

Simple and Powerful GMM Over-identification Tests with Accurate Size

Yixiao Sun and Min Seong Kim*

Department of Economics,
University of California, San Diego

This version: August 1, 2010

Abstract

The paper provides a new class of over-identification tests that are robust to heteroscedasticity and autocorrelation of unknown forms. The tests are based on the series long run variance estimator that is designed to pivotalize the moment restrictions. We show that when the number of terms used in the series long run variance estimator is fixed, the conventional J statistic, after a simple correction, is asymptotically F -distributed. We apply the idea of the F -approximation to the conventional kernel-based J tests. Simulations show that the J^* tests based on the finite sample corrected J statistic and the F approximation have virtually no size distortion, and yet are as powerful as the standard J tests.

JEL Classification: C12, C32

Keywords: F-distribution, Heteroscedasticity and Autocorrelation Robust, Long-run variance, Over-identification test, Robust standard error, Series Estimator

1 Introduction

In linear and nonlinear models with moment restrictions, it is standard practice to employ the generalized method of moments (GMM) to estimate the model parameters. The method, introduced by Hansen (1982), has become a leading estimation technique in empirical research. To test for the validity of the moment restrictions, we often use the popular J test associated with Sargan (1958) and Hansen (1982). In the time series setting, the J test can be made robust to heteroscedasticity and autocorrelation of unknown forms by using an appropriate long run variance (LRV) estimator. While the heteroscedasticity and

*Email: yisun@ucsd.edu and msk003@ucsd.edu. Sun gratefully acknowledges partial research support from NSF under Grant No. SES-0752443. Correspondence to: Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508.

autocorrelation robust (HAR) J test has widespread use, Altonji and Segal (1996), and Hall and Horowitz (1996), Lee and Kuan (2009), among others, have documented that it frequently over-rejects in finite samples.

In this paper, we introduce a new HAR J test called J^* test that is easy to use and accurate in size. The J^* test is based on a nonparametric series type LRV estimator considered in Phillips (2005) and Sun (2010a, 2010b). The basic idea behind the series LRV estimator is to project the time series onto a set of basis functions designed to represent the long-run behavior directly. The outer-product of the projection coefficients is a direct and asymptotically unbiased estimator of the LRV. The series LRV estimator is simply an average of the direct estimators. The basis functions are required to be orthonormal and integrate to zero on $[0, 1]$. The latter ‘zero mean’ condition ensures that the series LRV estimator is asymptotically invariant to the model parameters of interest. As a result, it does not suffer from the bias due to the estimation uncertainty of model parameters. This is in contrast with the conventional kernel LRV estimator where this type of bias is often present. See for example, Hannan (1957). Furthermore, the zero mean condition automatically centers the moment conditions under the local alternative hypothesis. As a result, the new J^* test has the same power advantage as the test proposed by Hall (2000).

The smoothing parameter in the series LRV estimator is K , the number of basis functions employed. When the number of basis functions K is fixed, the series LRV estimator is inconsistent and converges in distribution to a scaled Wishart distribution. It is now well known that in cases like robust hypothesis testing, consistent LRV estimates are not needed in order to produce asymptotically valid tests, see for example, Kiefer, Vogelsang and Bunzel (2000), Kiefer and Vogelsang (2002a, 2002b, 2005). Indeed, under the fixed- K asymptotics where K is held fixed, we show that the modified J statistic converges to an F distribution. This is the main innovation of the paper. The modification involves multiplying the conventional J -statistic J_T by $(K - q + 1)/K$, where q is the number of over-identifying moment conditions. More specifically, $J_T^* = (K - q + 1) J_T / K$ converges in distribution to $F(q, K - q + 1)$, the F distribution with degrees of freedom q and $K - q + 1$. For convenience, we refer to the test based on the J_T^* statistic and using critical values from the F distribution as the J^* test.

The multiplicative factor in J_T^* can be regarded as a finite sample correction under the large K asymptotics where K grows with the sample size at certain rate. Under the large K asymptotics, the series LRV estimator becomes consistent and the associated J_T^* converges to the standard χ_q^2/q distribution. This result can be also seen from the sequential asymptotics where K is held fixed as $T \rightarrow \infty$ followed by letting $K \rightarrow \infty$. When $K \rightarrow \infty$, both the multiplicative correction factor and the random denominator of the F distribution converge to one. In this case, the J^* test reduces to the conventional J test. However, when K is not large and $q \neq 1$, the finite sample correction factor can be much smaller than one and the random denominator in the F approximation may be far from degenerate. The difference between the J^* test and the J test can be substantially large.

The fixed- K asymptotics captures the randomness of the LRV estimator and the resulting J^* test is expected to have better finite sample size properties than the conventional J test. See Jansson (2004) and Sun, Phillips and Jin (2008) for related theoretical analysis for testing model parameters in location models. In a more general setting, Sun (2010a) shows that critical values from the fixed- K limiting distribution are higher order correct under the large K asymptotics.

We apply the idea behind the F limiting theory to the popular kernel-based J tests. The underlying LRV estimator is based on the nonparametric kernel method with the bandwidth parameter b equal to the ratio of the truncation lag (bandwidth) to the sample size. When b is fixed, the kernel LRV estimator converges to a double stochastic integral and the J_T statistic has a nonstandard limiting distribution. The key idea is to observe that the stochastic integral is equal to an infinite weighted sum of independent Wishart distributions. We can approximate the infinite weighted sum by a finite and unweighted sum, that is, we can approximate the stochastic integral by a Wishart distribution with an “equivalent degree of freedom.” With this approximation, the fixed- b limiting distribution of a modified J statistic becomes approximately F -distributed. As in the case with the series LRV estimator, the kernel J^* test is as easy to use as the standard J test, as the correction factor is easy to compute and the critical values are readily available from statistical tables and software packages.

A Monte Carlo study shows that both the series J^* tests and the kernel J^* tests have much smaller size distortion than the conventional J test. Among the Bartlett, Parzen, QS and Daniell kernels considered, the Parzen J^* test is most accurate in size. The series J^* test is as accurate in size as the Parzen J^* test. In terms of size-adjusted power, the series J^* test is as competitive as the conventional J tests. In fact, it is more powerful than the J tests based on the Parzen, QS and Daniell kernels in some scenarios. In view of its remarkable accuracy in size and competitiveness in power, we recommend the series J^* test for practical use. At the minimum, the kernel J^* tests should be used in place of the conventional J tests that often over reject.

The paper that is most closely related to the present one is Lee and Kuan (2009). Like this paper, they consider testing over-identifying restrictions in the framework of time series GMM and establish the nonstandard fixed- b asymptotics for the kernel-based J tests. Our main focus here is on the series J^* test, which is accurate in size and yet easy to use in practice. There is no need to simulate nonstandard critical values. As an extension, we also consider the kernel-based J^* tests that use the conventional J statistic. To obtain their fixed- b asymptotics, Lee and Kuan (2009) propose to modify the conventional J statistic in order to avoid a singularity problem. We show that the modification is unnecessary. In addition, the new F approximation to the sampling distribution of the kernel J^* statistic is very convenient in econometric applications.

The remainder of the paper is organized as follows. Section 2 describes the testing problem of concern and introduces the series LRV estimator. Section 3 studies the limiting F -distribution theory under the fixed- K asymptotics. Section 4 investigates the local asymptotic power of the series J^* test. The next section applies the idea of F -approximation to the kernel-based J^* tests. Section 6 presents simulation evidence and the last section concludes. Proofs are given in the Appendix.

2 GMM and Over-identification Test

In this section, we present the GMM estimation and the over-identification test in the time series setting. We also provide an overview of the series LRV estimator.

We are interested in $d \times 1$ vector of parameters $\theta \in \Theta \subseteq \mathbb{R}^d$. Let v_t denote a vector of observations. Let θ_0 denote the true value and assume that θ_0 is an interior point of Θ .

The moment conditions

$$Ef(v_t, \theta) = 0, \quad t = 1, 2, \dots, T \quad (1)$$

hold if and only if $\theta = \theta_0$ where $f(\cdot)$ is an $m \times 1$ vector of continuously differentiable functions. As a time series, $f(v_t, \theta_0)$ may exhibit autocorrelation of unknown forms. We assume that $m > d$ and $\text{rank } E[\partial f(v_t, \theta_0) / \partial \theta'] = d$. That is, we consider an over-identified model with the degree of over-identification $q = m - d$.

Define

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

then the GMM estimator of θ_0 is given by

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

where W_T is a weighting matrix.

To obtain an initial first step estimator, we often choose a simple weighting matrix W_o that does not depend on model parameters, leading to

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

As an example, we may set $W_o = I_m$ in the general GMM setting. In the IV regression, we may set $W_o = Z'Z/T$ where Z is the data matrix for the instruments. We assume that

$$W_o \xrightarrow{p} W_{o,\infty},$$

a positive definite matrix.

On the basis of $\tilde{\theta}_T$, we can use $\tilde{u}_t = f(v_t, \tilde{\theta}_T)$ to construct an optimal weighting matrix. According to Hansen (1982), the optimal weighting matrix W_T is the long run variance of the time series $f(v_t, \theta_0)$. Many nonparametric estimators of the LRV are available in the literature. For kernel estimators, see Andrews (1991) and Newey and West (1987, 1994).

In this paper, we consider a class of series LRV estimators considered by Phillips (2005) and Sun (2010a, 2010b). As we show later, the series estimators are especially convenient for constructing the J test. Let $\{\Phi_k(\cdot), k = 1, 2, \dots, K\}$ be a sequence of orthonormal basis functions on $L^2[0, 1]$. Define the projection coefficient

$$\Lambda_k(\theta) = \frac{1}{\sqrt{T}} \sum_{t=1}^T \Phi_k\left(\frac{t}{T}\right) f(v_t, \theta) \quad \text{for } k = 1, 2, \dots, K.$$

Then the series LRV estimator is the average of the outer product of these projection coefficients evaluated at $\theta = \tilde{\theta}_T$:

$$W_T(\tilde{\theta}_T) = \frac{1}{K} \sum_{k=1}^K \Lambda_k(\tilde{\theta}_T) \Lambda_k'(\tilde{\theta}_T). \quad (2)$$

In essence, each outer product $\Lambda_k(\tilde{\theta}_T) \Lambda_k'(\tilde{\theta}_T)$ is a direct estimator of the LRV and the series LRV estimator is a simple average of these direct estimators. Here K is the smoothing parameter underlying $W_T(\tilde{\theta}_T)$. To ensure that $W_T(\tilde{\theta}_T)$ is positive semidefinite, it is

necessary to assume that $K \geq m$. We maintain this assumption throughout the rest of the paper. The series LRV estimator belongs to the class of multiple-window estimators (e.g. Percival and Walden, 1993, ch. 7) and filter-bank estimators (e.g. Stoica and Moses, 2005, ch. 5). See Sun (2010a, 2010b) for more discussions.

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

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

The standard method of testing over-identifying restrictions is to take the second step estimator $\hat{\theta}_T$ of the parameter θ , and construct a test statistic J_T :

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

The above definition is slightly different from the standard one in that the GMM objective function is normalized by the degree of over-identification. The normalization does not have any impact on the properties of the test as long as critical values are appropriately adjusted.

In the following we also consider

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

That is, the test statistic with the updated weighting matrix.

3 Asymptotic Distribution of the Series J -statistic

In this section, we derive the asymptotic distribution of J_T under the specification that K is fixed. We show that the asymptotic distribution is invariant to the initial first step estimator used.

Denote

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

To analyze the asymptotic properties of $\tilde{\theta}_T$ and $\hat{\theta}_T$, we make the following high-level assumptions, which are the same as those in Kiefer and Vogelsang (2005) and Lee and Kuan (2009).

Assumption 1 $\text{plim}_{T \rightarrow \infty} \hat{\theta}_T = \theta_0$, $\text{plim}_{T \rightarrow \infty} \tilde{\theta}_T = \theta_0$.

Assumption 2 $\text{plim}_{T \rightarrow \infty} G_{[rT]}(\theta) = rG(\theta)$ uniformly in r and $\theta \in \Theta$ where $G(\theta) = E \partial f(v_j, \theta) / \partial \theta'$ and $G_0 = G(\theta_0)$ has rank d .

Assumption 3 $T^{1/2} g_{[rT]}(\theta_0) \xrightarrow{d} \Lambda B_m(r)$ where $\Lambda \Lambda' = \Omega = \sum_{j=-\infty}^{\infty} \Gamma_j$, $\Gamma_j = E f(v_t, \theta_0) f(v_{t-j}, \theta_0)'$ and $B_m(r)$ is a standard Brownian motion.

Under the above assumptions, $\hat{\theta}_T$ satisfies:

$$\hat{\theta}_T - \theta_0 = - \left[G_T(\tilde{\theta}_T)' W_T^{-1}(\tilde{\theta}_T) G_T(\tilde{\theta}_T) \right]^{-1} G_T(\theta_0)' W_T^{-1}(\tilde{\theta}_T) g_T(\theta_0).$$

where $\tilde{\theta}_T$ is between $\hat{\theta}_T$ and θ_0 . Similarly

$$\tilde{\theta}_T - \theta_0 = - \left[G_T (\bar{\theta}_T)' W_o^{-1} G_T (\bar{\theta}_T) \right]^{-1} G_T (\theta_0)' W_o^{-1} g_T (\theta_0).$$

and $\bar{\theta}_T$ is between $\tilde{\theta}_T$ and θ_0 .

To derive the limiting distribution of J_T when K is fixed, we first establish the limiting distribution of $W_T(\tilde{\theta}_T)$. Under Assumptions 1-3, we have

$$\sqrt{T} (\tilde{\theta}_T - \theta_0) \xrightarrow{d} - (G_0' W_{o,\infty}^{-1} G_0)^{-1} G_0' W_{o,\infty}^{-1} \Lambda B_m(1),$$

and for $\bar{\theta}_T^*$ between $\tilde{\theta}_T$ and θ ,

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \tilde{u}_t &= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t + \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \frac{\partial f(v_t, \bar{\theta}_T^*)}{\partial \theta'} (\tilde{\theta}_T - \theta_0) \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t - (G_{[Tr]} (\bar{\theta}_T^*)) \left[G_T (\bar{\theta}_T)' W_o^{-1} G_T (\bar{\theta}_T) \right]^{-1} G_T (\theta_0)' W_o^{-1} \sqrt{T} g_T (\theta_0) \\ &\xrightarrow{d} \Lambda B_m(r) - r G_0 [G_0' W_{o,\infty}^{-1} G_0]^{-1} G_0' W_{o,\infty}^{-1} \Lambda B_m(1). \end{aligned}$$

Hence by summation and integration by parts and the continuous mapping theorem, we have

$$\begin{aligned} &\frac{1}{\sqrt{T}} \sum_{t=1}^T \Phi_k \left(\frac{t}{T} \right) \tilde{u}_t \\ &\xrightarrow{d} \Lambda \int_0^1 \Phi_k(r) dB_m(r) - \left(\int_0^1 \Phi_k(r) dr \right) G_0 [G_0' W_{o,\infty}^{-1} G_0]^{-1} G_0' W_{o,\infty}^{-1} \Lambda B_m(1). \end{aligned}$$

Here the second term reflects the estimation uncertainty of $\tilde{\theta}_T$. To remove this term, we require $\int_0^1 \Phi_k(r) dr = 0$. In this case

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \Phi_k \left(\frac{t}{T} \right) \tilde{u}_t \xrightarrow{d} \Lambda \int_0^1 \Phi_k(r) dB_m(r)$$

and $W_T(\tilde{\theta}_T) \xrightarrow{d} W_\infty$ where

$$W_\infty = \Lambda \tilde{W}_\infty \Lambda' \text{ and } \tilde{W}_\infty = K^{-1} \sum_{k=1}^K \left[\int_0^1 \Phi_k(r) dB_m(r) \right] \left[\int_0^1 \Phi_k(r) dB_m(r) \right]'$$

Using the same argument for proving proposition 7.3 in Bilodeau and Brenner (1999), we can show that, when $K \geq m$, \tilde{W}_∞ is positive definite with probability one. As a consequence, the continuous mapping theorem can be applied to obtain $W_T^{-1}(\tilde{\theta}_T) \xrightarrow{d} W_\infty^{-1}$.

The zero mean assumption $\int_0^1 \Phi_k(r) dr = 0$ ensures that the estimation uncertainty in $\tilde{\theta}_T$ will not affect the asymptotic distribution of $W_T(\tilde{\theta}_T)$. This is an important point as the conventional kernel estimators often suffer from the bias due to the estimation error in $\tilde{\theta}_T$. See for example, Hannan (1957). In fact, we have $W_T(\bar{\theta}_T^*) \xrightarrow{d} W_\infty$ for any \sqrt{T} consistent

estimator $\hat{\theta}_T^*$. For example, we have $W_T(\hat{\theta}_T) \xrightarrow{d} W_\infty$. A direct implication is that J_T and \hat{J}_T have the same limiting distribution. More generally, the asymptotic distribution of J_T is invariant to the initial first step estimator $\check{\theta}_T$, provided that it is \sqrt{T} -consistent.

Next, we derive the asymptotic distribution of $\sqrt{T}g_T(\hat{\theta}_T) = T^{-1/2} \sum_{t=1}^T \hat{u}_t$, where $\hat{u}_t = f(v_t, \hat{\theta}_T)$. Under assumptions 1-3 and using the result $W_T^{-1}(\check{\theta}_T) \xrightarrow{d} W_\infty^{-1}$, we have

$$\sqrt{T} \left(\hat{\theta}_T - \theta_0 \right) \xrightarrow{d} - [G'_0 W_\infty^{-1} G_0]^{-1} G'_0 W_\infty^{-1} \Lambda B_m(1).$$

As a result,

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^T \hat{u}_t &= \frac{1}{\sqrt{T}} \sum_{t=1}^T u_t + \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial f(v_t, \check{\theta}_T)}{\partial \theta'} \left(\hat{\theta}_T - \theta_0 \right) \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T u_t - (G_T(\check{\theta}_T)) \left[G_T(\check{\theta}_T)' W_T^{-1} G_T(\check{\theta}_T) \right]^{-1} G_T(\theta_0)' W_T^{-1} \sqrt{T} g_T(\theta_0) \\ &\xrightarrow{d} \Lambda B_m(1) - G_0 [G'_0 W_\infty^{-1} G_0]^{-1} G'_0 W_\infty^{-1} \Lambda B_m(1). \end{aligned}$$

Now

$$J_T = \left(\frac{1}{\sqrt{T}} \sum_{s=1}^T \hat{u}_s \right)' \left(W_T(\check{\theta}_T) \right)^{-1} \left(\frac{1}{\sqrt{T}} \sum_{s=1}^T \hat{u}_s \right) / q \xrightarrow{d} J_\infty$$

where

$$\begin{aligned} J_\infty &= \left[B_m(1) - G_\Lambda \left[G'_\Lambda \tilde{W}_\infty^{-1} G_\Lambda \right]^{-1} G'_\Lambda \tilde{W}_\infty^{-1} B_m(1) \right]' \tilde{W}_\infty^{-1} \\ &\quad \times \left[B_m(1) - G_\Lambda \left[G'_\Lambda \tilde{W}_\infty^{-1} G_\Lambda \right]^{-1} G'_\Lambda \tilde{W}_\infty^{-1} B_m(1) \right] / q \end{aligned}$$

and

$$G_\Lambda = \Lambda^{-1} G_0.$$

Define the corrected J_T^* statistic:

$$J_T^* = \frac{(K - q + 1)}{K} J_T. \quad (3)$$

The following theorem gives the limiting distribution of J_T^* under the null of correct moment specifications.

Theorem 1 *Let Assumptions 1-3 hold. Further assume*

(i) $\int_0^1 \Phi_k(r) dr = 0$ for all $k = 1, 2, \dots, K$

(ii) $\int_0^1 \Phi_k(r) \Phi_j(r) dr = \delta_{kj}$

(iii) $\Phi_k(r)$ is continuously differentiable.

Then for a fixed $K \geq m$,

$$J_T^* \xrightarrow{d} F(q, K - q + 1). \quad (4)$$

There are several possible choices for $\Phi_k(\cdot)$. First, we can start with zero mean polynomials such as $r - 1/2, r^2 - 1/3, r^3 - 1/4, r^4 - 1/5$ and use the Gram-Schmidt procedure to orthonormalize them. Second, we can let $\Phi_k(r) = \sqrt{2} \cos(\pi kr)$. Following Phillips (2005), one may consider using $\Phi_k(r) = \sqrt{2} \sin(\pi kr)$ or $\sqrt{2} \sin(\pi(k - 0.5)r)$. However, these functions do not satisfy the zero mean condition. We have to rule them out as we do not observe the moment process but have to estimate it. Finally, we can let

$$\Phi_k(r) = \begin{cases} \sqrt{2} \cos(\pi kr), & k \text{ is even} \\ \sqrt{2} \sin(\pi(k + 1)r), & k \text{ is odd} \end{cases}$$

Assuming that K is even, we can write the resulting LRV estimator as

$$W_T(\hat{\theta}_T) = \frac{1}{0.5K} \sum_{k=1}^{0.5K} \text{Re}(\mathcal{A}_k \mathcal{A}_k^*).$$

where $*$ denotes transpose and complex conjugate and

$$\mathcal{A}_k = \frac{1}{\sqrt{T}} \sum_{t=1}^T \exp\left(-\frac{i2\pi kt}{T}\right) f(v_t, \hat{\theta}_T).$$

is the Finite Fourier transform (FFT) of the stochastic process $f(v_t, \hat{\theta}_T)$. We will use this estimator in our simulation study.

The zero mean assumption ensures that $\int_0^1 \Phi_k(r) dB_m(r)$ is independent of $B_m(1)$. The independence allows us to represent J_∞ as

$$J_\infty \stackrel{d}{=} B_q(1)' [\mathbb{W}(I_q, K) / K]^{-1} B_q(1) / q \quad (5)$$

where $\mathbb{W}(I_q, K)$ is a Wishart random variable that is independent of $B_q(1)$. It is well known that the right hand side of (5) is Hotelling's (1931) T^2 distribution. The theorem then follows from the relationship between the T^2 distribution and the F distribution. See also Proposition 8.2 in Bilodeau and Brenner (1999).

It follows from Theorem 1 that

$$J_T \stackrel{d}{\rightarrow} \frac{K}{K - q + 1} \frac{\chi_q^2 / q}{\chi_{K-q+1}^2 / (K - q + 1)}$$

where χ_q^2 and χ_{K-q+1}^2 are independent χ^2 random variables. Compared to the standard χ_q^2 / q approximation, the fixed- K asymptotic distribution contains two extra factors: the multiplicative factor $K / (K - q + 1)$ and the random denominator $\chi_{K-q+1}^2 / (K - q + 1)$. Since both shift the probability mass to the right, critical values from the F -approximation are larger than those from the conventional χ_q^2 / q approximation. When K is not very large relative to q , the degree of over-identification, the difference between the two sets of critical values can be considerably large.

When $q = 1$, $J_T \stackrel{d}{\rightarrow} \chi_1^2 / (\chi_K^2 / K)$. In this case, the multiplicative correction factor equals 1 and becomes irrelevant. The adjustment comes from only the random denominator. The multiplicative correction factor becomes more important when there are many moment restrictions, a scenario that often appears in econometric applications.

It is obvious that when $K \rightarrow \infty$, J_∞ becomes the standard χ_q^2/q distribution. That is, under the sequential large- K limit theory where K is held fixed as $T \rightarrow \infty$ followed by letting $K \rightarrow \infty$, J_T converges in distribution to χ_q^2/q . This is exactly the same as the conventional limiting distribution. In fact, we can use the result in Sun (2010c) and show that when $K \rightarrow \infty$ and $T \rightarrow \infty$ jointly such that $K/T \rightarrow 0$, $W_T(\hat{\theta}_T)$ is consistent for Ω under some regularity conditions. Under this joint large- K asymptotics, J_T also converges to the standard distribution χ_q^2/q . In other words, the sequential limiting distribution coincides with the joint limiting distribution.

4 Local Asymptotic Power of the Series J* Test

In the section, we study the test performance under a sequence of alternatives representing local departures from (1):

$$Ef(v_t, \theta_0) = \delta_0/\sqrt{T}.$$

This configuration is also known as the Pitman drift. Let $G_\Lambda = U\Sigma V'$ be the singular value decomposition (svd) of $G_\Lambda := \Lambda^{-1}G_0$, where

$$\Sigma = \begin{pmatrix} A \\ O \end{pmatrix},$$

A is a $d \times d$ diagonal matrix, O is a $q \times d$ matrix of zeros. By definition, $U\Sigma\Sigma'U'$ is the spectral decomposition of $G_\Lambda G'_\Lambda$. We parametrize δ_0 by

$$\delta_0 = \Lambda U \delta,$$

for some $\delta = (\delta'_1, \delta'_2) \in \mathbb{R}^m$ where $\delta_1 \in \mathbb{R}^d$ and $\delta_2 \in \mathbb{R}^q$. In other words, $\delta = U'\Lambda^{-1}\delta_0$ is a scaled version of δ_0 followed by a rotation.

Under the local alternatives, Assumptions 1 and 2 can be still valid. Assumption 3 has to be replaced by the following assumption.

Assumption 4 $T^{1/2}g_{[rT]}(\theta_0) \xrightarrow{d} r\delta_0 + \Lambda B_m(r)$ where $\Lambda\Lambda' = \Omega = \sum_{j=-\infty}^{\infty} \Gamma_j$, $\Gamma_j = Ef(v_t, \theta_0)f(v_{t-j}, \theta_0)'$ and $B_m(r)$ is a standard Brownian motion.

Under Assumptions 1, 2 and 4, we have, under the local alternatives:

$$\sqrt{T}(\tilde{\theta}_T - \theta_0) \xrightarrow{d} -[G'_0 W_{o,\infty}^{-1} G_0]^{-1} G'_0 W_{o,\infty}^{-1} [\delta_0 + \Lambda B_m(1)]$$

and

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \tilde{u}_t \xrightarrow{d} [r\delta_0 + \Lambda B_m(r)] - rG_0 [G'_0 W_{o,\infty}^{-1} G_0]^{-1} G'_0 W_{o,\infty}^{-1} [\delta_0 + \Lambda B_m(1)].$$

As a result,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \Phi_k \left(\frac{t}{T} \right) \tilde{u}_t \xrightarrow{d} \Lambda \int_0^1 \Phi_k(r) dB_m(r)$$

and $W_T(\tilde{\theta}_T) \xrightarrow{d} W_\infty$. Similarly $W_T(\hat{\theta}_T) \xrightarrow{d} W_\infty$. That is, $W_T(\tilde{\theta}_T)$ and $W_T(\hat{\theta}_T)$ have the same fixed- K limiting distributions as before.

Some comments are in order. Due to the zero mean condition, the limiting distribution of the series LRV estimator is invariant to the initial first step estimator under both the null and local alternative hypotheses. Imposing the zero mean condition is analogous to employing “centering” for kernel LRV estimators as suggested in Hall (2000) and used in Lee and Kuan (2009).

It is easy to see that under the local alternatives,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \hat{u}_t \xrightarrow{d} [\Lambda B_m(1) + \delta_0] - G_0 [G_0' W_\infty^{-1} G_0]^{-1} G_0' W_\infty^{-1} [\Lambda B_m(1) + \delta_0].$$

Therefore, under the local alternatives, we have

$$J_T \xrightarrow{d} J_\infty(\delta_0)$$

where

$$\begin{aligned} J_\infty(\delta_0) &= \left\{ B_m(1) + U\delta - G_\Lambda \left[G_\Lambda' \tilde{W}_\infty^{-1} G_\Lambda \right]^{-1} G_\Lambda' \tilde{W}_\infty^{-1} [B_m(1) + U\delta] \right\}' \tilde{W}_\infty^{-1} \\ &\quad \times \left\{ B_m(1) + U\delta - G_\Lambda \left[G_\Lambda' \tilde{W}_\infty^{-1} G_\Lambda \right]^{-1} G_\Lambda' \tilde{W}_\infty^{-1} [B_m(1) + U\delta] \right\} / q. \end{aligned}$$

The following theorem simplifies the limiting distribution of J_T under the local alternatives.

Theorem 2 *Let Assumptions 1, 2 and 4 hold. Further assume that the basis functions satisfy the assumptions in Theorem 1. Then under the local alternatives*

$$J_T^* \xrightarrow{d} F(q, K - q + 1, \|\delta_2\|^2)$$

where $F(q, K - q + 1, \|\delta_2\|^2)$ is the noncentral F distribution with degrees of freedom $(q, K - q + 1)$ and noncentrality parameter $\|\delta_2\|^2$.

The noncentrality parameter in Theorem 2 is the same as that in the conventional noncentral χ^2 distribution under local mis-specifications. In the notation of this paper, the noncentrality parameter in Hall (2005, Theorem 5.4) is given by the squared length of μ_O where

$$\mu_O = \left[I_m - G_\Lambda (G_\Lambda' G_\Lambda)^{-1} G_\Lambda' \right] \Lambda^{-1} \delta_0.$$

Note that

$$\begin{aligned} \mu_O &= \left[I_m - U \Sigma V' [V \Sigma' U' U \Sigma V']^{-1} V \Sigma' U' \right] \Lambda^{-1} \Lambda U \delta \\ &= U \left(I_m - \Sigma V' [V \Sigma' \Sigma V']^{-1} V \Sigma' \right) \delta \\ &= U \left(I_m - \Sigma (A' A)^{-1} \Sigma' \right) \delta = U \left[(\mathbf{0}', \delta_2')' \right], \end{aligned}$$

so $\|\mu_O\|^2 = \|\delta_2\|^2$. That is, we obtain the same noncentrality parameter but the route to it is different.

If $\delta_0 \in \mathcal{R}(G_0)$ such that $\delta_0 = G_0\alpha$ for some $\alpha \in \mathbb{R}^d$, then

$$\delta = U'\Lambda^{-1}\delta_0 = U'\Lambda^{-1}G_0\alpha = \Sigma V'\alpha.$$

where we have used $U'\Lambda^{-1}G_0 = \Sigma V'$, which follows from the definition $\Lambda^{-1}G_0 = U\Sigma V'$. In view of the definition of Σ , we know that $\delta_2 = 0$ and $\|\delta_2\|^2 = 0$. Hence the test has no power when the local departure is in the column space of G_0 . This result is qualitatively the same as the conventional J test using the chi-square approximation.

5 F -Approximation to Kernel-based J Statistic

The kernel LRV estimator and the associated J test are very popular in practice. In this section, we discuss the kernel J test under the so-called fixed- b asymptotics and show how the fixed- b asymptotic distribution can be approximated by an F distribution.

We maintain the following mild kernel conditions that hold for commonly used positive definite kernel functions.

Assumption 5 $\kappa(x) : \mathbb{R} \rightarrow [0, 1]$ is a symmetric and positive definite function with $\kappa(0) = 1$, $\int_0^\infty |\kappa(x)| dx < \infty$ and $\int_0^\infty \kappa^2(x) dx < \infty$.

For the given initial estimator $\tilde{\theta}_T$, the kernel LRV matrix estimator is

$$W_{T,\kappa}^0(\tilde{\theta}_T) = \frac{1}{T} \sum_{t=1}^T \sum_{\tau=1}^T \kappa_b\left(\frac{t-\tau}{T}\right) \left[\tilde{u}_t - \frac{1}{T} \sum_{s=1}^T \tilde{u}_s \right] \left[\tilde{u}_\tau - \frac{1}{T} \sum_{s=1}^T \tilde{u}_s \right]'$$

where as before $\tilde{u}_t = f(v_t, \tilde{\theta}_T)$ and $\kappa_b(r) = \kappa(r/b)$. Here we have followed Hall (2000) and centered the estimated moment condition \tilde{u}_t . Define the ‘‘centered’’ version of the kernel function

$$\tilde{\kappa}_b(t, \tau) := \tilde{\kappa}(t/b, \tau/b) = \kappa_b(t - \tau) - \int_0^1 \kappa_b(s - \tau) ds - \int_0^1 \kappa_b(t - s) ds + \int_0^1 \int_0^1 \kappa_b(r - s) dr ds.$$

While $\kappa_b(t - \tau)$ depends only on the difference between t and τ , the centered kernel function $\tilde{\kappa}_b(t, \tau)$ in general is not a function of only the difference $t - \tau$. By definition,

$$\int_0^1 \tilde{\kappa}_b(r, \tau) dr = \int_0^1 \tilde{\kappa}_b(t, s) ds = \int_0^1 \int_0^1 \tilde{\kappa}_b(r - s) dr ds = 0 \text{ for any } t \text{ and } \tau.$$

It is easy to show that $W_{T,\kappa}^0(\tilde{\theta}_T)$ is asymptotically equivalent to

$$W_{T,\kappa}(\hat{\theta}_T) = \frac{1}{T} \sum_{t=1}^T \sum_{\tau=1}^T \tilde{\kappa}_b\left(\frac{t}{T}, \frac{\tau}{T}\right) \tilde{u}_t \tilde{u}_\tau'.$$

The asymptotic equivalence holds regardless whether b is a fixed constant or goes to zero as $T \rightarrow \infty$.

Now using sum and integration by parts and the continuous mapping theorem, we have

$$W_{T,\kappa}(\hat{\theta}_T) \xrightarrow{d} W_{\infty,\kappa} \text{ and } W_{T,\kappa}(\hat{\theta}_T) \xrightarrow{d} W_{\infty,\kappa} \quad (6)$$

where

$$W_{\infty,\kappa} = \Lambda \tilde{W}_{\infty,\kappa} \Lambda' \text{ and } \tilde{W}_{\infty,\kappa} = \int_0^1 \int_0^1 \tilde{\kappa}_b(r, s) dB_m(r) dB_m'(s).$$

The asymptotic equivalence of $W_{T,\kappa}(\tilde{\theta}_T)$ and $W_{T,\kappa}(\hat{\theta}_T)$ implies that the over-identification test has the same asymptotic properties whether $W_{T,\kappa}(\tilde{\theta}_T)$ or $W_{T,\kappa}(\hat{\theta}_T)$ is used in constructing the testing statistic.

We use $W_{T,\kappa}(\tilde{\theta}_T)$ as the weighting matrix to construct the second step GMM estimator:

$$\hat{\theta}_{T,\kappa} = \arg \min_{\theta \in \Theta} g_T(\theta)' W_{T,\kappa}^{-1}(\tilde{\theta}_T) g_T(\theta).$$

As before, we can show that

$$\sqrt{T}(\hat{\theta}_{T,\kappa} - \theta_0) \xrightarrow{d} -[G_0' W_{\infty,\kappa}^{-1} G_0]^{-1} G_0' W_{\infty,\kappa}^{-1} \Lambda B_m(1)$$

and

$$\sqrt{T} g_T(\hat{\theta}_{T,\kappa}) \xrightarrow{d} \Lambda B_m(1) - G_0 [G_0' W_{\infty,\kappa}^{-1} G_0]^{-1} G_0' W_{\infty,\kappa}^{-1} \Lambda B_m(1). \quad (7)$$

Define the J statistic in the usual way as

$$J_{T,\kappa}(\delta_0) = T g_T(\hat{\theta}_{T,\kappa})' W_{T,\kappa}^{-1}(\tilde{\theta}_T) g_T(\hat{\theta}_{T,\kappa}) / q.$$

Since (6) and (7) hold jointly, we have

$$J_{T,\kappa} \xrightarrow{d} J_{\infty,\kappa}$$

where

$$\begin{aligned} J_{\infty,\kappa} &= \left[B_m(1) + U\delta - G_\Lambda [G_\Lambda' \tilde{W}_{\infty,\kappa}^{-1} G_\Lambda]^{-1} G_\Lambda' \tilde{W}_{\infty,\kappa}^{-1} B_m(1) \right]' \tilde{W}_{\infty,\kappa}^{-1} \\ &\quad \times \left[B_m(1) + U\delta - G_\Lambda [G_\Lambda' \tilde{W}_{\infty,\kappa}^{-1} G_\Lambda]^{-1} G_\Lambda' \tilde{W}_{\infty,\kappa}^{-1} B_m(1) \right] / q. \end{aligned}$$

The following theorem shows that $J_{\infty,\kappa}$ is pivotal under the null hypothesis.

Theorem 3 *Let Assumptions 1, 2, 4 and 5 hold. Then for a fixed b ,*

$$J_{T,\kappa}(\delta_0) \xrightarrow{d} J_{\infty,\kappa}(\delta_0)$$

where

$$J_{\infty,\kappa}(\delta_0) = [B_q(1) + \delta_2]' \left[\int_0^1 \int_0^1 \tilde{\kappa}_b(r, s) dB_q(r) dB_q'(s) \right]^{-1} [B_q(1) + \delta_2] / q$$

and $B_q(r)$ is a q -dimensional standard Brownian motion.

Under the null hypothesis, we have $\delta_0 = 0$ and

$$J_{\infty,\kappa} := J_{\infty,\kappa}(0) = B_q(1)' \left[\int_0^1 \int_0^1 \tilde{\kappa}_b(r, s) dB_q(r) dB_q'(s) \right]^{-1} B_q(1),$$

which is pivotal. Lee and Kuan (2009) consider the kernel-based over-identification test. To achieve pivotality under the fixed- b asymptotics, they propose to modify the usual J -statistic. It follows from Theorem 3 that the modification is unnecessary.

Under Assumption 5, $\tilde{\kappa}_b(r, s) \in L^2([0, 1] \times [0, 1])$. So it has a Fourier series representation:

$$\tilde{\kappa}_b(r, s) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \lambda_{ij} \phi_i(r) \phi_j(s)$$

where $\{\phi_i(r) \phi_j(s)\}$ is an orthonormal basis for $L^2([0, 1] \times [0, 1])$ and the convergence is in the L^2 space. Since $\int_0^1 \tilde{\kappa}_b(r, s) dr = \int_0^1 \tilde{\kappa}_b(r, s) ds = 0$, we know that

$$\int_0^1 \phi_i(r) dr = 0 \text{ for } i = 1, 2, \dots$$

In addition, by the symmetry of $\tilde{\kappa}_b(r, s)$, we can deduce that $\lambda_{ij} = \lambda_{ji}$.

Using the Fourier series representation, we can write

$$\tilde{W}_{\infty, \kappa} = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \lambda_{ij} \zeta_i \zeta_j'$$

where $\zeta_i = \int_0^1 \phi_i(s) dB_q(s) \sim iidN(0, I_q)$. To simplify the above representation, we note that

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \lambda_{ij} \zeta_i \zeta_j' = \lim_{N \rightarrow \infty} \sum_{i=1}^N \sum_{j=1}^N \lambda_{ij} \zeta_i \zeta_j' = \lim_{N \rightarrow \infty} \zeta' \Delta_N \zeta$$

where $\zeta_{N \times q} = (\zeta_1, \dots, \zeta_N)'$ and $\Delta_N = (\lambda_{ij})$ is an $N \times N$ symmetric matrix. Since $\kappa(\cdot)$ is positive definite, it can be shown that Δ_N is a positive definite matrix. Let $\Delta_N = HD_NH'$ be the spectral decomposition of Δ_N , where $D_N = diag(\lambda_1, \dots, \lambda_N)$ and $H'H = I_N$. Then

$$\lim_{N \rightarrow \infty} \zeta' \Delta_N \zeta = \lim_{N \rightarrow \infty} (\zeta' H) D_N (H' \zeta) = \sum_{i=1}^{\infty} \lambda_i \xi_i \xi_i'$$

where $H' \zeta = (\xi_1, \dots, \xi_N)'$ or $\xi_i = H' \zeta_i$. So $\tilde{W}_{\infty, \kappa}$ can be represented as

$$\tilde{W}_{\infty, \kappa} = \sum_{i=1}^{\infty} \lambda_i \xi_i \xi_i'. \quad (8)$$

By definition, $\xi_i \xi_i'$ follows a Wishart distribution $\mathbb{W}_q(I_q, 1)$, so $\sum_{i=1}^{\infty} \lambda_i \xi_i \xi_i'$ is an infinite weighted sum of independent Wishart distributions.

Using the representation in (8), we have

$$J_{\infty, \kappa} = {}^d \eta' \left(\sum_{i=1}^{\infty} \lambda_i \xi_i \xi_i' \right)^{-1} \eta/q \quad (9)$$

where $\xi_i \sim iidN(0, I_q)$, $\eta \sim N(0, I_q)$ and ξ_i is independent of η for all i . That is, $J_{\infty, \kappa}$ is equal in distribution to a quadratic form in standard normal variates with an independent and random weighting matrix.

We approximate the weighted sum of independent Wishart distributions by a simple Wishart distribution with an equivalent degree of freedom. That is, we approximate the distribution of $(\sum_{i=1}^{\infty} \lambda_i)^{-1} \sum_{i=1}^{\infty} \lambda_i \xi_i \xi_i'$ by that of $K^{-1} \sum_{i=1}^K \xi_i \xi_i'$. With this approximation, the distribution of $J_{\infty, \kappa}$ is approximately the same as that of J_{∞} .

Define

$$\mu_1 = \int_0^1 \tilde{\kappa}_b(r, r) dr \text{ and } \mu_2 = \int_0^1 \int_0^1 [\tilde{\kappa}_b(r, s)]^2 dr ds$$

and

$$K = \frac{\mu_1^2}{\mu_2} = \frac{\left[\int_0^1 \tilde{\kappa}_b(r, r) dr \right]^2}{\int_0^1 \int_0^1 [\tilde{\kappa}_b(r, s)]^2 dr ds}.$$

Theorem 4 (i) As $b \rightarrow 0$, $K = \lceil (bc_2)^{-1} \rceil + o(b^{-1})$ where $\lceil x \rceil$ denotes the smallest integer that is larger than x .

(ii) Let $J_{\infty, \kappa}^*$ be the corrected variate defined by

$$J_{\infty, \kappa}^* = \frac{\mu_1 (K - q + 1)}{K} J_{\infty, \kappa}. \quad (10)$$

Then, as $b \rightarrow 0$, we have

$$P(J_{\infty, \kappa}^* < z) = P(F_{q, K-q+1} < z) + o(b).$$

The parameter K is the “equivalent degree of freedom (EDF)” of the kernel LRV estimator. To the first order, the EDF is proportional to $1/b$ where the asymptotic variance of the kernel LRV estimator is proportional to b . Hence, as b decreases, i.e. as the degree of smoothing increases, the equivalent degree of freedom increases and the variance decreases. In other words, the higher the degree of freedom, the larger the degree of smoothing and the smaller the variance.

Compared to the correction factor in (3), the correction factor in (10) has an additional multiplicative constant μ_1 . This constant captures the effect of centering in constructing the kernel LRV estimator. For the series LRV, no centering is needed, and as a result this factor does not appear.

In finite samples, the correction factor may not be positive. It is easy to see that Theorem 4 holds with an asymptotically equivalent correction:

$$J_{\infty, \kappa}^* = \exp\{-b[c_1 + (q-1)c_2]\} J_{\infty, \kappa}$$

where

$$\exp\{-b[c_1 + (q-1)c_2]\} = \frac{\mu_1 (K - q + 1)}{K} + o(b).$$

The new correction factor $\exp\{-b[c_1 + (q-1)c_2]\}$ is guaranteed to be positive. We will use this version in the simulation study. Equivalently, we can correct the test statistic and define

$$J_{T, \kappa}^* = \exp\{-b[c_1 + (q-1)c_2]\} J_{T, \kappa}.$$

Then $J_{T, \kappa}^*$ is approximately distributed as $F_{q, K-q+1}$. We call the over-identification test based on this asymptotic approximation the kernel J^* test.

While the correction is motivated from the fixed- b asymptotics, it is entirely justifiable under the small- b asymptotics under which $b \rightarrow 0$ at a certain rate with the sample size. The

reason is that as $b \rightarrow 0$, $J_{T,\kappa}^*$ converges to $J_{T,\kappa}$ and critical values from $F_{q,K-q+1}$ converge to those from χ_q^2/q . So when $b \rightarrow 0$ as $T \rightarrow \infty$, the kernel J^* test is asymptotically equivalent to the conventional kernel J test.

6 Simulation Study

In this section, we study the finite sample performance of the J^* tests. We consider the following data generating process:

$$y_t = x_t\theta + \gamma z_{1,t} + \varepsilon_{yt}$$

where x_t is a scalar process generated by

$$x_t = \sum_{i=1}^m z_{it} + \varepsilon_{xt}.$$

We assume that $z_t = (z_{1t}, \dots, z_{mt})'$ follows either a VAR(1) process

$$z_t = \rho z_{t-1} + \sqrt{1 - \rho^2} e_{zt},$$

or a VMA(1) process

$$z_t = \rho e_{z,t-1} + \sqrt{1 - \rho^2} e_{zt},$$

where

$$e_{zt} = \left[\frac{e_{zt}^1 + e_{zt}^0}{\sqrt{2}}, \dots, \frac{e_{zt}^m + e_{zt}^0}{\sqrt{2}} \right]'$$

and $[e_{zt}^0, e_{zt}^1, \dots, e_{zt}^m]' \sim iidN(0, I_{m+1})$. By construction, the variance of z_{it} for any $i = 1, 2, \dots, m$ is 1. Due to the presence of the common shocks e_{zt}^0 , the correlation coefficient between z_{it} and z_{jt} for $i \neq j$ is 0.5. The DGP for $\varepsilon_t = (\varepsilon_{yt}, \varepsilon_{xt})'$ is the same as that for z_t except the dimensionality difference. The two vector processes ε_t and z_t are independent from each other. The model we consider reduces to that of Hall (2000) when $\rho = 0$ and reduces to Lee and Kuan (2009) when $m = 2$.

In the notation of this paper, we have $d = 1$, $f(v_t, \theta) = z_t(y_t - \theta x_t)$ where $v_t = [y_t, x_t, z_t']'$. It is also easy to verify that

$$Ef(v_t; \theta_0) = \gamma [1, 0.5\ell'_{m-1}]', \quad G_0 = Ez_t x_t = [1 + .5(m-1)]\ell_m$$

and

$$\Omega = \frac{1 - \rho^4}{2(1 - \rho^2)^2} (I_m + \ell_m \ell'_m)$$

where $\ell_m = (1, 1, \dots, 1)' \in \mathbb{R}^m$ is a column vector of ones. We set $\gamma = c/\sqrt{T}$ with $c = [0, 30]$ so that $\delta_0 = c [1, 0.5\ell'_{m-1}]'$.

We take $\rho = -0.8, -0.5, 0.0, 0.5, 0.8$ and 0.95 . We consider $m = 2$ and 5 and the corresponding degrees of over-identification are $q = 1$ and $q = 4$. For each test, we consider two significance levels $\alpha = 5\%$ and $\alpha = 10\%$ and two different sample sizes $T = 100, 200$. For the kernel-based tests, we consider four kernel functions: Bartlett (B), Parzen (P), Quadratic Spectral (Q) and Daniell (D) kernels. The number of simulation replications is 20000.

We examine the finite sample performances of J tests and J^* tests for different LRV estimator and reference distribution combinations. For each LRV estimator, we consider two versions of the over-identification test. The first version is the standard J test that uses χ_q^2/q as the reference distribution. These tests are referred to as S-J (series J test), B-J, P-J, Q-J, and D-J in the tables and figures below. The second version is the J^* test that is based on the modified J -statistic and uses $F(q, K - q + 1)$ as the reference distribution. These tests are labeled as S- J^* (series J^* test), B- J^* , P- J^* , Q- J^* and D- J^* . For all the tests, the initial estimator is the IV estimator with weight matrix $W_o = (Z'Z/T)$ where $Z = (z_1, \dots, z_T)'$. We use MSE optimal smoothing parameters implemented using the VAR(1) plug-in procedure in Andrews (1991) and Sun (2010b). For the smoothing parameter b , the formula is given in Andrews (1991). For the smoothing parameter K , the formula is given in Sun (2010b) and is reproduced here:

$$K_{MSE} = \left(\frac{\text{tr} [(I_{m^2} + \mathbb{K}_{mm})(\Omega \otimes \Omega)]}{4\text{vec}(B)'\text{vec}(B)} \right)^{1/5} T^{4/5},$$

where $B = -\pi^2/6 \times \Omega^{(2)}$, $\Omega^{(2)}$ is the asymptotic bias of the series LRV estimator, and \mathbb{K}_{mm} is the $m^2 \times m^2$ commutation matrix.

Table 1 gives the empirical size of the different testing methods for the VAR(1) case with sample size $T = 100$. First, as it is clear from the table, the conventional J tests can have a large size distortion. The size distortion increases with both the serial dependence and the degree of over-identification. The size distortion can be very severe. For example, when $\rho = 0.95$, $q = 4$ and $\alpha = 5\%$, the empirical size of the conventional kernel-based J test can be as high as 70%, which is far from 5%, the nominal size of the test. Second, the size distortion of the J^* test is substantially smaller than the corresponding J test. This is because the J^* test employs the asymptotic approximation that captures the estimation uncertainty of the LRV estimator. Third, among the J^* tests, the Parzen kernel-based test and the series-based test have the smallest size distortion. The empirical size of these two tests is also very close to the nominal size. The J^* test based on the Bartlett kernel tends to be over-sized, although it is significantly less size distorted than the conventional J test. To sum up, the table shows that the finite sample correction combined with an F approximation is very effective in reducing the size distortion of conventional J tests.

Table 2 presents the simulated empirical size for the VMA(1) case. The qualitative observations for the VAR(1) error remain valid. When the degree of over-identification is 1, the conventional J tests are relatively accurate in size. The J^* tests are even more accurate in size. When the degree of over-identification is 4, the conventional J tests have large size distortion. In contrast, the J^* tests have virtually no size distortion. This table provides further evidence that the F -distribution provides an accurate approximation to the sampling distribution of the finite sample corrected J -statistic.

Table 3 reports results for the case when the degree of over-identification is 4 and the sample size is 200. Compared to the cases with sample size 100, all tests become more accurate in size. This is well expected. While the conventional J tests remain over-sized, the J^* tests are remarkably accurate in size. As before, the series J^* test and the Parzen J^* test have virtually no size distortion. Other J^* tests are only slightly over-sized with the exception of the Bartlett J^* test, which can still be somewhat over-sized.

Figures 1-2 present the finite sample power in the VAR(1) case when the degree of over-identification is 1 and $\alpha = 5\%$. We compute the power using the 5% empirical finite

sample critical values obtained from the null distribution. So the finite sample power is size-adjusted and power comparisons are meaningful. The parameter configuration is the same as those for Table 1 except the DGP is generated under the local alternatives. Since the size-adjusted power of a J^* test is the same as that for the corresponding J test, we report the power of the series J^* test and that of the kernel J tests. We do so without the loss of generality. It is clear from Figure 1 that when the number of over-identifying restriction is 1, the power curves are very close to each other. In terms of the size-adjusted power, no test stands out. Figure 2 reports the same power comparison but with 4 over-identifying restrictions. Again, when $|\rho|$ is in the intermediate range, the power curves are indistinguishable. However, when $|\rho|$ is large, the series J^* test (or series J test) and the Bartlett J^* test (or the Bartlett J test) are more powerful than other J^* tests and J tests. The power advantage remains when the sample size is increased to 200. This is shown in Figure 3 where the parameter configuration is the same as Figure 2 except the sample size.

We do not report all power curves but summarize the main results here. For all the DGP's, the over-identification tests have more or less the same power with the exception that the series-based test and the Bartlett-kernel-based test are more powerful in some scenarios.

7 Conclusions

The paper provides an easy-to-implement over-identification test that is accurate in size. The test is based on the series LRV estimator that involves projecting the moment conditions onto a series of orthonormal basis functions. Since the basis functions have only low frequency components, the projections capture the long run behavior of the moment process and can therefore be used to estimate the asymptotic variance of its sample mean. An advantage of using the series LRV estimator is that the conventional J statistic, after a finite sample correction, is asymptotically F distributed. This result completely removes the computation burden of simulating critical values that is often required under the non-standard asymptotics for kernel LRV estimators. Our simulation demonstrates that the resulting series J^* test using F critical values has virtually no size distortion in almost all cases considered.

The paper uses the MSE criterion to select the smoothing parameter for the J^* test. The MSE criterion may be a reasonable choice but is not most suitable for hypothesis testing problems. Sun, Phillips and Jin (2008), and Sun (2010a,b,c,d) consider selecting the smoothing parameter to maximize the local asymptotic power while controlling for the size of the test. It is interesting to extend their approach to the current setting.

8 Appendix

8.1 Additional Technical Results

Lemma 1 *As $b \rightarrow 0$, we have*

$$(a) \mu_1 = \sum_{i=1}^{\infty} \lambda_i = \int_0^1 \tilde{\kappa}_b(r, r) dr = 1 - bc_1 + o(b).$$

$$(b) \mu_2 = \sum_{i=1}^{\infty} (\lambda_i)^2 = \int_0^1 \int_0^1 \tilde{\kappa}_b^2(r, s) dr ds = bc_2 + o(b).$$

$$(c) \mu_j = \sum_{i=1}^{\infty} (\lambda_i)^j = O(b^{j-1}), \text{ for } j \geq 2.$$

Proof of Lemma 1. Parts (a) and (b): We only need to show $\sum_{i=1}^{\infty} \lambda_i = \int_0^1 \tilde{\kappa}_b(r, r) dr$ and $\sum_{i=1}^{\infty} (\lambda_i)^2 = \int_0^1 \int_0^1 \tilde{\kappa}_b(r, s) dr ds$ as the last equalities in parts (a) and (b) hold by Lemma 2 of Sun (2010d). By definition

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \lambda_{ij} \zeta_i \zeta_j' = \sum_{i=1}^{\infty} \lambda_i \xi_i \xi_i'$$

where $\zeta_i \sim iidN(0, I_q)$ and $\xi_i \sim iidN(0, I_q)$. Taking expectations on both sides yields

$$\sum_{i=1}^{\infty} \lambda_{ii} = \sum_{i=1}^{\infty} \lambda_i. \quad (11)$$

It follows from the Fourier series expansion of $\tilde{\kappa}_b(r, s)$ that

$$\int_0^1 \tilde{\kappa}_b(r, r) dr = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \lambda_{ij} \int_0^1 \phi_i(r) \phi_j(r) dr = \sum_{i=1}^{\infty} \lambda_{ii}. \quad (12)$$

Combining (11) and (12) yields $\int_0^1 \tilde{\kappa}_b(r, r) dr = \sum_{i=1}^{\infty} \lambda_i$.

To show that $\sum_{i=1}^{\infty} (\lambda_i)^2 = \int_0^1 \int_0^1 \tilde{\kappa}_b(r, s) dr ds$, we note that

$$\begin{aligned} \sum_{i=1}^{\infty} (\lambda_i)^2 &= \lim_{N \rightarrow \infty} tr(\Delta_N^2) = \lim_{N \rightarrow \infty} \sum_{i=1}^N \sum_{j=1}^N \lambda_{ij}^2 \\ &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \lambda_{ij}^2 = \int_0^1 \int_0^1 \tilde{\kappa}_b^2(r, s) dr ds \end{aligned}$$

where the last equality follows from Parseval's identity.

Part (c): Note that

$$\begin{aligned} \mu_j &= \sum_{i=1}^{\infty} (\lambda_i)^j = \lim_{N \rightarrow \infty} tr(\Delta_N^j) = \sum_{i_1, i_2, \dots, i_j} \lambda_{i_1 i_2} \lambda_{i_2 i_3} \cdots \lambda_{i_{j-1} i_j} \lambda_{i_j i_1} \\ &= \int_0^1 \int_0^1 \cdots \int_0^1 \int_0^1 \tilde{\kappa}_b(r_1, r_2) \tilde{\kappa}_b(r_2, r_3) \cdots \tilde{\kappa}_b(r_{j-1}, r_j) \tilde{\kappa}_b(r_j, r_1) dr_1 dr_2 \cdots dr_j, \end{aligned}$$

so

$$\begin{aligned} |\mu_j| &\leq \int_0^1 \int_0^1 \cdots \int_0^1 \int_0^1 |\tilde{\kappa}_b(r_1, r_2)| |\tilde{\kappa}_b(r_2, r_3)| \cdots |\tilde{\kappa}_b(r_{j-1}, r_j)| dr_1 dr_2 \cdots dr_j \\ &\leq \left(\int_0^1 \sup_s |\tilde{\kappa}_b(r, s)| dr \right)^{j-1}. \end{aligned}$$

In view of the definition

$$\tilde{\kappa}_b(r, s) = k_b(r - s) - \int_0^1 k_b(r - p)dp - \int_0^1 k_b(s - q)dq + \int_0^1 \int_0^1 k_b(p - q)dpdq,$$

we have

$$\begin{aligned} \int_0^1 \sup_s |\tilde{\kappa}_b(r, s)| dr &\leq 4 \sup_s \int_0^1 |k_b(r - s)| dr \\ &= 4 \sup_{s \in [0,1]} \left(\int_{-s}^{1-s} |k_b(x)| dx \right) \\ &= 4b \sup_{s \in [0,1]} \left(\int_{-s/b}^{(1-s)/b} |k(x)| dx \right) \\ &\leq 4b \int_{-\infty}^{\infty} |k(x)| dx = O(b). \end{aligned}$$

Therefore, $\mu_j = O(b^{j-1})$. ■

Lemma 2 Let $\{\nu_i := \nu_i(b)\}$ be a nonnegative weighting sequence satisfying

$$\sum_{i=1}^{\infty} \nu_i = 1 \text{ and } \sum_{i=1}^{\infty} (\nu_i)^j = O(b^{j-1}) \text{ for } j \geq 2.$$

Define

$$\sum_{i=1}^{\infty} \nu_i \xi_i \xi_i' = \begin{pmatrix} \nu_{11} & \nu_{12} \\ \nu_{21} & \nu_{22} \end{pmatrix} \text{ and } \nu_{11 \cdot 2} = \nu_{11} - \nu_{12} \nu_{22}^{-1} \nu_{21}.$$

where $\xi_i \sim iidN(0, I_q)$, ν_{11} is a scalar, and ν_{22} is a $(q-1) \times (q-1)$ matrix. As $b \rightarrow 0$, we have

- (a) $E\nu_{11 \cdot 2} = 1 - (q-1) \sum_{i=1}^{\infty} \nu_i^2 + o(b)$,
- (b) $E(\nu_{11 \cdot 2})^2 = 1 - 2(q-2) \sum_{i=1}^{\infty} \nu_i^2 + o(b)$,
- (c) $var(\nu_{11 \cdot 2}) = 2 \sum_{i=1}^{\infty} \nu_i^2 + o(b)$.

Proof of Lemma 2. Part (a): Let $\xi_i' = (\xi_{i1}', \xi_{i2}')$ where $\xi_{i1} \in \mathbb{R}$ and $\xi_{i2} \in \mathbb{R}^{q-1}$. Then

$$\begin{aligned} &E\nu_{11 \cdot 2} \\ &= E \sum_{i=1}^{\infty} \nu_i \xi_{i1} \xi_{i1}' - E \left(\sum_{i=1}^{\infty} \nu_i \xi_{i1} \xi_{i2}' \right) \left(\sum_{i=1}^{\infty} \nu_i \xi_{i2} \xi_{i2}' \right)^{-1} \left(\sum_{i=1}^{\infty} \nu_i \xi_{i2} \xi_{i1}' \right) \\ &= \sum_{i=1}^{\infty} \nu_i - Etr \left[\left(\sum_{i=1}^{\infty} \nu_i \xi_{i2} \xi_{i2}' \right)^{-1} \left(\sum_{i=1}^{\infty} \nu_i \xi_{i2} \xi_{i1}' \right) \left(\sum_{i=1}^{\infty} \nu_i \xi_{i1} \xi_{i2}' \right)' \right] \\ &= \sum_{i=1}^{\infty} \nu_i - Etr \left[\left(\sum_{i=1}^{\infty} \nu_i \xi_{i2} \xi_{i2}' \right)^{-1} \left(\sum_{i=1}^{\infty} \nu_i^2 \xi_{i2} \xi_{i2}' \right) \right]. \end{aligned}$$

But

$$\begin{aligned}
& Etr \left[\left(\sum_{i=1}^{\infty} \nu_i \xi_{i2} \xi'_{i2} \right)^{-1} \left(\sum_{i=1}^{\infty} \nu_i^2 \xi_{i2} \xi'_{i2} \right) \right] \\
&= Etr \left[I_{q-1} - \sum_{i=1}^{\infty} \nu_i (I_{q-1} - \xi_{i2} \xi'_{i2}) \right]^{-1} \left(\sum_{i=1}^{\infty} \nu_i^2 \xi_{i2} \xi'_{i2} \right) \\
&= (q-1) \sum_{i=1}^{\infty} \nu_i^2 + Etr \sum_{j=1}^{\infty} \left[\sum_{i=1}^{\infty} \nu_i (I_{q-1} - \xi_{i2} \xi'_{i2}) \right]^j \left(\sum_{i=1}^{\infty} \nu_i^2 \xi_{i2} \xi'_{i2} \right) \\
&= (q-1) \sum_{i=1}^{\infty} \nu_i^2 + O \left[\sum_{j=1}^{\infty} C^j \sum_{i=1}^{\infty} \nu_i^{j+2} \right] \\
&= (q-1) \sum_{i=1}^{\infty} \nu_i^2 + O \left[\sum_{j=1}^{\infty} C^j b^{j+1} \right] \\
&= (q-1) \sum_{i=1}^{\infty} \nu_i^2 + o(b). \tag{13}
\end{aligned}$$

where C is a fixed constant and C^j is a bound for $Etr (I_{q-1} - \xi_{i_1 2} \xi'_{i_1 2}) \dots (I_{q-1} - \xi_{i_j 2} \xi'_{i_j 2}) \xi_{i_2} \xi'_{i_2}$. So

$$E\nu_{11 \cdot 2} = 1 - (q-1) \sum_{i=1}^{\infty} \nu_i^2 + o(b)$$

Part (b): We write

$$E(\nu_{11} - \nu_{12} \nu_{22}^{-1} \nu_{21})^2 = E\nu_{11}^2 + E\nu_{12} \nu_{22}^{-1} \nu_{21} \nu_{12} \nu_{22}^{-1} \nu_{21} - 2E\nu_{11} \nu_{12} \nu_{22}^{-1} \nu_{21}.$$

Using the same argument as in (13), we have

$$\begin{aligned}
E\nu_{11}^2 &= \left(\sum_{i=1}^{\infty} \nu_i \right)^2 + 2 \sum_{i=1}^{\infty} \nu_i^2, \\
E\nu_{11} \nu_{12} \nu_{22}^{-1} \nu_{21} &= \left(\sum_{i=1}^{\infty} \nu_i \right) Etr \left(\sum_{k=1}^{\infty} \nu_k \xi_{k2} \xi'_{k2} \right)^{-1} \left(\sum_{i=1}^{\infty} \nu_i^2 \xi_{i2} \xi'_{i2} \right) \\
&\quad + 2Etr \left(\sum_{i=1}^{\infty} \nu_i \xi_{i2} \xi'_{i2} \right)^{-1} \left(\sum_{i=1}^{\infty} \nu_i^3 \xi_{i2} \xi'_{i2} \right) \\
&= \sum_{i=1}^{\infty} \nu_i^2 (q-1) + o(b),
\end{aligned}$$

and

$$\begin{aligned}
& E\nu_{12}\nu_{22}^{-1}\nu_{21}\nu_{12}\nu_{22}^{-1}\nu_{21} \\
&= Etr \left(\left[\sum_{i=1}^{\infty} \nu_i \xi_{i2} \xi'_{i2} \right]^{-1} \left[\sum_{j=1}^{\infty} \nu_j^2 \xi_{j2} \xi'_{j2} \right] \right) tr \left(\left[\sum_{i=1}^{\infty} \nu_i \xi_{i2} \xi'_{i2} \right]^{-1} \left[\sum_{j=1}^{\infty} \nu_j^2 \xi_{j2} \xi'_{j2} \right] \right) \\
&+ 2Etr \left\{ \left[\sum_{i=1}^{\infty} \nu_i \xi_{i2} \xi'_{i2} \right]^{-1} \left[\sum_{j=1}^{\infty} \nu_j^2 \xi_{j2} \xi'_{j2} \right] \left[\sum_{i=1}^{\infty} \nu_i \xi_{i2} \xi'_{i2} \right]^{-1} \left[\sum_{j=1}^{\infty} \nu_j^2 \xi_{j2} \xi'_{j2} \right] \right\} \\
&= o(b).
\end{aligned}$$

Hence

$$E(\nu_{11.2})^2 = 1 - 2(q-2) \sum_{i=1}^{\infty} \nu_i^2 + o(b).$$

Part (c) follows from parts (a) and (b). ■

8.2 Proof of the Main Results

Proof of Theorem 1. Under Assumptions (i)-(iii) in the theorem, $K\tilde{W}_{\infty}$ follows a Wishart distribution $\mathbb{W}(I_m, K)$ as

$$K\tilde{W}_{\infty} = \sum_{k=1}^K \xi_k \xi'_k.$$

where $\xi_k = \int_0^1 \Phi_k(r) dB_m(r)$ is iid $N(0, I_m)$. In addition, for any k , ξ_k and $B_m(1)$ are independent because both are normal and

$$\begin{aligned}
E \int_0^1 \Phi_k(r) dB_m(r) B'_m(1) &= E \int_0^1 \Phi_k(r) dB_m(r) \int_0^1 dB'_m(s) \\
&= I_m \int_0^1 \Phi_k(r) dr = 0.
\end{aligned}$$

Consequently, \tilde{W}_{∞}^{-1} is independent of $B_m(1)$.

Let $G_{\Lambda} = U_{m \times m} \Sigma_{m \times d} V'_{d \times d}$ be the singular value decomposition of $G_{\Lambda} = \Lambda^{-1} G_0$. By definition, $UU' = UU' = I_m$, $VV' = V'V = I_d$ 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. Now

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

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

$$\begin{aligned} J_\infty &\stackrel{d}{=} \left\{ B_m(1) - \Sigma \left[\Sigma' \left(\tilde{W}_\infty^{-1} \right) \Sigma \right]^{-1} \Sigma' \tilde{W}_\infty^{-1} B_m(1) \right\}' \tilde{W}_\infty^{-1} \\ &\times \left\{ B_m(1) - \Sigma \left[\Sigma' \left(\tilde{W}_\infty^{-1} \right) \Sigma \right]^{-1} \Sigma' \tilde{W}_\infty^{-1} B_m(1) \right\} / q \end{aligned}$$

where $\stackrel{d}{=}$ signifies “is equal to in distribution” and \tilde{W}_∞^{-1} remains independent of $B_m(1)$.

Let

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

where C_{11} is a $d \times d$ matrix, C_{22} is a $q \times q$ matrix and $C_{12} = C_{21}'$. With this partition of \tilde{W}_∞^{-1} , we have

$$\begin{aligned} &\Sigma \left[\Sigma' \left(\tilde{W}_\infty^{-1} \right) \Sigma \right]^{-1} \Sigma' \\ &= \begin{pmatrix} A \\ O \end{pmatrix} \left\{ \begin{pmatrix} A' & O' \end{pmatrix} \tilde{W}_\infty^{-1} \begin{pmatrix} A \\ O \end{pmatrix} \right\}^{-1} \begin{pmatrix} A' & O' \end{pmatrix} \\ &= \begin{bmatrix} C_{11}^{-1} & O_{12} \\ O_{21} & O_{22} \end{bmatrix} \end{aligned} \tag{14}$$

where O_{ij} are matrices of zeros with compatible dimensions. Letting $B_m(1) = [B_d'(1), B_q'(1)]'$ and plugging (14) into J_∞ yields

$$\begin{aligned} J_\infty &\stackrel{d}{=} \left\{ B_m(1) - \begin{bmatrix} C_{11}^{-1} & O_{12} \\ O_{21} & O_{22} \end{bmatrix} \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} B_m(1) \right\}' \tilde{W}_\infty^{-1} \\ &\times \left\{ B_m(1) - \begin{bmatrix} C_{11}^{-1} & O_{12} \\ O_{21} & O_{22} \end{bmatrix} \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} B_m(1) \right\} / q \\ &= \left\{ B_m(1) - \begin{bmatrix} I_d & C_{11}^{-1}C_{12} \\ O_{21} & O_{22} \end{bmatrix} B_m(1) \right\}' \tilde{W}_\infty^{-1} \\ &\times \left\{ B_m(1) - \begin{bmatrix} I_d & C_{11}^{-1}C_{12} \\ O_{21} & O_{22} \end{bmatrix} B_m(1) \right\} / q \\ &= \begin{pmatrix} -C_{11}^{-1}C_{12}B_q(1) \\ B_q(1) \end{pmatrix}' \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \begin{pmatrix} -C_{11}^{-1}C_{12}B_q(1) \\ B_q(1) \end{pmatrix} / q \\ &= \begin{pmatrix} -C_{11}^{-1}C_{12}B_q(1) \\ B_q(1) \end{pmatrix}' \begin{pmatrix} O \\ [C_{22} - C_{21}C_{11}^{-1}C_{12}] B_q(1) \end{pmatrix} / q \\ &= B_q(1)' [C_{22} - C_{21}C_{11}^{-1}C_{12}] B_q(1) / q. \end{aligned}$$

To see the distribution of J_∞ , we note that

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

so by the partitioned inverse formula

$$C_{22} - C_{21}C_{11}^{-1}C_{12} = (C^{22})^{-1}.$$

But by definition $KC^{22} \sim \mathbb{W}(I_q, K)$. Hence

$$\frac{1}{K} J_\infty \stackrel{d}{=} B_q(1)' [KC^{22}]^{-1} B_q(1)/q,$$

or

$$\frac{(K-q+1)}{K} J_\infty \stackrel{d}{=} \frac{\chi_q^2/q}{\chi_{K-q-1}^2/(K-q+1)} = F(q, K-q+1).$$

The distributional equivalence follows from Proposition 8.2 in Bilodeau and Brenner (1999) where the notation F_c denotes the canonical F distribution (Bilodeau and Brenner (1999), page 42).

We have therefore shown that when K is fixed,

$$\frac{(K-q+1)}{K} J_T \xrightarrow{d} F(q, K-q+1).$$

■

Proof of Theorem 2. We use the same notation as in the proof of Theorem 1.

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

Using the same argument as before, we can write

$$\begin{aligned} J_\infty(\delta_0) & \stackrel{d}{=} \left\{ [B_m(1) + \delta] - \Sigma \left[\Sigma' \left(\tilde{W}_\infty^{-1} \right) \Sigma \right]^{-1} \Sigma' \tilde{W}_\infty^{-1} [B_m(1) + \delta] \right\}' \tilde{W}_\infty^{-1} \\ & \times \left\{ [B_m(1) + \delta] - \Sigma \left[\Sigma' \left(\tilde{W}_\infty^{-1} \right) \Sigma \right]^{-1} \Sigma' \tilde{W}_\infty^{-1} [B_m(1) + \delta] \right\} / q \\ & = (B_q(1) + \delta_2) [C_{22} - C_{21} C_{11}^{-1} C_{12}] (B_q(1) + \delta_2) / q. \end{aligned}$$

Hence

$$\frac{(K-q+1)}{K} J_\infty(\delta_0) \stackrel{d}{=} F(q, K-q+1, \|\delta_2\|^2),$$

as desired. ■

Proof of Theorem 3. We prove the result under the null that $\delta_0 = 0$. The proof under the local alternatives is similar and is omitted here. As before, letting $G_\Lambda = U\Sigma V'$ be the svd of G_Λ , we have

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

and

$$J_{\infty, \kappa} \stackrel{d}{=} \left[B_m(1) - \Sigma \left[\Sigma' \tilde{W}_{\infty, \kappa}^{-1} \Sigma \right]^{-1} \Sigma' \tilde{W}_{\infty, \kappa}^{-1} B_m(1) \right]' \\ \times \tilde{W}_{\infty, \kappa}^{-1} \left[B_m(1) - \Sigma \left[\Sigma' \tilde{W}_{\infty, \kappa}^{-1} \Sigma \right]^{-1} \Sigma' \tilde{W}_{\infty, \kappa}^{-1} B_m(1) \right] / q.$$

Denote

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

By definition $(C_{\kappa}^{22})^{-1} = C_{22, \kappa} - C_{21, \kappa} C_{11, \kappa}^{-1} C_{12, \kappa}$. Using exactly the same calculation as in the proof of Theorem 1, we have

$$J_{\infty, \kappa} \stackrel{d}{=} B_q(1)' \left[C_{22, \kappa} - C_{21, \kappa} C_{11, \kappa}^{-1} C_{12, \kappa} \right] B_q(1) / q \\ = B_q(1)' (C_{\kappa}^{22})^{-1} B_q(1) / q \\ = B_q(1)' \left[\int_0^1 \int_0^1 \tilde{\kappa} \left(\frac{r}{b}, \frac{s}{b} \right) dB_q(r) dB_q'(s) \right]^{-1} B_q(1) / q$$

as desired. ■

Proof of Theorem 4. (i) It follows from Lemma 1 that

$$K = \frac{\mu_1^2}{\mu_2} = \frac{[1 - bc_1 + o(b)]^2}{[bc_2 + o(b)]} = \frac{1}{bc_2} \frac{[1 - bc_1 + o(b)]^2}{1 + o(1)} \\ = \frac{1}{bc_2} (1 + o(1)) = \frac{1}{bc_2} + o\left(\frac{1}{bc_2}\right).$$

(ii) Define

$$T_{\nu} = \frac{(K - q + 1)}{Kq} \eta' \left(\sum_{i=1}^{\infty} \nu_i \xi_i \xi_i' \right)^{-1} \eta$$

for some weighting sequence $\nu = (\nu_1, \nu_2, \dots)$ satisfying $\sum_{i=1}^{\infty} \nu_i = 1$. Then $J_{\infty, \kappa}^* = T_{\nu}$ when

$$\nu = \nu_{\lambda} =: (\lambda_1 / \mu_1, \lambda_2 / \mu_1, \dots),$$

and $F_{q, K-q+1} = T_{\nu}$ when

$$\nu = \nu_K = (K^{-1}, K^{-1}, \dots, K^{-1}, 0, \dots).$$

Let \mathcal{H} be an orthonormal matrix such that $\mathcal{H} = (\eta / \|\eta\|, \Pi)'$ where Π is a $q \times (q - 1)$ matrix, then

$$T_{\nu} = \frac{(K - q + 1)}{Kq} (\mathcal{H}\eta)' \left(\sum_{i=1}^{\infty} \nu_i \mathcal{H}\xi_i (\mathcal{H}\xi_i)' \right)^{-1} (\mathcal{H}\eta) \\ = \frac{(K - q + 1)}{Kq} \|\eta\|^2 e_1' \left(\sum_{i=1}^{\infty} \nu_i \mathcal{H}\xi_i (\mathcal{H}\xi_i)' \right)^{-1} e_1$$

where $e_1 = (1, 0, 0, \dots, 0, 0)'$. Note that $\|\eta\|^2$ is independent of $\mathcal{H}\xi_i$ and $\mathcal{H}\xi_i$ has the same distribution as ξ_i , we can write

$$T_\nu \stackrel{d}{=} \frac{(K - q + 1) \|\eta\|^2}{K} e_1' \left(\sum_{i=1}^{\infty} \nu_i \xi_i \xi_i' \right)^{-1} e_1.$$

As a result,

$$\begin{aligned} & P(T_\nu < z) \\ &= P \left(\frac{(K - q + 1) \|\eta\|^2}{K} e_1' \left(\sum_{i=1}^{\infty} \nu_i \xi_i \xi_i' \right)^{-1} e_1 < z \right) \\ &= EG_q \left(z \frac{K}{K - q + 1} \left[e_1' \left(\sum_{i=1}^{\infty} \nu_i \xi_i \xi_i' \right)^{-1} e_1 \right]^{-1} \right), \end{aligned}$$

where $G_q(\cdot)$ is the CDF of the distribution χ_q^2/q .

Define

$$\sum_{i=1}^{\infty} \nu_i \xi_i \xi_i' = \begin{pmatrix} \nu_{11} & \nu_{12} \\ \nu_{21} & \nu_{22} \end{pmatrix}$$

where ν_{11} is a scalar and ν_{22} is a $(q - 1) \times (q - 1)$ matrix. Then

$$\left[e_1' \left(\sum_{i=1}^{\infty} \nu_i \xi_i \xi_i' \right)^{-1} e_1 \right]^{-1} = \nu_{11} - \nu_{12} \nu_{22}^{-1} \nu_{21} = \nu_{11.2},$$

and so

$$P(T_\nu < z) = EG_q \left(z \frac{K}{K - q + 1} \nu_{11.2} \right).$$

By Lemma 1, the weighting sequence ν_λ satisfies the assumptions in Lemma 2. It is easy to see that the weighting sequence ν_K also satisfies the assumptions. Hence, by Lemma 2, when $\nu = \nu_\lambda$ or ν_K , we have

$$\begin{aligned} E\nu_{11.2} &= \sum_{i=1}^{\infty} \nu_i - \sum_{i=1}^{\infty} (\nu_i)^2 (q - 1) + o(b) \\ &= 1 - \frac{\mu_2}{\mu_1^2} (q - 1) + o(b) = 1 - bc_2 (q - 1) + o(b) \end{aligned}$$

and

$$\begin{aligned} \text{var}(\nu_{11.2}) &= \sum_{i=1}^{\infty} (\nu_i)^2 [(2q - 2) - 2q + 4] + o(b) \\ &= 2 \frac{\mu_2}{\mu_1^2} + o(b) = 2bc_2 + o(b). \end{aligned}$$

That is, $\nu_{11.2}$ concentrates around $E\nu_{11.2}$. Let $\tilde{\nu}_{11.2}$ be a random variable between $\nu_{11.2}$ and $E\nu_{11.2}$. A Taylor expansion yields:

$$\begin{aligned}
& P(T_\nu < z) \\
&= G_q \left(z \frac{K}{K-q+1} E\nu_{11.2} \right) + \frac{1}{2} G_q'' \left(z \frac{K}{K-q+1} E\nu_{11.2} \right) \text{var}(\nu_{11.2}) \\
&\quad + E \left[\frac{1}{2} G_q'' \left(z \frac{K}{K-q+1} \tilde{\nu}_{11.2} \right) - \frac{1}{2} G_q'' \left(z \frac{K}{K-q+1} E\nu_{11.2} \right) \right] (\nu_{11.2} - E\nu_{11.2})^2 \\
&= G_q \left(z \frac{K}{K-q+1} E\nu_{11.2} \right) + \frac{1}{2} G_q'' \left(z \frac{K}{K-q+1} E\nu_{11.2} \right) \text{var}(\nu_{11.2}) + o(b). \quad (15)
\end{aligned}$$

This result, combined with the observation that $E\nu_{11.2}$ and $\text{var}(\nu_{11.2})$ are the same across $\nu = \nu_\lambda$ and $\nu = \nu_K$, up to order $o(b)$, yields

$$P(J_{\infty, \kappa}^* < z) = P(F_{q, K-q+1} \leq z) + o(b).$$

■

References

- [1] Altonji, J. G. and L. M. Segal (1996): “Small-sample Bias in GMM Estimation of Covariance Structures.” *Journal of Business & Economic Statistics* 14, 353–366.
- [2] Andrews, D. W. K. (1991): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation.” *Econometrica*, 59, 817–854.
- [3] Bilodeau, M. and D. Brenner (1999): *Theory of Multivariate Statistics*, Springer, New York.
- [4] Hall, A. R. (2000): “Covariance Matrix Estimation and the Power of the Over-identifying Restrictions Test.” *Econometrica* 68, 1517–1527.
- [5] Hall, A. R. (2005): *Generalized Method of Moments*. New York: Oxford University Press.
- [6] Hall, P. and J. L. Horowitz (1996): “Bootstrap Critical Values for Tests Based on Generalized-method-of-moments Estimators.” *Econometrica* 64, 891–916.
- [7] Hannan, E. J. (1957): “The Variance of the Mean of a Stationary Process.” *Journal of the Royal Statistical Society, Series B*, Vol. 19, No. 2, pp. 282–285
- [8] Hansen, L. P. (1982): “Large Sample Properties of Generalized Method of Moments Estimators.” *Econometrica* 50, 1029–1054.
- [9] Hotelling H. (1931): “The Generalization of Student’s Ratio.” *Annals of Mathematical Statistics*, 2, 360–378.
- [10] Jansson, M. (2004): “On the Error of Rejection Probability in Simple Autocorrelation Robust Tests.” *Econometrica*, 72, 937–946.
- [11] Kiefer, N. M., T. J. Vogelsang and H. Bunzel (2000): “Simple Robust Testing of Regression Hypotheses.” *Econometrica*, 68, 695–714.
- [12] Kiefer, N. M. and T. J. Vogelsang (2002a): “Heteroskedasticity-autocorrelation Robust Testing Using Bandwidth Equal to Sample Size.” *Econometric Theory*, 18, 1350–1366.
- [13] ——— (2002b): “Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation.” *Econometrica*, 70, 2093–2095.
- [14] ——— (2005): “A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests.” *Econometric Theory*, 21, 1130–1164.
- [15] Lee, W. and C. Kuan (2009): “Testing Over-identifying Restrictions without Consistent Estimation of the Asymptotic Covariance Matrix.” Working paper.
- [16] Newey, W. K. and K. D. West (1987): “A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica*, 55, 703–708.
- [17] Newey, W. K. and K. D. West (1994): “Automatic Lag Selection in Covariance Matrix Estimation.” *The Review of Economic Studies*, 61, 631–654.

- [18] Percival D. and A. T. Walden (1993): *Spectral Analysis For Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge University Press.
- [19] Phillips, P. C. B. (2005): “HAC Estimation by Automated Regression.” *Econometric Theory*, 21, 116–142.
- [20] Sargan, J. D. (1958): “The Estimation of Economic Relationships Using Instrumental Variables.” *Econometrica* 26 (1958), pp. 393–415
- [21] Stoica P. and R. Moses (2005): *Spectral Analysis of Signals*. Pearson Prentice Hall, New Jersey.
- [22] Sun, Y., P. C. B. Phillips, and S. Jin (2008): “Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing.” *Econometrica*, 76, 175–194.
- [23] Sun, Y. (2010a): “Autocorrelation Robust Inference Using Nonparametric Series Method and Testing-optimal Smoothing Parameter.” Working paper, Department of Economics, UC San Diego.
- [24] Sun, Y. (2010b): “Robust Trend Inference with Series Variance Estimator and Testing-optimal Smoothing Parameter.” Working paper, Department of Economics, UC San Diego.
- [25] Sun, Y. (2010c): “Calibrating Optimal Smoothing Parameter in Heteroscedasticity and Autocorrelation Robust Tests.” Working paper, Department of Economics, UC San Diego.
- [26] Sun, Y. (2010d): “Let’s Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroscedasticity and Autocorrelation Robust Inference.” Working paper, Department of Economics, UC San Diego.

Table 1: Empirical size of different J tests and J* tests for the VAR(1) case with nominal size α , degree of overidentification q , and sample size $T = 100$

	S-J	S-J*	B-J	B-J*	P-J	P-J*	Q-J	Q-J*	D-J	D-J*
$q = 1, \alpha = 5\%$										
$\rho=-0.8$	0.095	0.061	0.126	0.085	0.117	0.063	0.112	0.065	0.109	0.064
$\rho=-0.5$	0.071	0.058	0.083	0.067	0.078	0.058	0.076	0.059	0.077	0.059
$\rho=0.00$	0.052	0.047	0.054	0.049	0.056	0.048	0.054	0.047	0.054	0.047
$\rho=0.50$	0.073	0.060	0.084	0.068	0.079	0.060	0.078	0.061	0.078	0.061
$\rho=0.8$	0.097	0.062	0.128	0.086	0.118	0.064	0.113	0.066	0.111	0.065
$\rho=0.95$	0.160	0.065	0.243	0.120	0.243	0.077	0.228	0.082	0.221	0.080
$q = 1, \alpha = 10\%$										
$\rho=-0.8$	0.164	0.126	0.207	0.160	0.193	0.128	0.187	0.130	0.183	0.131
$\rho=-0.5$	0.133	0.121	0.153	0.132	0.144	0.119	0.142	0.120	0.142	0.122
$\rho=0.00$	0.104	0.098	0.110	0.100	0.109	0.099	0.108	0.098	0.108	0.098
$\rho=0.50$	0.133	0.119	0.148	0.132	0.141	0.118	0.140	0.120	0.141	0.120
$\rho=0.8$	0.166	0.127	0.207	0.160	0.194	0.130	0.188	0.132	0.185	0.133
$\rho=0.95$	0.254	0.145	0.348	0.222	0.345	0.160	0.330	0.166	0.319	0.164
$q = 4, \alpha = 5\%$										
$\rho=-0.8$	0.271	0.042	0.329	0.083	0.361	0.049	0.335	0.056	0.342	0.059
$\rho=-0.5$	0.135	0.047	0.153	0.060	0.158	0.050	0.147	0.052	0.148	0.052
$\rho=0.00$	0.086	0.044	0.094	0.045	0.096	0.044	0.090	0.044	0.093	0.045
$\rho=0.50$	0.142	0.050	0.160	0.063	0.165	0.051	0.153	0.052	0.155	0.053
$\rho=0.8$	0.283	0.044	0.342	0.087	0.371	0.052	0.347	0.058	0.352	0.061
$\rho=0.95$	0.506	0.081	0.650	0.113	0.730	0.039	0.710	0.066	0.718	0.072
$q = 4, \alpha = 10\%$										
$\rho=-0.8$	0.370	0.100	0.438	0.170	0.461	0.115	0.436	0.125	0.440	0.125
$\rho=-0.5$	0.214	0.100	0.241	0.124	0.243	0.106	0.231	0.107	0.233	0.107
$\rho=0.00$	0.153	0.093	0.162	0.095	0.165	0.092	0.159	0.093	0.161	0.093
$\rho=0.50$	0.223	0.105	0.249	0.127	0.251	0.109	0.239	0.112	0.240	0.111
$\rho=0.8$	0.382	0.104	0.449	0.172	0.473	0.119	0.448	0.129	0.451	0.131
$\rho=0.95$	0.625	0.187	0.743	0.234	0.797	0.105	0.782	0.150	0.787	0.159

S-J: J test based on the series LRV estimator and χ^2 approximation; S-J*: J* test based on the series LRV estimator and F approximation.

B-J: J test based on the Bartlett kernel LRV estimator and χ^2 approximation; B-J*: J* test based on the Bartlett kernel LRV estimator and F approximation.

P-J, P-J*, Q-J, Q-J*, D-J and D-J* are defined analogously.

Table 2: Empirical size of different J tests and J* tests for the VMA(1) case with nominal size α , degree of overidentification q , and sample size $T = 100$

	S-J	S-J*	B-J	B-J*	P-J	P-J*	Q-J	Q-J*	D-J	D-J*
$q = 1, \alpha = 5\%$										
$\rho=-0.8$	0.066	0.056	0.076	0.064	0.073	0.057	0.071	0.057	0.071	0.057
$\rho=-0.5$	0.063	0.053	0.071	0.059	0.068	0.053	0.067	0.053	0.066	0.054
$\rho=0.00$	0.053	0.048	0.055	0.049	0.057	0.048	0.056	0.048	0.055	0.048
$\rho=0.50$	0.067	0.057	0.075	0.063	0.074	0.058	0.071	0.058	0.070	0.058
$\rho=0.8$	0.069	0.058	0.079	0.065	0.074	0.059	0.074	0.059	0.073	0.059
$\rho=0.95$	0.061	0.055	0.067	0.058	0.065	0.055	0.065	0.055	0.065	0.055
$q = 1, \alpha = 10\%$										
$\rho=-0.8$	0.125	0.112	0.140	0.122	0.133	0.112	0.131	0.113	0.132	0.113
$\rho=-0.5$	0.121	0.109	0.136	0.119	0.131	0.109	0.129	0.110	0.128	0.110
$\rho=0.00$	0.102	0.097	0.107	0.098	0.108	0.097	0.106	0.097	0.105	0.096
$\rho=0.50$	0.123	0.112	0.135	0.122	0.130	0.112	0.129	0.113	0.129	0.113
$\rho=0.8$	0.126	0.113	0.140	0.124	0.134	0.113	0.132	0.114	0.132	0.114
$\rho=0.95$	0.116	0.109	0.125	0.115	0.123	0.109	0.122	0.109	0.121	0.110
$q = 4, \alpha = 5\%$										
$\rho=-0.8$	0.129	0.050	0.144	0.060	0.146	0.051	0.137	0.052	0.140	0.053
$\rho=-0.5$	0.118	0.048	0.131	0.058	0.135	0.051	0.126	0.051	0.130	0.051
$\rho=0.00$	0.084	0.043	0.094	0.044	0.096	0.044	0.091	0.044	0.091	0.044
$\rho=0.50$	0.122	0.048	0.136	0.058	0.137	0.051	0.127	0.051	0.132	0.051
$\rho=0.8$	0.130	0.047	0.146	0.057	0.149	0.048	0.139	0.049	0.141	0.049
$\rho=0.95$	0.104	0.047	0.115	0.052	0.118	0.048	0.112	0.048	0.113	0.049
$q = 4, \alpha = 10\%$										
$\rho=-0.8$	0.203	0.104	0.228	0.122	0.227	0.107	0.218	0.109	0.220	0.108
$\rho=-0.5$	0.190	0.101	0.213	0.114	0.214	0.103	0.204	0.103	0.205	0.103
$\rho=0.00$	0.152	0.090	0.162	0.094	0.165	0.093	0.158	0.093	0.160	0.092
$\rho=0.50$	0.197	0.102	0.216	0.117	0.218	0.105	0.207	0.107	0.211	0.108
$\rho=0.8$	0.203	0.101	0.229	0.120	0.229	0.104	0.217	0.108	0.221	0.106
$\rho=0.95$	0.174	0.099	0.190	0.109	0.192	0.102	0.183	0.103	0.186	0.102

S-J: J test based on the series LRV estimator and χ^2 approximation; S-J*: J* test based on the series LRV estimator and F approximation.

B-J: J test based on the Bartlett kernel LRV estimator and χ^2 approximation; B-J*: J* test based on the Bartlett kernel LRV estimator and F approximation.

P-J, P-J*, Q-J, Q-J*, D-J and D-J* are defined analogously.

Table 3: Empirical size of different J tests and J* tests for different DGPs with nominal size α , degree of overidentification 4, and sample size $T = 200$

	S-J	S-J*	B-J	B-J*	P-J	P-J*	Q-J	Q-J*	D-J	D-J*
$q = 4, \alpha = 5\%, \text{VAR}(1)$ case										
$\rho=-0.8$	0.185	0.048	0.221	0.082	0.218	0.054	0.203	0.056	0.205	0.057
$\rho=-0.5$	0.098	0.054	0.116	0.065	0.108	0.055	0.103	0.056	0.106	0.056
$\rho=0.00$	0.067	0.049	0.072	0.048	0.072	0.048	0.070	0.048	0.070	0.049
$\rho=0.50$	0.096	0.053	0.113	0.064	0.107	0.053	0.102	0.054	0.103	0.055
$\rho=0.8$	0.184	0.048	0.221	0.080	0.218	0.054	0.203	0.056	0.205	0.055
$\rho=0.95$	0.385	0.054	0.513	0.107	0.566	0.049	0.539	0.067	0.547	0.069
$q = 4, \alpha = 10\%, \text{VAR}(1)$ case										
$\rho=-0.8$	0.274	0.109	0.323	0.160	0.316	0.117	0.298	0.120	0.298	0.120
$\rho=-0.5$	0.168	0.112	0.190	0.130	0.180	0.113	0.176	0.114	0.179	0.114
$\rho=0.00$	0.127	0.099	0.133	0.100	0.134	0.099	0.130	0.099	0.131	0.099
$\rho=0.50$	0.170	0.112	0.193	0.127	0.183	0.112	0.178	0.113	0.179	0.113
$\rho=0.8$	0.278	0.106	0.325	0.157	0.317	0.115	0.300	0.118	0.301	0.117
$\rho=0.95$	0.498	0.126	0.620	0.214	0.660	0.118	0.635	0.143	0.638	0.148
$q = 4, \alpha = 5\%, \text{VMA}(1)$ case										
$\rho=-0.8$	0.089	0.051	0.103	0.061	0.098	0.052	0.093	0.053	0.094	0.052
$\rho=-0.5$	0.088	0.053	0.099	0.060	0.095	0.053	0.091	0.053	0.093	0.054
$\rho=0.00$	0.067	0.047	0.072	0.047	0.072	0.047	0.070	0.047	0.070	0.048
$\rho=0.50$	0.088	0.051	0.099	0.059	0.095	0.051	0.091	0.052	0.092	0.052
$\rho=0.8$	0.091	0.053	0.105	0.061	0.100	0.053	0.095	0.054	0.097	0.054
$\rho=0.95$	0.077	0.050	0.085	0.054	0.083	0.050	0.081	0.050	0.080	0.051
$q = 4, \alpha = 10\%, \text{VMA}(1)$ case										
$\rho=-0.8$	0.152	0.104	0.172	0.119	0.165	0.106	0.160	0.106	0.161	0.106
$\rho=-0.5$	0.151	0.107	0.169	0.119	0.163	0.108	0.157	0.108	0.159	0.108
$\rho=0.00$	0.126	0.098	0.133	0.099	0.133	0.098	0.130	0.098	0.131	0.099
$\rho=0.50$	0.151	0.106	0.168	0.118	0.162	0.107	0.157	0.108	0.159	0.107
$\rho=0.8$	0.157	0.108	0.176	0.121	0.169	0.108	0.163	0.109	0.164	0.109
$\rho=0.95$	0.136	0.102	0.149	0.108	0.146	0.103	0.143	0.103	0.142	0.103

S-J: J test based on the series LRV estimator and χ^2 approximation; S-J*: J* test based on the series LRV estimator and F approximation.

B-J: J test based on the Bartlett kernel LRV estimator and χ^2 approximation; B-J*: J* test based on the Bartlett kernel LRV estimator and F approximation.

P-J, P-J*, Q-J, Q-J*, D-J and D-J* are defined analogously.

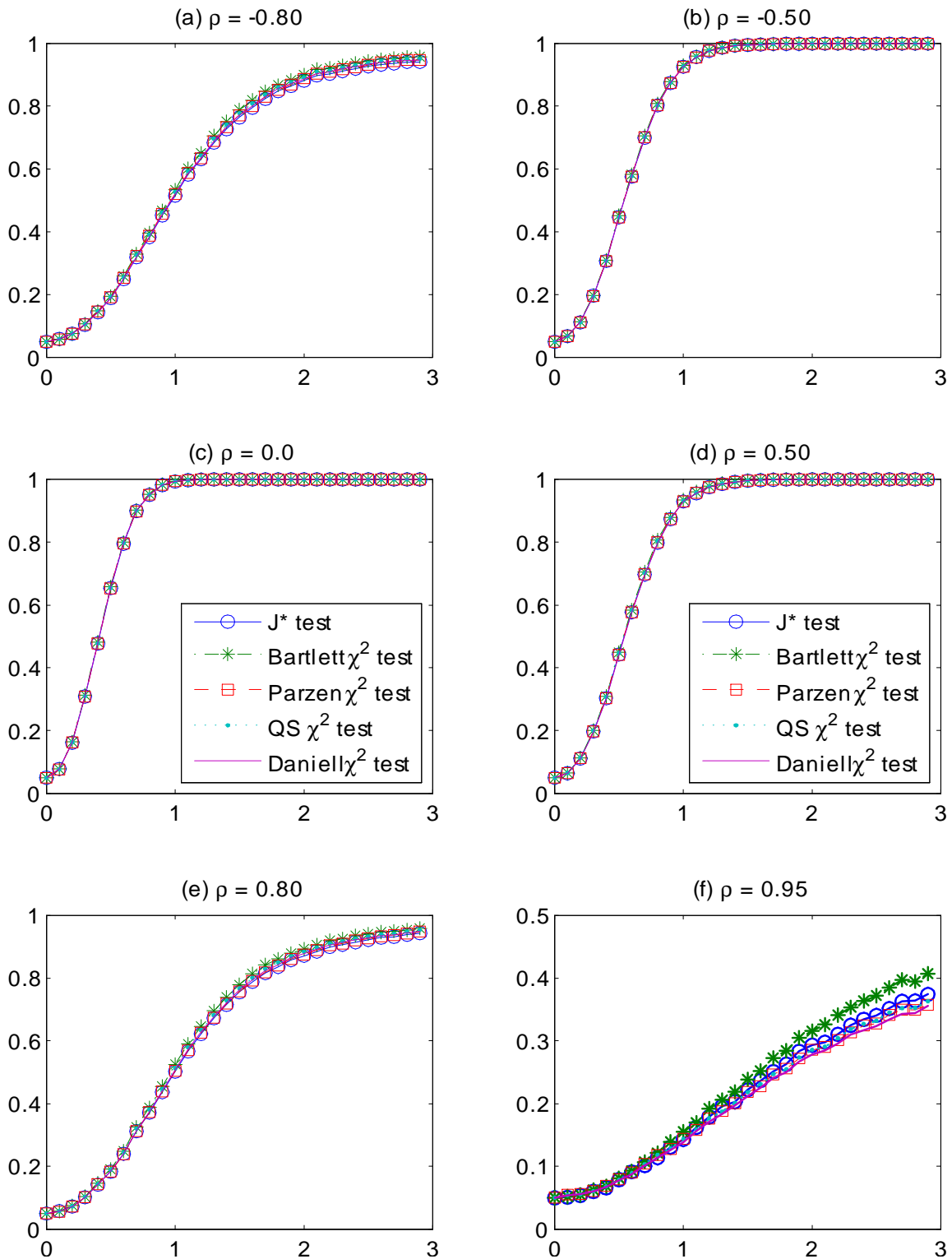


Figure 1: Size adjusted power of the series J^* test and kernel J tests under the VAR(1) case with sample size $T = 100$ and degree of overidentification $q = 1$.

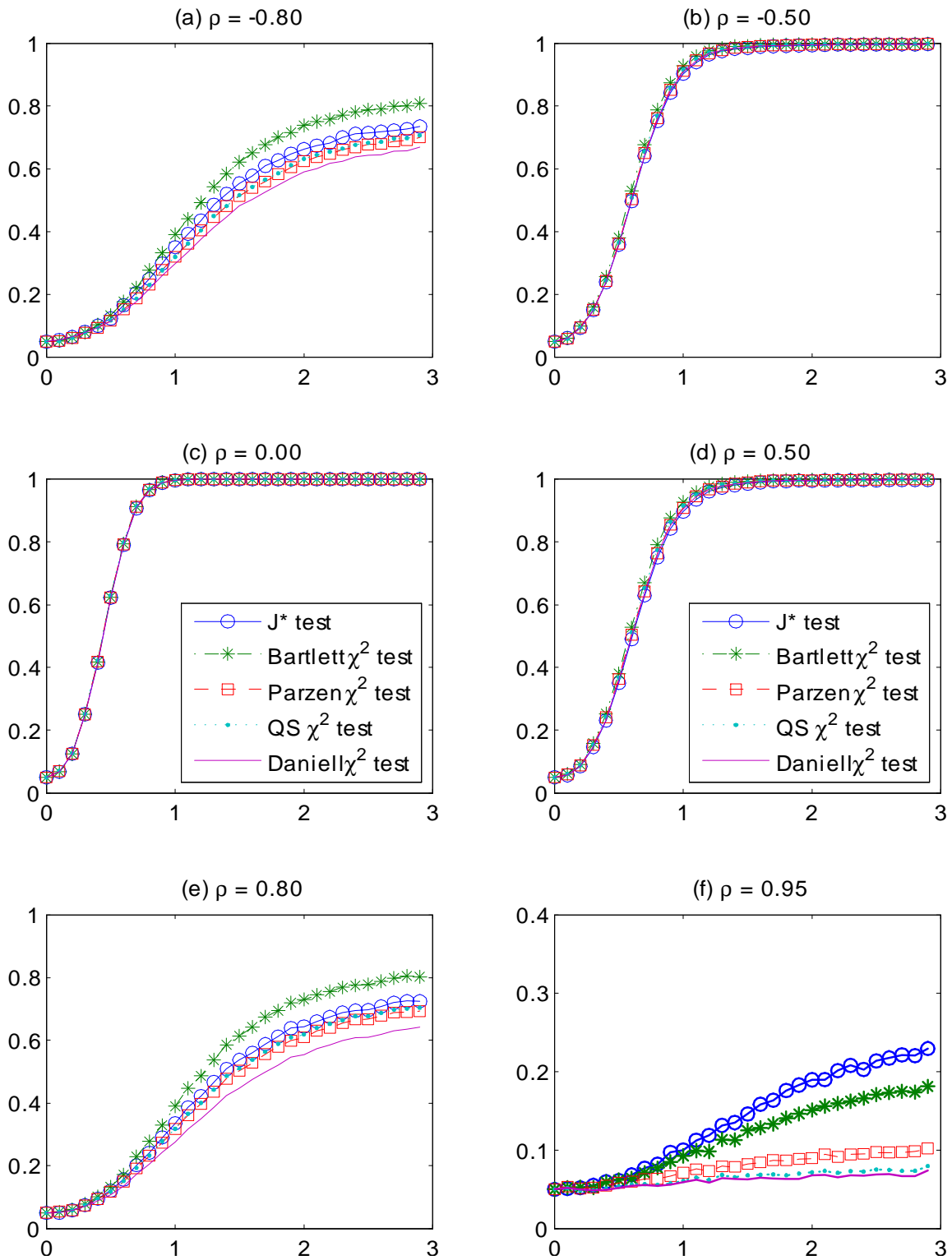


Figure 2: Size adjusted power of the series J^* test and kernel J tests under the VAR(1) case with sample size $T = 100$ and degree of overidentification $q = 4$.

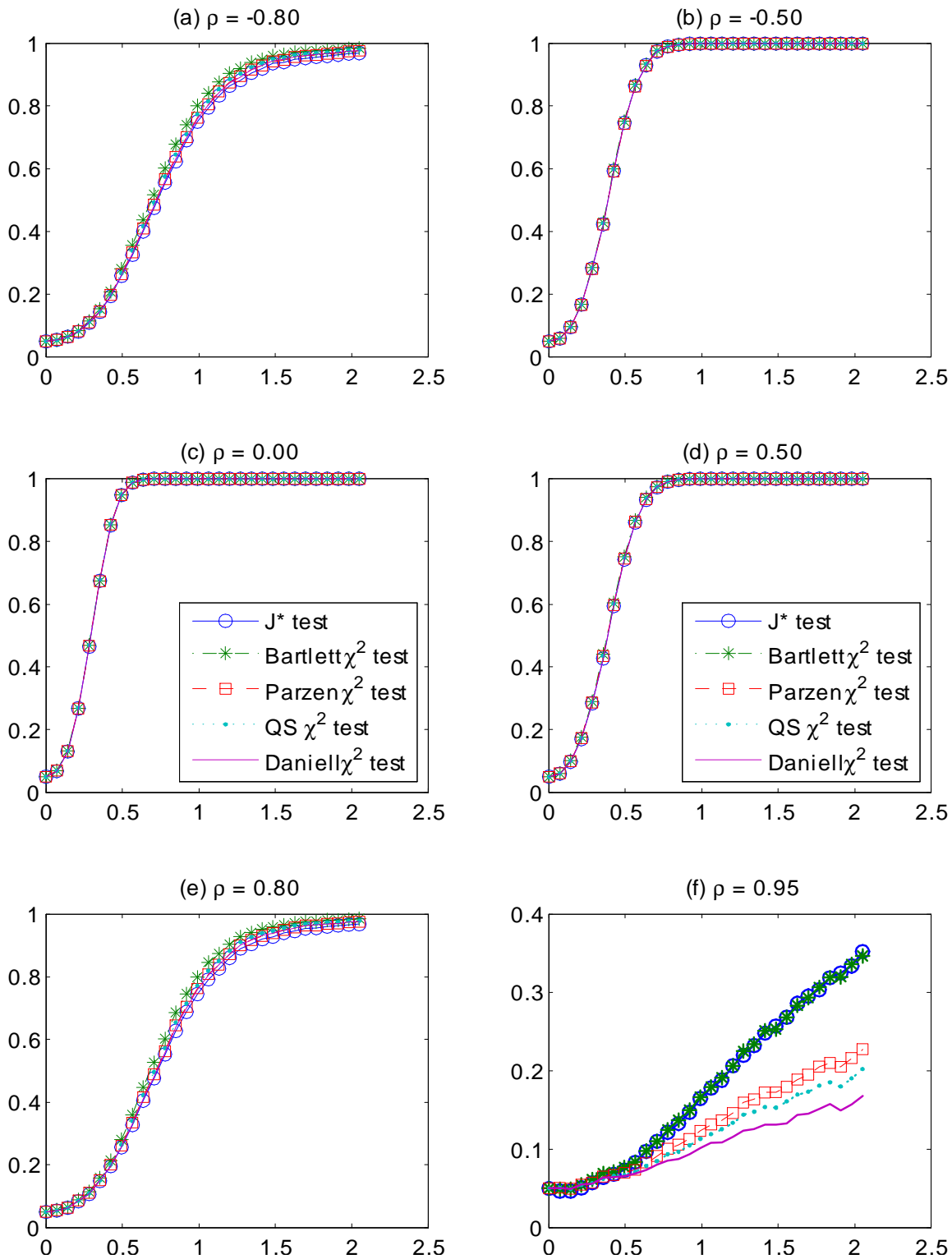


Figure 3: Size adjusted power of the series J^* test and kernel J tests under the VAR(1) case with sample size $T = 200$ and degree of overidentification $q = 4$.