* 21, 1130–1164.

[11] Müller, U. K. (2007): "A Theory of Robust Long-Run Variance Estimation." *Journal of Econometrics* 141, 1331–1352.

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

[13] Newey, W. K. and K. D. West (1987): "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55, 703–708.

[14] Phillips, P. C. B. (2005): "HAC Estimation by Automated Regression." *Econometric Theory* 21, 116–142.

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

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

[17] 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.

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

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

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

[21] 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.

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

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

[24] 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.

[25] Zhang, X. (2015): "Fixed-smoothing Asymptotics in the Generalized Empirical Likelihood Estimation Framework." Department of Statistics, University of Missouri–Columbia.