UC San Diego
Recent Work

Title

Fixed-smoothing Asymptotics and Asymptotic F and t Tests in the Presence of Strong Autocorrelation

Permalink

https://escholarship.org/uc/item/8479f4s2

Author

Sun, Yixiao

Publication Date

2014-05-27

Peer reviewed

Fixed-smoothing Asymptotics and Asymptotic F and t Tests in the Presence of Strong Autocorrelation

Yixiao Sun[*]
Department of Economics,
University of California, San Diego

May 26, 2014

Abstract

New asymptotic approximations are established for the Wald and t statistics in the presence of unknown but strong autocorrelation. The asymptotic theory extends the usual fixed-smoothing asymptotics under weak dependence to allow for near unit root and weak unit root processes. As the locality parameter that characterizes the neighborhood of the autoregressive root increases from zero to infinity, the new fixed-smoothing asymptotic distribution changes smoothly from the unit-root fixed-smoothing asymptotics to the usual fixed-smoothing asymptotics under weak dependence. Simulations show that the new approximation is more accurate than the usual fixed-smoothing approximation.

1 Introduction

In order to robustify statistical inferences to heteroskedasticity and autocorrelation, it is now a standard practice to make use of heteroskedasticity and autocorrelation robust (HAR) standard errors in time series analysis. The last decade of research on this topic has been largely focused on more accurate approximations to the test statistics constructed on the basis of HAR standard errors and variances. There are now two types of asymptotic approximations: the conventional increasing-smoothing asymptotic approximation and the relatively new fixed-smoothing asymptotic approximation. For kernel HAR inference, the fixed-smoothing asymptotics is the so-called fixed-b asymptotics of Kiefer, Vogelsang and Bunzel (2000, KVB hereafter) and Kiefer and Vogelsang (2002a, 2002b, 2005, KV hereafter). For series HAR inference, the fixed-smoothing asymptotics is the so-called fixed-K asymptotics of Sun (2011, 2013a). While the fixed-smoothing asymptotic approximation is more accurate than the conventional increasing-smoothing asymptotics approximation, i.e. the chi-square approximation (Jansson, 2004; Sun, Phillips and Jin, 2008), the quality of the fixed-smoothing approximation is not completely satisfactory when the underlying time series has strong autocorrelation. To confront with this problem, the paper develops a new fixed-smoothing asymptotics that accommodates strong autocorrelation.

We start by considering a multivariate time series whose mean value is the parameter of interest. Each component of the time series follows an autoregressive process with the autoregressive parameter ρ changing with the sample size T. More specifically, $\rho = 1 - c_m/T$ for some positive sequence c_m. Depending on the limiting thought experiments employed, c_m may be held fixed as the sample size $T \to \infty$ or grow with the sample size. The former case is the conventional local-to-unit-root specification and the latter case specifies a moderate deviation from the unit root. Park (2003) refers to these two cases as "near unit root" and "weak unit root," respectively. For more discussions, see Giraitis and Phillips (2006), Phillips and Magdalinos (2007), and Phillips, Magdalinos and Giraitis (2010).

We establish the fixed-smoothing asymptotics for the Wald and t statistics when c_m is held fixed. This leads to our pivotal near-unit-root fixed-smoothing asymptotics. As $c_m \to \infty$, we show that the near-unit-root fixed-smoothing asymptotics converges to the fixed-smoothing asymptotics under weak dependence. Depending on the value of c_m, the near-unit-root fixed-smoothing asymptotics thus provides a smooth transition from the usual stationary fixed-smoothing asymptotics to the unit-root fixed-smoothing asymptotics.

The near-unit-root fixed-smoothing asymptotic distribution is nonstandard but can be simulated. The critical values from this distribution are larger than the corresponding ones from the usual fixed-smoothing asymptotic distribution, which in turn are larger than those

1

from the conventional chi-square distribution. A direct implication is that statistical inference based on the chi-square approximation and stationary fixed-smoothing approximation may lead to the finding of statistical significance that does not actually exist.

In the case of series variance estimation, we can judiciously design a set of basis functions such that the near-unit-root fixed-smoothing asymptotic distribution becomes a standard F or t distribution. The designing process involves projection and orthogonalization. The F and t approximations are very handy in empirical applications, as the F and t critical values are readily available from statistical programs and software packages. There is no need to simulate nonstandard critical values. The possibility of deriving a standard approximation under the fixed-smoothing asymptotics is an advantage of using series variance estimators. For this type of estimators, we have the complete freedom in choosing the basis functions. This is in contrast with the kernel variance estimators where the basis functions are implicitly given and cannot be changed.

Monte Carlo simulations show that the chi-square tests have the largest size distortion, the tests based on the near-unit-root fixed-smoothing asymptotics have the smallest size distortion, and the tests based on the usual fixed-smoothing asymptotics are in the middle. The near-unit-root fixed-smoothing tests, which include tests based on F and t approximation, achieves triple robustness in the following sense: it is asymptotically valid regardless of whether autocorrelation is present or not; whether the autocorrelation is strong or not and whether the level of smoothing is held fixed or is allowed to increase with the sample size.

The rest of the paper is organized as follows. Section 2 describes the basic setting and the problem at hand. Section 3 develops the fixed-smoothing asymptotics in the presence of near-unit roots. This section also establishes the behavior of the near-unit-root fixed-smoothing asymptotic distribution as $c_m \to \infty$. Section 4 presents the F and t approximations based on a set of judiciously designed basis functions. Section 5 discusses the applicability of our fixed-smoothing approximation for location models to more general models. The subsequent section reports simulation evidence, and the last section provides some concluding discussions. Proofs of the main results are given in the Appendix.

2   The Basic Setting and the Problem

Assume that p-dimensional time series y_t follows the process:

$$y_t = \theta + e_t, t = 1, 2, \ldots, T \tag{1}$$

2

where $y_t = (y_{1t}, \ldots, y_{pt})'$, $\theta = (\theta_1, \ldots, \theta_p)'$, and $e_t = (e_{1t}, \ldots, e_{pt})'$ is a zero mean process. The location model, as simple as it is, is empirically relevant in a number of situations. For example, the data might consist of a multivariate time series of forecasting loss that are produced by different forecasting methods. We can test equal predictive accuracy of these forecasting methods by examining whether the loss differential series has mean zero. There is also a large and active literature on inference for the mean of simulated time series. See for example Alexopoulos (2006, 2007) and references therein. More importantly, our points can be made more clearly in the simple location model.

We are interested in testing the null $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta \neq \theta_0$. The OLS estimator of θ is the average of $\{y_t\}$, i.e., $\hat{\theta} = \bar{y} := T^{-1} \sum_{t=1}^{T} y_t$. The F test version of the Wald statistic based on the OLS estimator is then given by

$$F_T = D_T \left(\hat{\theta} - \theta_0 \right)' \hat{\Omega}^{-1} D_T \left(\hat{\theta} - \theta_0 \right) / p = D_T^2 \left(\hat{\theta} - \theta_0 \right)' \hat{\Omega}^{-1} \left(\hat{\theta} - \theta_0 \right) / p$$

where D_T is a real-valued scaling factor that characterizes the rate of convergence of $\hat{\theta}$ to θ_0, and $\hat{\Omega}$ is an estimator of the asymptotic variance of $D_T(\hat{\theta} - \theta_0)$. Usually $D_T = \sqrt{T}$ but the exact magnitude of D_T is not important in practice, as it will be canceled out. When $p = 1$, we can construct the t-statistic as follows:

$$t_T = \frac{D_T \left(\hat{\theta} - \theta_0 \right)}{\sqrt{\hat{\Omega}}}.$$

Many nonparametric estimators of the asymptotic variance matrix are available in the literature. In this paper, we consider a class of quadratic variance estimators, which includes the conventional kernel variance estimators of Andrews (1991), Newey and West (1987), Politis (2011), and the series variance estimators of Phillips (2005), Müller (2007), and Sun (2006, 2011, 2013a) as special cases. The quadratic variance estimators take the following form:

$$\hat{\Omega} = \frac{D_T^2}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h \left(\frac{t}{T}, \frac{s}{T} \right) \hat{e}_t \hat{e}_s' \tag{2}$$

where $\hat{e}_t = y_t - \hat{\theta} = e_t - \bar{e}$ for $\bar{e} = T^{-1} \sum_{t=1}^{T} e_t$ and $Q_h(r, s)$ is a weighting function that depends on the smoothing parameter h. A reasonable choice of $Q_h(r, s)$ should satisfy: (i) $Q_h(\frac{t}{T}, \frac{s}{T})$ decays to zero as $|t - s|/T$ approaches 1; (ii) $Q_h(\frac{t}{T}, \frac{s}{T})$ increases to 1 as $|t - s|/T$ approaches 0. The speed of change is controlled by h. For conventional kernel variance estimators, $Q_h(r, s) = k((r - s)/b)$ and we take $h = 1/b$, where $k(\cdot)$ is a kernel function. For the series variance estimators, $Q_h(r, s) = K^{-1} \sum_{j=1}^{K} \phi_j(r) \phi_j(s)$ and we take $h = K$, where $\{\phi_j(r)\}$ are basis functions on $\mathbb{L}^2[0, 1]$ satisfying $\int_0^1 \phi_j(r) \, dr = 0$. We parametrize h in such a way that h indicates the level or amount of smoothing in both cases.

Define

$$Q^*_{T,h}(r,s) = Q_h(r,s) - \frac{1}{T}\sum_{\tau_1=1}^{T} Q_h\left(\frac{\tau_1}{T},s\right) - \frac{1}{T}\sum_{\tau_2=1}^{T} Q_h\left(r,\frac{\tau_2}{T}\right) + \frac{1}{T}\sum_{\tau_1=1}^{T}\sum_{\tau_2=1}^{T} Q_h\left(\frac{\tau_1}{T},\frac{\tau_2}{T}\right),$$

then

$$\hat{\Omega} = \frac{D_T^2}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} Q^*_{T,h}\left(\frac{t}{T},\frac{s}{T}\right) e_t e_s', \tag{3}$$

where the demeaning operation on e_t and e_s has been moved to the weighting function. The Wald statistic is then equal to

$$F_T = \left(\sum_{t=1}^{T} e_t\right)'\left[\sum_{t=1}^{T}\sum_{s=1}^{T} Q^*_{Th}\left(\frac{t}{T},\frac{s}{T}\right) e_t e_s'\right]^{-1}\left(\sum_{t=1}^{T} e_t\right)/p.$$

Similarly, the t-statistic becomes

$$t_T = \frac{\sum_{t=1}^{T} e_t}{\left[\sum_{t=1}^{T}\sum_{s=1}^{T} Q^*_{Th}\left(\frac{t}{T},\frac{s}{T}\right) e_t e_s\right]^{1/2}}.$$

Note that the scaling factor D_T has been canceled out in both F_T and t_T.

The question is how to approximate the sampling distributions of F_T and t_T. If $T^{-1/2}\sum_{t=1}^{[Tr]} e_t$ converges weakly to a Brownian motion process, then under some conditions on Q_h, it can be shown that, for a fixed h :

$$F_T \to^d F_\infty := W_p(1)'\left[\int_0^1\int_0^1 Q^*_h(r,s)\,dW_p(r)\,dW_p'(s)\right]^{-1} W_p(1)/p, \tag{4}$$

$$t_T \to^d t_\infty := \frac{W_p(1)}{\sqrt{\int_0^1\int_0^1 Q^*_h(r,s)\,dW_p(r)\,dW_p'(s)}}, \tag{5}$$

where $W_p(r)$ is a $p \times 1$ vector of standard Wiener processes and

$$Q^*_h(r,s) = Q_h(r,s) - \int_0^1 Q_h(\tau_1,s)\,d\tau_1 - \int_0^1 Q_h(r,\tau_2)\,d\tau_2 + \int_0^1\int_0^1 Q_h(\tau_1,\tau_2)\,d\tau_1 d\tau_2$$

is the 'continuous' version of $Q^*_{T,h}(r,s)$. See for example, Kiefer and Vogelsang (2005) for the kernel case and Sun (2013a) for the series case.

For both variance estimators, $W_p(1)$ is independent of $\int_0^1\int_0^1 Q^*_h(r,s)\,dW_p(r)\,dW_p'(s)$. So F_∞ resembles an F distribution in the sense that both can be written as a quadratic form in standard normals with an independent weighting matrix. Similarly, t_∞ resembles a t distribution in the sense that both can be written as a standard normal scaled by an independent random variable.

4

The asymptotic approximation obtained under a fixed h as $T \to \infty$ is the fixed-smoothing asymptotic approximation. This approximation improves the traditional chi-square or normal approximation, which is obtained under the increasing-smoothing asymptotics where h increases with T but at a slower rate. However, the fixed-smoothing asymptotic approximation is still not satisfactory when the underlying process has strong autocorrelation. Our goal is to establish a further improved approximation when e_t may be strongly autocorrelated.

3 Fixed-Smoothing Asymptotics under the Local-to-Unity Specification

To model strong autocorrelation, we maintain the following assumption on $\{e_t\}$.

Assumption 1 *(i) For some positive sequence* $\{c_m\}$

$$e_t = \rho_{T,m} e_{t-1} + u_t \ \text{where } e_0 = O_p(1) \ \text{and } \rho_{T,m} = 1 - \frac{c_m}{T};$$

(ii) $T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \|E u_t u_s'\|$ *is bounded uniformly in* T*;*
(iii) $\{u_t\}$ *satisfies a FCLT:*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t \to^d \Lambda W_p(r), \ \text{as } T \to \infty, \tag{6}$$

where $W_p(r)$ *is a* $p \times 1$ *vector of standard Wiener processes and* $\Lambda = \Omega^{1/2}$ *is the matrix square root of the long run variance matrix* Ω *of* u_t *:*

$$\Omega = \Lambda\Lambda' = \sum_{j=-\infty}^{\infty} E u_t u_{t-j}'.$$

The data generating process in Assumption 1 is similar to that used in Phillips, Magdalinous and Giraitis (2010), which establishes a smooth transition between the conventional unit root distribution and the standard norm distribution. When c_m is fixed, each component of $\{e_t\}$ has a local-to-unity root in the conventional sense (Phillips, 1987). When $c_m \to \infty$ as $T \to \infty$, each component of $\{e_t\}$ has a moderate unit root in the sense that the root belongs to a larger neighborhood of unity than the conventional local-to-unity roots. We could allow for different c_m's for different components of $\{e_t\}$. For notational simplicity, we assume that all components have the same local-to-moderate unit root. The FCLT assumption holds for serially correlated and heterogeneously distributed data that satisfy certain regularity conditions on moments and the dependence structure over time.

These primitive regularity conditions can be found in Phillips and Durlauf (1986), Phillips and Solo (1992), Davidson (1994), among others.

Under Assumption 1, we have

$$\frac{1}{\sqrt{T}} e_{[Tr]} \to \Lambda J_{c_m}(r)$$

for each fixed c_m, where $J_{c_m}(r)$ is the Ornstein-Uhlenbeck (OU) process defined by

$$dJ_{c_m}(r) = -c_m J_{c_m}(r)\, dr + dW_p(r)$$

with $J_{c_m}(0) = 0$ or $J_{c_m}(r) = \int_0^r e^{-c_m(r-s)} dW_p(s)$.

Assumption 2 (i) *For kernel variance estimators, the kernel function $k(\cdot)$ satisfies: for any $b \in (0,1]$, $k_b(x) := k(x/b)$ is symmetric, continuous, piecewise monotonic, and piecewise continuously differentiable on $[-1,1]$. (ii) For the series variance estimator, $\{\phi_j(\cdot)\}$ are piecewise monotonic, continuously differentiable, and orthonormal in $\mathbb{L}^2[0,1]$ and $\int_0^1 \phi_j(x)\, dx = 0$ for each j.*

Assumption 2 is similar to Assumption 1 in Sun (2013b). Under this assumption, we can replace $Q^*_{Th}(r,s)$ in (3) by $Q^*_h(r,s)$ and all our asymptotic results continue to hold. For our theoretical development, we can assume that such a replacement has been made. Furthermore, under Assumption 2, $Q^*_h(r,s)$ has the following unified representation for both types of variance estimators we consider:

$$Q^*_h(r,s) = \sum_{j=1}^{\infty} \lambda_j \Phi_j(r)\Phi_j(s), \tag{7}$$

where $\{\Phi_j(r)\}$ is a sequence of continuously differentiable functions satisfying $\int_0^1 \Phi_j(r)\, dr = 0$. The right hand side of (7) converges absolutely and uniformly over $(r,s) \in [0,1] \times [0,1]$.

The representation holds trivially for the case of series variance estimation, as

$$Q^*_h(r,s) = Q_h(r,s) = \frac{1}{K}\sum_{j=1}^{K} \phi_j(r)\phi_j(s),$$

and so we can take $\{\lambda_j\} = \{K^{-1}, \ldots, K^{-1}, 0, \ldots, 0, \ldots\}$ and $\Phi_j = \phi_j$. For the case of kernel variance estimation, Sun (2013b) proves that the representation holds with the following choices of Φ_j and λ_j:

$$\Phi_j(r) = \begin{cases} \cos\left(\frac{1}{2}\pi j r\right), & j \text{ is even} \\ \sin\left(\frac{1}{2}\pi(j+1)r\right) - \int_0^1 \sin\left(\frac{1}{2}\pi(j+1)\tau\right) d\tau, & j \text{ is odd} \end{cases} \tag{8}$$

and

$$\lambda_j = \begin{cases} \tilde{\lambda}_{j/2}, & j \text{ is even} \\ \tilde{\lambda}_{(j+1)/2}, & j \text{ is odd} \end{cases}$$

where $\{\tilde{\lambda}_j\}$ are the Fourier transforms of $k_b(\cdot)$, i.e., $\tilde{\lambda}_j = \int_0^1 k(x/b) \cos(j\pi x) dx$. For easy reference, we define $\Phi_0(r) \equiv 1$.

Theorem 1 *Let Assumptions 1 and 2 hold. Then for fixed c_m and h,*

$$F_T \to^d F_\infty(c_m) := \left[\int_0^1 J_{c_m}(r) dr \right]'$$

$$\times \left[\int_0^1 \int_0^1 Q_h^*(r,s) J_{c_m}(r) J'_{c_m}(s) drds \right]^{-1} \left[\int_0^1 J_{c_m}(r) dr \right] / p.$$

If we further assume that $Q_h(r,s)$ is positive definite, then

$$t_T \to^d t_\infty(c_m) := \frac{\int_0^1 J_{c_m}(r) dr}{\left[\int_0^1 \int_0^1 Q_h^*(r,s) J_{c_m}(r) J'_{c_m}(s) drds \right]^{1/2}}.$$

Since

$$cov(J_{c_m}(r), J_{c_m}(s)) = \frac{1}{2c_m} \{\exp[-c_m |r-s|] - \exp[-c_m(r+s)]\} I_p,$$

where I_p is the $p \times p$ identity matrix, we have

$$cov\left(\int_0^1 \Phi_{j_1}(r) J_{c_m}(r) dr, \int_0^1 \Phi_{j_2}(r) J_{c_m}(r) dr \right)$$

$$= \frac{1}{2c_m} \int_0^1 \int_0^1 \Phi_{j_1}(r) \Phi_{j_2}(s) \{\exp[-c_m |r-s|] - \exp[-c_m(r+s)]\} drds \times I_p$$

for all $j_1 \geq 0$ and $j_2 \geq 0$. In general, the covariance matrix is not zero. A direct implication is that the weighting matrix $\int_0^1 \int_0^1 Q_h^*(r,s) J_{c_m}(r) J'_{c_m}(s) drds$ is not independent of $\int_0^1 J_{c_m}(r) dr$. Hence the limiting distribution $F_\infty(c_m)$ cannot be written as a quadratic form in standard normals with an independent random weighting matrix. This is in contrast with F_∞ which takes such a form.

The limiting distribution depends not only on the kernel/basis functions, the smoothing parameter h, the number of joint hypotheses p but also on the local-to-unit-root parameter c_m. In the extreme case when $c_m = 0$, e_t has unit roots and $F_\infty(c_m)$ becomes

$$F_\infty(0) := \left[\int_0^1 W_p(r) dr \right]' \left[\int_0^1 \int_0^1 Q_h^*(r,s) W_p(r) W'_p(s) drds \right]^{-1} \left[\int_0^1 W_p(r) dr \right] / p.$$

This case seems to be of only theoretical interest. The empirically more relevant case is when c_m is large. The theorem below characterizes the behaviors of $F_\infty(c_m)$ and $t_\infty(c_m)$ as $c_m \to \infty$.
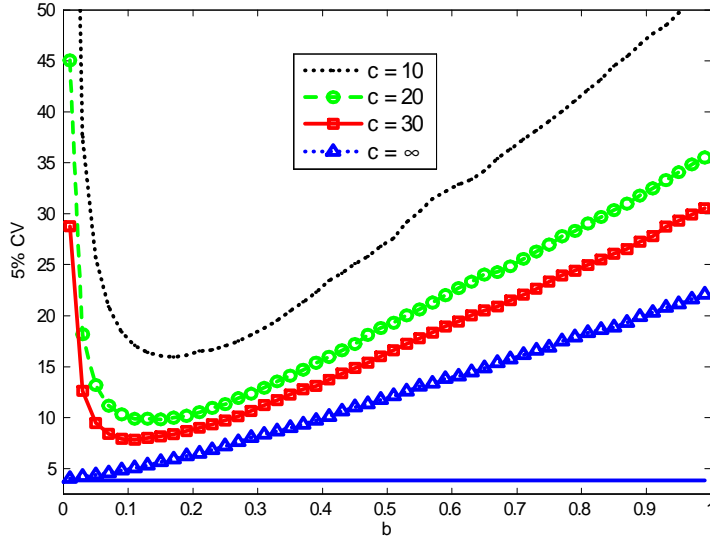
7

Figure 1: Critical values from $F_\infty(c)$ with $p = 1$ and the Bartlett kernel

Theorem 2 *As $c_m \to \infty$ for a fixed h, we have*

$$F_\infty(c_m) = F_\infty(\infty) + o_p(1) \ \ and \ t_\infty(c_m) = t_\infty(\infty) + o_p(1)$$

where $F_\infty(\infty) := F_\infty$ and $t_\infty(\infty) := t_\infty$ defined in (4) and (5) respectively.

Theorem 2 shows that the sequential limits when $T \to \infty$ followed by $c_m \to \infty$ lead to the usual fixed-smoothing asymptotics under weak dependence. Depending on the value of c_m, the limiting distributions $F_\infty(c_m)$ and $t_\infty(c_m)$ provide a smooth transition from the usual fixed-smoothing asymptotics to the unit-root fixed smoothing asymptotics.

Figures 1–3 present the 5% critical values from $F_\infty(c_m)$ with $p = 1$ for two kernel variance estimators and one series variance estimator. For the series variance estimator, we take K to be even and use $\phi_{2j-1}(x) = \sqrt{2}\cos(2j\pi x)$, $\phi_{2j}(x) = \sqrt{2}\sin(2j\pi x)$, $j = 1, \ldots, K/2$ as the basis functions. In the figure, '$c = \infty$' corresponds to the usual fixed-smoothing asymptotics, and the solid line indicates the critical value from the χ_1^2 distribution. For any given variance estimator, the critical values increase monotonically as c_m approaches zero. That is, the more persistent the underlying processes are, the higher the critical values will be. While the critical values from the usual fixed-smoothing asymptotic distribution are larger than those from the chi-square distribution, they are still not large enough when the process is highly autocorrelated. Figures 1–3 also demonstrate the smooth transition from the near-unit-root fixed-smoothing asymptotics to the usual fixed-smoothing asymptotics.
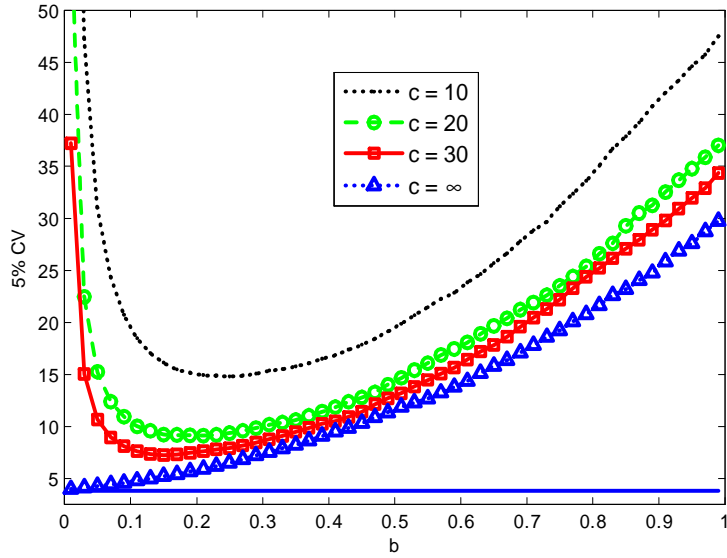
8

Figure 2: Critical values from $F_\infty(c)$ with $p = 1$ and the Parzen kernel
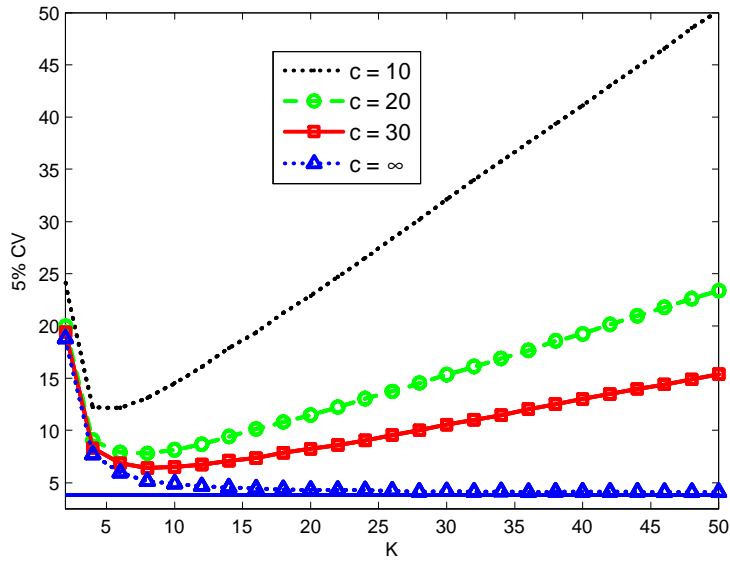

Figure 3: Critical values from $F_\infty(c)$ with $p = 1$ and the series LRV estimator

9

4 F Approximation in the Case of Series Variance Estimation

In this section, we consider the series variance estimation and show that $F_\infty(c_m)$ can be approximated by an F distribution for some (transformed) basis functions.

In the case of series variance estimation, we have

$$pF_\infty(c_m) = \left[\int_0^1 \phi_0(r) J_{c_m}(r)\, dr\right]'$$

$$\times \left\{\frac{1}{K}\sum_{\ell=1}^K \left[\int_0^1 \phi_\ell(r) J_{c_m}(r)\, dr\right]\left[\int_0^1 \phi_\ell(r) J_{c_m}(r)\, dr\right]'\right\}^{-1}\left[\int_0^1 \phi_0(r) J_{c_m}(r)\, dr\right]/p$$

for $\phi_0(r) \equiv 1$. We want to select the basis functions such that $\int_0^1 \phi_\ell(r) J_{c_m}(r)\, dr \sim iidN(0, I_p)$ across $\ell = 0, 1, \ldots, K$. Note that $\int_0^1 \phi_\ell(r) J_{c_m}(r)\, dr$ for $\ell = 0, 1, \ldots, K$ are jointly normal, it suffices to select $\{\phi_\ell(r), \ell = 1, \ldots, K\}$ such that $\{\int_0^1 \phi_\ell(r) J_{c_m}(r)\, dr, \ell = 0, 1, \ldots, K\}$ are not correlated with each other. But

$$cov\left(\int_0^1 \phi_{\ell_1}(r) J_{c_m}(r)\, dr, \int_0^1 \phi_{\ell_2}(r) J_{c_m}(r)\, dr\right)$$

$$= \int_0^1 \int_0^1 \phi_{\ell_1}(r)\phi_{\ell_2}(s)\kappa_{c_m}(r,s)drds \times I_p$$

where

$$\kappa_{c_m}(r,s) := \{\exp[-c_m|r-s|] - \exp[-c_m(r+s)]\}/(2c_m)$$

is the covariance kernel of $J_{c_m}(\cdot)$. When $c_m = 0$, we define

$$\kappa_{c_m}(r,s)|_{c_m=0} = \lim_{c_m \to 0}\kappa_{c_m}(r,s) = \min(r,s),$$

which is the covariance kernel of the standard Brownian motion. So it suffices to select the basis functions to satisfy:

$$\int_0^1 \int_0^1 \phi_{\ell_1}(r)\phi_{\ell_2}(s)\kappa_{c_m}(r,s)drds = \delta_{\ell_1\ell_2} \text{ for all } \ell_1, \ell_2 = 1, \ldots, K; \tag{9}$$

$$\int_0^1 \int_0^1 \phi_\ell \kappa_{c_m}(r,s)drds = 0 \text{ for all } \ell = 1, 2, \ldots, K; \tag{10}$$

$$\int_0^1 \phi_\ell(r)\, dr = 0 \text{ for all } \ell = 1, 2, \ldots, K. \tag{11}$$

The last set of equations maintains the 'zero mean' conditions.

Instead of directly searching for the appropriate basis functions, for computational reasons we consider the discrete analogue of (9)–(11). Let $A = (a_{ij})$ be the $T \times T$ matrix

10

with elements $a_{ij} = \kappa_{c_m}(i/T, j/T)$. By definition, A is a positive definite symmetric matrix. For any two vectors $\ell_1, \ell_2 \in \mathbb{R}^T$, we define the inner product:

$$\langle \ell_1, \ell_2 \rangle = \ell_1' A \ell_2 / T^2. \tag{12}$$

Then $\mathbb{R}^T$ is a Hilbert space with the above weighted inner product. Let

$$\boldsymbol{\phi}_\ell = [\phi_\ell(1/T), \phi_\ell(2/T), \ldots, \phi_\ell(T/T)]'$$

be the basis vector associated with the basis function $\phi_\ell(\cdot)$ for $\ell = 0, 1, \ldots, K$. The finite sample versions of (9)–(11) are:

$$\langle \boldsymbol{\phi}_{\ell_1}, \boldsymbol{\phi}_{\ell_2} \rangle = \delta_{\ell_1 \ell_2} \text{ for all } \ell_1, \ell_2 = 1, \ldots, K; \tag{13}$$

$$\boldsymbol{\phi}_\ell' A \boldsymbol{\phi}_0 = 0 \text{ for all } \ell = 1, 2, \ldots, K; \tag{14}$$

$$\boldsymbol{\phi}_\ell' \boldsymbol{\phi}_0 = 0 \text{ for all } \ell = 1, 2, \ldots, K. \tag{15}$$

Note that (13) is different from the usual orthonormality in the Euclidian sense. In general, the basis vectors $\{\boldsymbol{\phi}_\ell\}$ do not satisfy (13) even if the basis functions $\{\phi_\ell\}$ are orthonormal in $\mathbb{L}^2[0,1]$. However, given any candidate basis functions or vectors, we can make them satisfy the above conditions via transformation and orthogonalization. We follow the steps below:

(i) Let $V_T = [\boldsymbol{\phi}_0, A\boldsymbol{\phi}_0]$ be a $T \times 2$ matrix and compute

$$\tilde{\boldsymbol{\phi}}_\ell = \left[I_T - V_T \left(V_T' V_T \right)^{-1} V_T' \right] \boldsymbol{\phi}_\ell$$

for $\ell = 1, \ldots, K$.

(ii) Employ the Gram-Schmidt scheme to orthogonalize the vectors $\{\tilde{\boldsymbol{\phi}}_\ell\}_{\ell=1}^K$ under the inner product $\langle \cdot, \cdot \rangle$. More specifically, we let

$$\tilde{\boldsymbol{q}}_1 = \tilde{\boldsymbol{\phi}}_1,$$
$$\tilde{\boldsymbol{q}}_2 = \tilde{\boldsymbol{\phi}}_2 - \frac{\langle \tilde{\boldsymbol{\phi}}_2, \tilde{\boldsymbol{q}}_1 \rangle}{\langle \tilde{\boldsymbol{q}}_1, \tilde{\boldsymbol{q}}_1 \rangle} \tilde{\boldsymbol{q}}_1,$$
$$\ldots$$
$$\tilde{\boldsymbol{q}}_K = \tilde{\boldsymbol{\phi}}_K - \frac{\langle \tilde{\boldsymbol{\phi}}_K, \tilde{\boldsymbol{q}}_{K-1} \rangle}{\langle \tilde{\boldsymbol{q}}_{K-1}, \tilde{\boldsymbol{q}}_{K-1} \rangle} \tilde{\boldsymbol{q}}_{K-1} - \ldots - \frac{\langle \tilde{\boldsymbol{\phi}}_K, \tilde{\boldsymbol{q}}_1 \rangle}{\langle \tilde{\boldsymbol{q}}_1, \tilde{\boldsymbol{q}}_1 \rangle} \tilde{\boldsymbol{q}}_1.$$

Note that $\tilde{\boldsymbol{\phi}}_\ell$ is the projection of $\boldsymbol{\phi}_\ell$ onto the orthogonal complement of the space spanned by $\boldsymbol{\phi}_0$ and $A\boldsymbol{\phi}_0$. By construction $\tilde{\boldsymbol{\phi}}_\ell' A \boldsymbol{\phi}_0 = 0$ and $\tilde{\boldsymbol{\phi}}_\ell' \boldsymbol{\phi}_0 = 0$. Since $\tilde{\boldsymbol{q}}_\ell$ is a linear combination of $\tilde{\boldsymbol{\phi}}_\ell$, we have $\tilde{\boldsymbol{q}}_\ell' A \boldsymbol{\phi}_0 = 0$ and $\tilde{\boldsymbol{q}}_\ell' \boldsymbol{\phi}_0 = 0$. Also by construction, $\langle \tilde{\boldsymbol{q}}_{\ell_1}, \tilde{\boldsymbol{q}}_{\ell_2} \rangle = 0$ for $\ell_1 \neq \ell_2$. Let $\boldsymbol{q}_\ell = \tilde{\boldsymbol{q}}_\ell / \sqrt{\langle \tilde{\boldsymbol{q}}_\ell, \tilde{\boldsymbol{q}}_\ell \rangle}$, then $\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_K\}$ is a set of bases in $\mathbb{R}^T$ that satisfies the conditions in (13)–(15).

11

Let $\tilde{\Phi} := \left(\tilde{\phi}_1, \tilde{\phi}_2, \ldots, \tilde{\phi}_K \right)$ and $\tilde{Q} := \left(\tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_K \right)$. Then $\tilde{\Phi} = \tilde{Q}\tilde{R}$ where R is the upper triangular matrix given by

$$
\tilde{R} = \begin{bmatrix}
1 & \frac{\langle \tilde{\phi}_2, \tilde{q}_1 \rangle}{\langle \tilde{q}_1, \tilde{q}_1 \rangle} & \cdots & \frac{\langle \tilde{\phi}_K, \tilde{q}_1 \rangle}{\langle \tilde{q}_1, \tilde{q}_1 \rangle} \\
0 & 1 & \cdots & \cdots \\
\cdots & \cdots & \ddots & \frac{\langle \tilde{\phi}_K, \tilde{q}_{K-1} \rangle}{\langle \tilde{q}_{K-1}, \tilde{q}_{K-1} \rangle} \\
0 & \cdots & 0 & 1
\end{bmatrix}
$$

To represent $\tilde{\Phi}$ in terms of $Q := (q_1, q_2, \ldots, q_K)$, we let $\mathcal{D} = diag(\langle \tilde{q}_1, \tilde{q}_1 \rangle^{1/2}, \ldots, \langle \tilde{q}_K, \tilde{q}_K \rangle^{1/2})$ be a $K \times K$ diagonal matrix. Then $\tilde{Q} = Q\mathcal{D}$ and

$$
\tilde{\Phi} = QR
$$

where $R = \mathcal{D}\tilde{R}$. The above decomposition is related to the QR decomposition but the columns of Q are not orthonormal in the usual Euclidean sense but instead they are orthonormal under the inner product defined in (12).

In a matrix programming environment, we can compute Q easily. Let R^* be the upper triangular factor of the Cholesky decomposition of $\tilde{\Phi}' A \tilde{\Phi}/T^2$, that is, $\tilde{\Phi}' A \tilde{\Phi}/T^2 = (R^*)' R^*$. Then we can let $Q = \tilde{\Phi}(R^*)^{-1}$, which satisfies

$$
Q'AQ/T^2 = \left(R^{*\prime} \right)^{-1} \tilde{\Phi}' A \tilde{\Phi} \left(R^* \right)^{-1} /T^2 = \left(R^{*\prime} \right)^{-1} \left(R^* \right)' R^* \left(R^* \right)^{-1} = I_K,
$$

as desired.

Using the basis vectors $\{q_1, \ldots, q_K\}$, we can construct the variance estimator:

$$
\hat{\Omega}_q = \frac{D_T^2}{T^2} \frac{1}{K} \sum_{\ell=1}^{K} \left(\sum_{t=1}^{T} q_{\ell t} \hat{e}_t \right) \left(\sum_{s=1}^{T} q_{\ell s} \hat{e}_s \right)'
$$

$$
= \frac{D_T^2}{T^2} \frac{1}{K} \sum_{\ell=1}^{K} \left(\sum_{t=1}^{T} q_{\ell t} e_t \right) \left(\sum_{s=1}^{T} q_{\ell s} e_s \right)'
$$

where $q_{\ell t}$ is an element of q_ℓ so that $q_\ell = (q_{\ell 1}, \ldots, q_{\ell T})'$ and the second equality holds because $\sum_{t=1}^{T} q_{\ell t} = q_\ell' \phi_0 = 0$. The associated F_T statistic is

$$
F_{T,q} = D_T \left(\hat{\theta} - \theta_0 \right)' \hat{\Omega}_q^{-1} D_T \left(\hat{\theta} - \theta_0 \right) /p
$$

$$
= \left(\frac{1}{T} \sum_{t=1}^{T} \frac{e_t}{\sqrt{T}} \right)' \left[\frac{1}{K} \sum_{j=1}^{K} \left(\frac{1}{T} \sum_{t=1}^{T} q_{\ell t} \frac{e_t}{\sqrt{T}} \right) \left(\frac{1}{T} \sum_{s=1}^{T} q_{\ell s} \frac{e_s}{\sqrt{T}} \right)' \right]^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \frac{e_t}{\sqrt{T}} \right) /p
$$

and the t_T statistic is

$$
t_{T,q} = \frac{D_T \left(\hat{\theta} - \theta_0 \right)}{\sqrt{\hat{\Omega}_q}} = \left[\frac{1}{T} \sum_{t=1}^{T} \frac{e_t}{\sqrt{T}} \right] \left[\frac{1}{K} \sum_{j=1}^{K} \left(\frac{1}{T} \sum_{t=1}^{T} q_{\ell t} \frac{e_t}{\sqrt{T}} \right)^2 \right]^{-1/2}.
$$

As $T \to \infty$ for a fixed c_m and h, we have

$$\left(\frac{1}{\sqrt{\langle \phi_0, \phi_0 \rangle}} \frac{1}{T} \sum_{t=1}^{T} \frac{e_t}{\sqrt{T}}, \frac{1}{T} \sum_{t=1}^{T} q_{1t} \frac{e_t}{\sqrt{T}}, \ldots, \frac{1}{T} \sum_{t=1}^{T} q_{Kt} \frac{e_t}{\sqrt{T}} \right) \to^d \Lambda \left(\eta_0, \eta_1, \ldots, \eta_K \right)$$

where $(\eta_0, \eta_1, \ldots, \eta_K)$ are jointly normal. Since $\langle \boldsymbol{q}_\ell, \phi_0 \rangle = 0$ for $\ell = 1, ..., K$ and $\langle \boldsymbol{q}_{\ell_1}, \boldsymbol{q}_{\ell_2} \rangle = \delta_{\ell_1, \ell_2}$, we know that $\eta_\ell \sim iidN(0, I_p)$ for $\ell = 0, 1, \ldots, K$. Hence

$$\langle \phi_0, \phi_0 \rangle^{-1} F_{T,q} \to^d F_\infty (c_m) = \eta_0' \left\{ \frac{1}{K} \sum_{\ell=1}^{K} \eta_\ell \eta_\ell' \right\}^{-1} \eta_0/p.$$

Note that $pF_\infty (c_m)$ follows Hotelling's $T^2_{p,K}$ distribution. Using the relationship between the T^2 distribution and the F distribution, we have

$$\begin{aligned}
F^*_{T,q} &:= \frac{K - p + 1}{K} \langle \phi_0, \phi_0 \rangle^{-1} F_{T,q} \\
&\to^d \frac{K - p + 1}{K} F_\infty (c_m) =^d \frac{K - p + 1}{Kp} T^2_{p,K} =^d F_{p,K-p+1},
\end{aligned} \tag{16}$$

where $F_{p,K-p+1}$ is the F distribution with degrees of freedom $(p, K - p + 1)$. Similarly,

$$t^*_{T,q} := \langle \phi_0, \phi_0 \rangle^{-1} t_{T,q} \to^d t_K, \tag{17}$$

where t_K is the t distribution with degrees of freedom K.

Note that

$$\begin{aligned}
\langle \phi_0, \phi_0 \rangle &= \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \kappa_{c_m} \left(\frac{t}{T}, \frac{s}{T} \right) \\
&\to \int_0^1 \int_0^1 \kappa_{c_m} (r, s) \, dr ds = \frac{1}{c_m^2} - \frac{1}{2c_m^3} \left(e^{-2c_m} - 4e^{-c_m} + 3 \right),
\end{aligned}$$

as $T \to \infty$, we can replace $\langle \phi_0, \phi_0 \rangle$ by the above limit in the definitions of $F^*_{T,q}$ and $t^*_{T,q}$ and still get the same limit distributions.

Our asymptotic development justifies the use of standard F and t approximations in hypothesis testing, even if the underlying process has strong autocorrelation. More specifically, we can employ the modified test statistic $\tilde{F}_T$ and a standard F distribution as the reference distribution to test joint hypotheses. When there is a single hypothesis, we can employ the modified t statistic and the standard t distribution. This leads to our asymptotic F test and t test.

The limiting distributions $F_{p,K-p+1}$ and t_K are exactly the same as what Sun (2013a) obtained in the absence of strong autocorrelation. More specifically, when ρ is a fixed constant less than 1 in absolute value, Sun (2013a) employs orthonormal basis functions,

13

say $\{\phi_{o\ell}(\cdot)\}$, in $\mathbb{L}^2[0,1]$ with $\int_0^1 \phi_{o\ell}(r)dr = 0$ to construct the series variance estimator, and establishes the following weak convergence results for the associated F and t statistics:

$$F_T \to^d F_\infty(\infty) \text{ and } t_T \to^d t_\infty(\infty)$$

where

$$\frac{K-p+1}{K} F_\infty(\infty) =^d F_{p,K-p+1} \text{ and } t_\infty(\infty) =^d t_K.$$

That is

$$\frac{K-p+1}{K} F_T \to^d F_{p,K-p+1} \text{ and } t_T \to^d t_K.$$

To relate the above approximations to those in (16) and (17), we consider the case when c_m is large. As $c_m \to \infty$, we have

$$\int_0^1 \int_0^1 \phi_{o\ell_1}(r) \phi_{o\ell_2}(s) c_m^2 \kappa_{c_m}(r,s) dr ds = \int_0^1 \phi_{o\ell_1}(r) \phi_{o\ell_2}(r) dr + o(1),$$

$$\int_0^1 \int_0^1 \phi_{o\ell_1}(r) c_m^2 \kappa_{c_m}(r,s) dr ds = \int_0^1 \phi_{o\ell_1}(r) dr + o(1).$$

So if $\{\phi_{o\ell}(\cdot)\}$ is a set of orthonormal basis functions in $\mathbb{L}^2[0,1]$ with $\int_0^1 \phi_{o\ell}(r)dr = 0$, then $\{c_m\phi_{o\ell}(r)\}$ will satisfy (9)–(11) approximately. That is, as $c_m \to \infty$, the effects of transformation and orthogonalization on $\{c_m\phi_{o\ell}(\cdot)\}$ becomes negligible. We can just use $\{\phi_{o\ell}(\cdot)\}$ as the basis functions to construct the series variance estimator, leading to the modified F statistic F_{T,ϕ_o}^*. We have

$$F_{T,\phi_o}^* := \frac{K-p+1}{K} \langle \phi_0, \phi_0 \rangle^{-1} F_{T,\phi_o} \to^d \frac{K-p+1}{K} F_\infty^{\phi_o}(c_m)$$

where

$$F_\infty^{\phi_o}(c_m) = \left[\int_0^1 J_{c_m}(r) dr\right]'$$

$$\times \left\{\frac{1}{K}\sum_{\ell=1}^K \left[\int_0^1 \phi_{o\ell}(r) J_{c_m}(r) dr\right] \left[\int_0^1 \phi_{o\ell}(r) J_{c_m}(r) dr\right]'\right\}^{-1} \left[\int_0^1 J_{c_m}(r) dr\right]/p$$

$$= \left[\int_0^1 \int_0^1 c_m^2 \kappa_{c_m}(r,s) dr ds\right] \left[\frac{\int_0^1 c_m J_{c_m}(r) dr}{\sqrt{\int_0^1 \int_0^1 c_m^2 \kappa_{c_m}(r,s) dr ds}}\right]' \times$$

$$\left\{\frac{1}{K}\sum_{\ell=1}^K \left[\int_0^1 \phi_{o\ell}(r) c_m J_{c_m}(r) dr\right] \left[\int_0^1 \phi_{o\ell}(r) c_m J_{c_m}(r) dr\right]'\right\}^{-1}$$

$$\times \left[\frac{\int_0^1 c_m J_{c_m}(r) dr}{\sqrt{\int_0^1 \int_0^1 c_m^2 \kappa_{c_m}(r,s) dr ds}}\right]/p$$

$$\to^d \eta_0' \left\{\frac{1}{K}\sum_{\ell=1}^K \eta_\ell \eta_\ell'\right\}^{-1} \eta_0/p =^d F_\infty(\infty)$$

14

as $c_m \to \infty$, where we have used $\lim_{c_m \to \infty} \int_0^1 \int_0^1 c_m^2 \kappa_{c_m}(r,s) dr ds = 1$. So F_{T,ϕ_o}^* converges weakly to $F_{p,K-p+1}$ under the sequential limits. This is consistent with Theorem 2.

5   Extension to a General Setting

In the previous section, we use the simple multivariate location model to highlight the effect of strong autocorrelation on distributional approximations. Hypothesis testing in location models, as simple as it seems, includes more general testing problems as special cases.

Consider an M-estimator, $\hat{\theta}_T$, of a $n \times 1$ parameter vector θ_0 that satisfies

$$\hat{\theta}_T = \arg\min_{\theta \in \Theta} Q_T(\theta) = \arg\min_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^T \rho(\theta, Z_t)$$

where Θ is a compact parameter space, and $\rho(\theta, Z_t)$ is the criterion function based on observation Z_t. M-estimators are a broad class of estimators and include, for example, the maximum likelihood estimator (MLE), ordinary least squares (OLS) estimator, quantile regression estimator as special cases.

Suppose we want to test the null hypothesis that $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. Then by the usual identification assumption for the M-estimator, under the null hypothesis and additional regularity assumptions, $\theta = \theta_0$ is the unique minimizer of

$$Q(\theta) = E\rho(\theta, Z_t).$$

That is

$$Es_t(\theta) = 0 \text{ where } s_t(\theta) = \frac{\partial \rho(\theta, Z_t)}{\partial \theta},$$

if and only if $\theta = \theta_0$. So the null hypothesis $H_0 : \theta = \theta_0$ is equivalent to the hypothesis that the multivariate process $s_t(\theta_0)$ has mean zero. We have just converted a general testing problem into testing for zero mean of a multivariate process. The latter problem is exactly the testing problem we consider in the previous sections. All results there remain valid if the multivariate process $s_t(\theta_0)$ satisfies the assumptions imposed on y_t.

The above extension applies to hypothesis testing that involves the whole parameter vector θ. Suppose we are only interested in some linear combinations of θ such that the null hypothesis is $H_0 : R\theta = r$ and the alternative hypothesis is $H_1 : R\theta \neq r$, where R is a $p \times n$ matrix. Under the usual regularity conditions, we have

$$\sqrt{T}\left(R\hat{\theta}_T - r\right) = -\frac{1}{\sqrt{T}} \sum_{t=1}^T RH_T(\bar{\theta}_T)^{-1} s_t(\theta_0) \text{ for } H_T(\theta) = \frac{1}{T} \sum_{t=1}^T \frac{\partial s_t(\theta)}{\partial \theta'}$$

15

and some value $\bar{\theta}_T$ between θ_0 and $\hat{\theta}_T$. Let $\tilde{s}_t(\theta_1, \theta_2) = -RH_T(\theta_1)^{-1}s_t(\theta_2)$ be the transformed score process, then the F-test version of the Wald statistic is

$$\tilde{F}_T = T\left(R\hat{\theta}_T - r\right)'\tilde{\Omega}^{-1}\left(R\hat{\theta}_T - r\right)/p, \tag{18}$$

where

$$\tilde{\Omega} = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}Q_{Th}^*\left(\frac{t}{T}, \frac{\tau}{T}\right)\tilde{s}_t\left(\hat{\theta}_T, \hat{\theta}_T\right)\tilde{s}_\tau'\left(\hat{\theta}_T, \hat{\theta}_T\right)$$

Under the assumptions given in KV (2005) and Sun (2013a), $\tilde{F}_T$ is asymptotically equivalent to

$$F_T = \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_t\right]'\hat{\Omega}^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_t\right]/p, \tag{19}$$

where $e_t = -RH^{-1}(\theta_0)s_t(\theta_0)$, $H(\theta_0) = \text{plim}_{T\to\infty}H_T(\bar{\theta}_T)$, and

$$\hat{\Omega} = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}Q_{Th}^*\left(\frac{t}{T}, \frac{s}{T}\right)(e_t - \bar{e})(e_s - \bar{e})'.$$

It follows that the F_T statistic can be viewed as a Wald statistic for testing whether the mean of the multivariate process e_t is zero. So the asymptotic approximation in Theorem 1 applies to the Wald statistic. The same observation remains valid for the t statistic.

More generally, consider a standard GMM framework with the moment conditions

$$Ef(Z_t, \theta_0) = 0, \ t = 1, 2, \ldots, T$$

where $\theta_0 \in \Theta \subseteq \mathbb{R}^n$ and $f(\cdot)$ is an $m \times 1$ vector of twice continuously differentiable functions with $m \geq n$ and rank $E\left[\partial f(Z_t, \theta_0)/\partial\theta'\right] = n$. The GMM estimator of θ_0 is then given by

$$\hat{\theta}_T = \arg\min_{\theta\in\Theta}\left(\frac{1}{T}\sum_{t=1}^{T}f(Z_t, \theta)\right)'\mathcal{W}_T\left(\frac{1}{T}\sum_{t=1}^{T}f(Z_t, \theta)\right)$$

where $\mathcal{W}_T$ is an $m \times m$ positive semidefinite weighting matrix and $\text{plim}_{T\to\infty}\mathcal{W}_T = \mathcal{W}_\infty$.

Suppose we test $H_0 : R\theta = r$ against the alternative $H_1 : R\theta \neq r$. Let $G_T(\theta) = \text{plim}_{T\to\infty}\left[\frac{1}{T}\sum_{t=1}^{T}\partial f(Z_t, \theta)/\partial\theta'\right]$. Then under the usual regularity conditions for GMM estimation, we have

$$\sqrt{T}\left(R\hat{\theta}_T - r\right) = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}R\left[G_T(\hat{\theta}_T)'\mathcal{W}_TG_T\left(\hat{\theta}_T\right)\right]^{-1}G_T(\hat{\theta}_T)'\mathcal{W}_Tf(Z_t, \theta_0) + o_p(1) \tag{20}$$

As before, we let $\tilde{s}_t(\theta_1, \theta_2) = R\left[G_T(\theta_1)'\mathcal{W}_TG_T(\theta_1)\right]^{-1}G_T(\theta_1)'\mathcal{W}_Tf(Z_t, \theta_2)$. Then the Wald statistic can be computed in the same way as in (18), which is asymptotically equivalent to F_T in (18) with

$$e_t = -R\left(G_0'\mathcal{W}_\infty G_0\right)^{-1}G_0'\mathcal{W}_\infty f(Z_t, \theta_0).$$

With this new definition of e_t, the asymptotic approximations in the previous sections are applicable to both the Wald statistic and t statistic.

6 Implementation and Simulation

6.1 Implementation

The near-unit-root fixed-smoothing asymptotic approximations $F_\infty(c_m)$ and $t_\infty(c_m)$ depend on the near-unit-root parameter c_m, which cannot be consistently estimated. There are many ways to gauge the value of c_m in the literature on optimal unit root testing. Sun (2014) constructs confidence intervals for c_m and uses the method of Bonferroni bound to obtain tests with good size properties. Here for simplicity we use the OLS estimator:

$$\hat{\rho}_i = \frac{\sum_{t=2}^{T} \hat{e}_{it}\hat{e}_{i,t-1}}{\sum_{t=2}^{T} \hat{e}_{i,t-1}^2}$$

where for the location model, we let $\hat{e}_{it} = y_{it} - \bar{y}_{i,.}$, for the model estimated by the M estimator, we let $\hat{e}_{it}$ be the i-th component of $RH_T^{-1}(\hat{\theta}_T)s_t(\hat{\theta}_T)$, for the model estimated by the GMM estimator, we let $\hat{e}_{it}$ be the i-th component of $R\left[\hat{G}_T'\mathcal{W}_T\hat{G}_T\right]^{-1}\hat{G}_T\mathcal{W}_T f(Z_t, \hat{\theta}_T)$ with $\hat{G}_T = G_T(\hat{\theta}_T)$. Given the average of the estimated autoregressive parameters: $\hat{\rho} = p^{-1}\sum_{i=1}^{p}\hat{\rho}_i$, we take c_m to satisfy $1 - \hat{c}_m/T = \hat{\rho}$, that is, $\hat{c}_m = T(1 - \hat{\rho})$. To reduce the randomness of $\hat{\rho}$ and hence $\hat{c}_m$, we can discretize the interval $[-1, 1]$ and use the grid point that is closest to $\hat{\rho}$ as the autoregressive parameter. Let $\tilde{\rho}$ be this grid point, we can let $\hat{c}_m = T(1 - \tilde{\rho})$.

It is important to point out that while we propose using $F_\infty(c_m)$ or $t_\infty(c_m)$ as the reference distributions to perform hypothesis testing, it does not mean that we literally treat c_m as fixed. Whether we hold c_m fixed or let it increase with the sample size can be viewed as different asymptotic specifications to obtain approximations to the finite sample distribution. The near-unit-root fixed-smoothing asymptotics does not require that we fix the value of c_m in finite samples. In fact, if e_t is stationary with a fixed autoregressive coefficient, then with probability approaching one, $\hat{c}_m$ will increase with the sample size. This will not cause any problem, as we have shown that as c_m increases, $F_\infty(c_m)$ will become very close to $F_\infty(\infty)$, the fixed-smoothing asymptotic distribution under weak dependence. So the near-unit-root fixed-smoothing approximation is asymptotically valid under weak dependence. To some extent, $F_\infty(c_m)$ is a more robust approximation than $F_\infty(\infty)$.

6.2 Simulations

For the Monte Carlo experiments, we first consider a simple multivariate location model with 4 time series:

$$y_t = \theta + u_t \tag{21}$$

where the error u_t follows either a VAR(1) or VMA(1) process:

$$u_t = Au_{t-1} + \varepsilon_t \text{ or } u_t = A\varepsilon_{t-1} + \varepsilon_t$$

where $A = \rho\mathbb{I}_4$, $\varepsilon_t = (v_{1t} + \mu f_t, v_{2t} + \mu f_t, ..., v_{nt} + \mu f_t)' / \sqrt{1 + \mu^2}$ and $(v_t, f_t)'$ is a multivariate Gaussian white noise process with unit variance. In the case of VAR(1), we set $u_0 \sim N(0, \mathbb{I}_4)$. Under this specification, the four time series all follow the same VAR(1) or VMA(1) process with $\varepsilon_t \sim iidN(0, \Sigma)$ for

$$\Sigma = \frac{1}{1 + \mu^2}\mathbb{I}_4 + \frac{\mu^2}{1 + \mu^2}\mathbb{J}_4,$$

where $\mathbb{J}_4$ is a 4×4 matrix of ones. The parameter μ determines the degree of dependence among the time series considered. When $\mu = 0$, the four series are uncorrelated with each other. The large μ is, the larger the correlation is.

For the model parameters, we take $\rho = -0.5, 0, 0.1, 0.3, 0.5, 0.7, 0.9$, and 0.95 and set $\mu = 0$ and 1. We set the intercepts to zero as the tests we consider are invariant to them. For each test, we consider two significance levels $\alpha = 5\%$ and $\alpha = 10\%$ and two different sample sizes $T = 200$ and 400.

We consider the following null hypotheses:

$$H_{01} : \theta_1 = 0,$$
$$H_{02} : \theta_1 = \theta_2 = 0,$$
$$H_{03} : \theta_1 = \theta_2 = \theta_3 = 0,$$
$$H_{04} : \theta_1 = \theta_2 = \theta_3 = \theta_4 = 0,$$

where $p = 1, 2, 3, 4$, respectively. The corresponding matrix R is the first p rows of the identity matrix $\mathbb{I}_4$.

We examine the finite sample performance of three different groups of tests. The first group of tests consists of the standard fixed-smoothing Wald tests of KVB (2000) and Sun (2013a). The KVB test employs the Bartlett kernel variance estimator based with $b = 1$, and the test in Sun (2013a) employs the series variance estimator with $K = 6, 12$, and 24. The basis functions used are $\phi_{2j-1}(x) = \sqrt{2}\cos(2j\pi x)$, $\phi_{2j}(x) = \sqrt{2}\sin(2j\pi x), j = 1, \ldots, K/2$. For both the KVB test and the test in Sun (2013a), we use standard fixed-smoothing critical values, which are obtained from simulations (the KVB test) or directly from the standard F distributions (the test in Sun (2013a)). The second group of tests is similar to the first group but uses critical values from the near-unit-root fixed-smoothing asymptotic distributions developed in Section 3. The critical values are obtained via simulations with 999 simulation replications. The third group of tests is based on the F approximation in Section 4. The basis functions/vectors are transformed and orthogonalized

18

version of $\phi_{2j-1}(x) = \sqrt{2}\cos(2j\pi x)$, $\phi_{2j}(x) = \sqrt{2}\sin(2j\pi x)$, $j = 1, \ldots, K/2$. These three groups of tests will be referred to as the "Stationary Fixed-smoothing Tests", "Near-Unity Fixed-smoothing Tests", and "Near-Unity F Tests." We could add the conventional chi-square tests but a large literature has already shown that the chi-square tests have much larger size distortion than the corresponding fixed-smoothing tests.

Table 1 reports the empirical null rejection probability of the three groups of tests for VAR(1) error. The sample size is 200 and the number of simulation replications is 10000. A few observations can be drawn from the table. First, it is clear that the stationary fixed-smoothing asymptotic approximation works very well when the processes are not strongly autocorrelated. When the processes become more autocorrelated, the size distortion starts to increase. This is especially true when the number of joint hypotheses is more than one. The size distortion of the stationary fixed-smoothing tests can be as high as 0.795. This happens when $K = 24$ and $p = 4$. Even for the KVB test with $b = 1$, the size distortion is 0.488 when $p = 4$. Second, the near-unity fixed-smoothing asymptotic tests outperform the stationary fixed-smoothing tests in terms of size accuracy. When the processes are not strongly autocorrelated, the near-unity fixed-smoothing asymptotic tests have more or less the same size properties as the stationary fixed-smoothing asymptotic tests. However, when the processes approach unit root nonstationarity, the near-unity fixed-smoothing asymptotic tests succeed in reducing the size distortion. Comparing the tests in the second group, the series test with $K = 6$ and the KVB test have the most accurate size. Third, the near-unity asymptotic F tests are also more accurate than the stationary fixed-smoothing tests, which in the case of series variance estimation use also the F critical values. When K is relatively small, the near-unity asymptotic F test is as accurate as the test based on nonstandard simulated critical values. However, when K is large, the near-unity asymptotic F test has somewhat larger size distortion.

Table 2 is the same as Table 1 except that the sample size is $T = 400$. The qualitative observations we make for Table 1 apply to Table 2. As expected, all tests become more accurate. We omit the Tables for the VMA(1) cases as all three groups of tests have similar good size properties for both sample sizes $T = 200$ and 400. There is no adverse effect of using the near-unity fixed-smoothing asymptotic distribution when the underlying processes are not strongly persistent. This is consistent with our theoretical findings.

Figure 4 is a representative figure for size-adjusted power curves. The sample size is 200 and the number of joint hypotheses is 2. Of course, size-adjustment is not feasible in practice; and this is exactly the reason we want to develop an asymptotically valid test that has accurate size in finite samples. Since tests based on the same statistic have the same size-adjusted power, a test in the first group has the same power as the corresponding

test in the second group. So it suffices to report the power curves of the tests in the first and second groups. In Figure 4, "K6", "K12", "K24" and "KVB" are the tests in the first group while "K6*", "K12*", "K24*" are the tests in the second group. It is clear from the figure that the power of the series test increases with K, as shown in Sun (2013a). What is new here is that the power of the fixed-smoothing nonstandard test is close to that of the fixed-smoothing F test. Transformation and orthogonalization underlying the F test do not lead to power loss.

7   Conclusions

The paper develops a new fixed-smoothing asymptotic theory that accommodates strongly autocorrelated time series while also working very well in the absence of strong autocorrelation. Our proposed near-unity fixed-smoothing tests achieve triple robustness in the following sense: they are asymptotically valid regardless of whether the amount of smoothing is fixed or increases with the sample size; whether there is temporal dependence of unknown form or not; and whether the temporal dependence is strong or not when presented.

The near-unity fixed-smoothing asymptotic distribution is nonstandard. By choosing the basis functions appropriately, the nonstandard distribution can be reduced to a standard F approximation. See Sun and Kim (2013) for an implementation of this idea in a different setting. An alternative way to characterize the strong autocorrelation is to model the moment process as a fractional process. In this case, we can follow Sun (2004) to develop a new asymptotic approximation. Our testing procedures can be combined with prewhitening. This may lead to a test with potentially very accurate size, even if the underlying processes are highly autocorrelated. We leave all these to future research.

Table 1: Empirical size of 5% Fixed-smoothing Asymptotic Tests with T = 200 under VAR(1) Errors

ρ	Stationary Fixed-smoothing				Near-Unity Fixed-smoothing				Near-Unity F Tests		
	K6	K12	K24	KVB	K6	K12	K24	KVB	K6*	K12*	K24*
							$p = 1$				
-0.5	0.048	0.051	0.050	0.043	0.039	0.045	0.043	0.044	0.048	0.050	0.049
0.0	0.049	0.051	0.052	0.048	0.039	0.045	0.042	0.045	0.049	0.050	0.047
0.1	0.049	0.051	0.052	0.049	0.039	0.044	0.042	0.045	0.048	0.050	0.046
0.3	0.049	0.052	0.054	0.050	0.040	0.044	0.043	0.045	0.048	0.050	0.045
0.5	0.049	0.055	0.062	0.053	0.040	0.046	0.044	0.047	0.048	0.050	0.044
0.7	0.051	0.063	0.087	0.060	0.041	0.049	0.050	0.047	0.048	0.051	0.047
0.9	0.076	0.138	0.254	0.094	0.053	0.066	0.071	0.055	0.059	0.076	0.088
0.95	0.143	0.264	0.415	0.136	0.083	0.103	0.111	0.073	0.091	0.124	0.146
							$p = 2$				
-0.5	0.051	0.050	0.050	0.041	0.052	0.043	0.044	0.044	0.052	0.047	0.046
0.0	0.052	0.050	0.051	0.049	0.053	0.043	0.043	0.046	0.052	0.048	0.045
0.1	0.052	0.050	0.052	0.051	0.053	0.043	0.043	0.045	0.051	0.048	0.045
0.3	0.053	0.051	0.056	0.056	0.053	0.044	0.043	0.044	0.051	0.049	0.044
0.5	0.053	0.053	0.067	0.062	0.054	0.044	0.043	0.045	0.050	0.049	0.043
0.7	0.055	0.064	0.101	0.074	0.051	0.045	0.047	0.048	0.050	0.049	0.046
0.9	0.083	0.169	0.353	0.137	0.059	0.073	0.086	0.063	0.062	0.078	0.095
0.95	0.157	0.359	0.590	0.229	0.085	0.120	0.142	0.085	0.096	0.142	0.184
							$p = 3$				
-0.5	0.050	0.050	0.048	0.036	0.047	0.050	0.048	0.043	0.050	0.051	0.044
0.0	0.051	0.051	0.049	0.050	0.046	0.050	0.048	0.044	0.051	0.050	0.043
0.1	0.051	0.052	0.050	0.052	0.046	0.050	0.049	0.045	0.051	0.049	0.043
0.3	0.051	0.053	0.054	0.058	0.044	0.051	0.049	0.046	0.052	0.049	0.042
0.5	0.052	0.055	0.065	0.069	0.042	0.051	0.050	0.046	0.053	0.049	0.040
0.7	0.053	0.066	0.110	0.090	0.045	0.055	0.059	0.048	0.052	0.050	0.043
0.9	0.079	0.189	0.428	0.201	0.063	0.085	0.100	0.062	0.058	0.079	0.098
0.95	0.152	0.427	0.710	0.354	0.090	0.150	0.172	0.088	0.089	0.150	0.205
							$p = 4$				
-0.5	0.051	0.050	0.048	0.033	0.047	0.050	0.043	0.051	0.052	0.051	0.044
0.0	0.051	0.051	0.050	0.051	0.046	0.052	0.042	0.052	0.051	0.050	0.043
0.1	0.050	0.051	0.051	0.054	0.047	0.052	0.043	0.052	0.050	0.050	0.042
0.3	0.050	0.052	0.055	0.062	0.048	0.051	0.044	0.051	0.050	0.048	0.041
0.5	0.050	0.055	0.066	0.075	0.049	0.052	0.046	0.051	0.049	0.048	0.041
0.7	0.052	0.065	0.117	0.108	0.049	0.052	0.053	0.053	0.050	0.048	0.043
0.9	0.072	0.202	0.493	0.275	0.065	0.083	0.096	0.078	0.057	0.074	0.102
0.95	0.129	0.473	0.795	0.488	0.081	0.154	0.188	0.130	0.081	0.155	0.229

Table 2: Empirical size of 5% Fixed-smoothing Asymptotic Tests with T = 400 under VAR(1) Errors

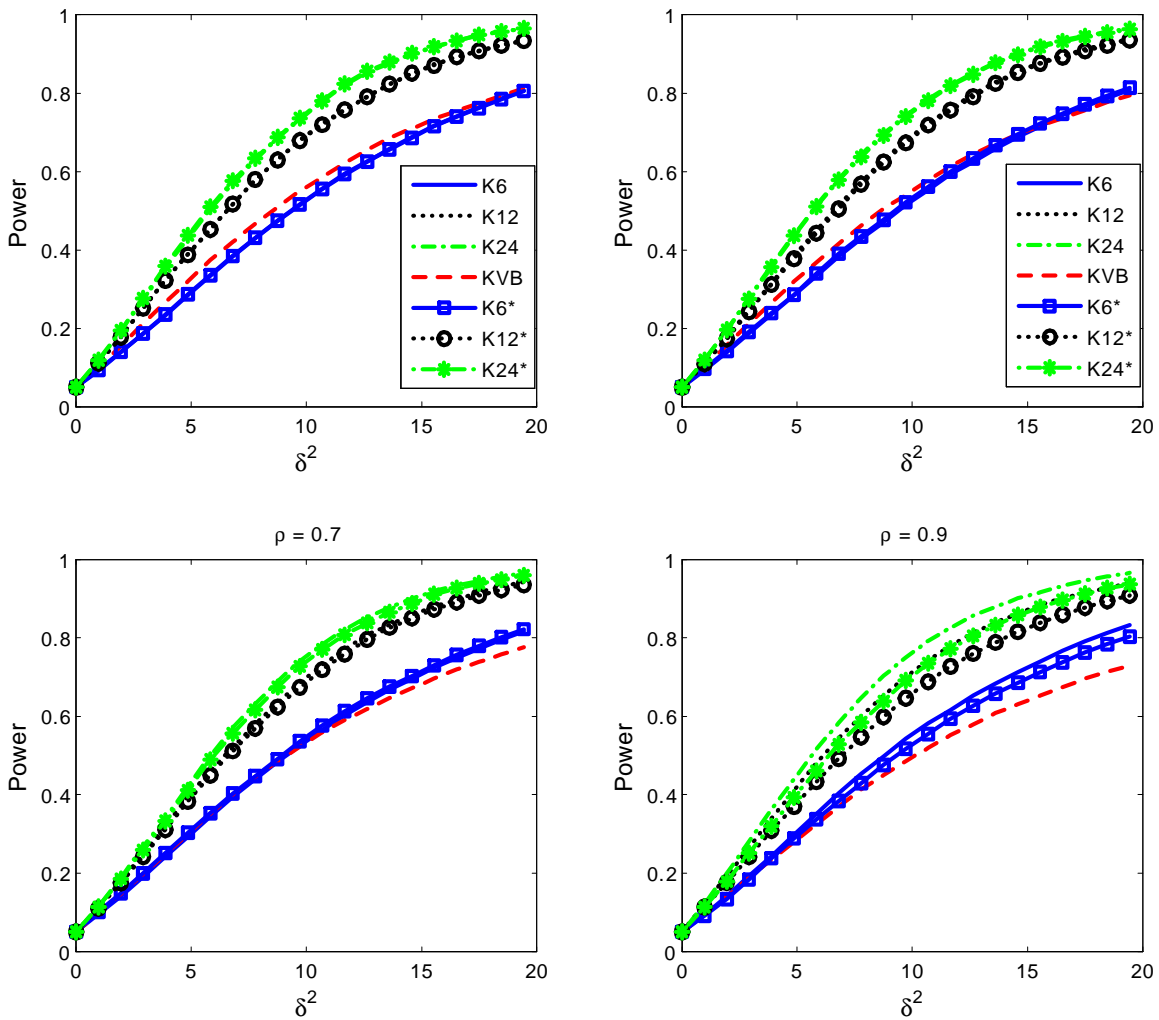| ρ	Stationary Fixed-smoothing				Near-Unity Fixed-smoothing				Near-Unity F Tests		
	K6	K12	K24	KVB	K6	K12	K24	KVB	K6*	K12*	K24*
						$p=1$					
-0.5	0.051	0.047	0.047	0.048	0.049	0.039	0.042	0.037	0.049	0.047	0.046
0.0	0.051	0.048	0.046	0.050	0.044	0.040	0.044	0.035	0.049	0.046	0.045
0.1	0.050	0.048	0.046	0.051	0.044	0.040	0.045	0.035	0.049	0.046	0.045
0.3	0.050	0.048	0.047	0.052	0.044	0.040	0.045	0.036	0.049	0.046	0.044
0.5	0.050	0.049	0.049	0.054	0.045	0.041	0.044	0.037	0.049	0.046	0.044
0.7	0.051	0.050	0.057	0.056	0.046	0.041	0.044	0.037	0.049	0.046	0.044
0.9	0.058	0.076	0.134	0.072	0.047	0.051	0.058	0.043	0.051	0.053	0.059
0.95	0.080	0.138	0.252	0.090	0.056	0.071	0.086	0.050	0.062	0.079	0.095
						$p=2$					
-0.5	0.049	0.047	0.047	0.043	0.057	0.048	0.045	0.045	0.050	0.046	0.047
0.0	0.048	0.047	0.047	0.048	0.059	0.046	0.045	0.046	0.050	0.046	0.046
0.1	0.048	0.047	0.047	0.048	0.060	0.047	0.046	0.045	0.050	0.047	0.045
0.3	0.048	0.047	0.048	0.050	0.062	0.046	0.047	0.046	0.049	0.046	0.045
0.5	0.049	0.048	0.050	0.053	0.062	0.046	0.048	0.044	0.050	0.045	0.045
0.7	0.050	0.050	0.058	0.058	0.059	0.047	0.048	0.047	0.048	0.045	0.044
0.9	0.058	0.081	0.166	0.088	0.062	0.053	0.062	0.058	0.051	0.054	0.059
0.95	0.081	0.171	0.355	0.134	0.071	0.077	0.095	0.072	0.061	0.080	0.104
						$p=3$					
-0.5	0.049	0.047	0.048	0.042	0.043	0.039	0.043	0.044	0.050	0.047	0.047
0.0	0.048	0.048	0.048	0.047	0.041	0.038	0.044	0.042	0.050	0.047	0.046
0.1	0.048	0.048	0.049	0.049	0.041	0.038	0.044	0.043	0.050	0.047	0.046
0.3	0.048	0.048	0.050	0.052	0.042	0.038	0.043	0.045	0.050	0.048	0.046
0.5	0.049	0.049	0.052	0.057	0.044	0.037	0.040	0.045	0.051	0.047	0.045
0.7	0.051	0.051	0.062	0.067	0.046	0.037	0.039	0.048	0.049	0.048	0.045
0.9	0.058	0.084	0.192	0.119	0.048	0.042	0.059	0.058	0.051	0.054	0.063
0.95	0.082	0.188	0.438	0.200	0.060	0.063	0.087	0.071	0.061	0.081	0.112
						$p=4$					
-0.5	0.049	0.048	0.048	0.040	0.058	0.043	0.048	0.045	0.051	0.047	0.047
0.0	0.050	0.049	0.049	0.049	0.051	0.044	0.047	0.048	0.052	0.048	0.048
0.1	0.050	0.049	0.049	0.051	0.053	0.044	0.047	0.049	0.052	0.048	0.047
0.3	0.050	0.049	0.050	0.055	0.051	0.045	0.046	0.048	0.052	0.048	0.046
0.5	0.051	0.049	0.052	0.061	0.046	0.044	0.045	0.049	0.051	0.047	0.045
0.7	0.050	0.051	0.063	0.076	0.044	0.044	0.047	0.044	0.049	0.046	0.045
0.9	0.055	0.088	0.216	0.154	0.049	0.051	0.060	0.052	0.051	0.053	0.063
0.95	0.072	0.206	0.505	0.285	0.057	0.081	0.102	0.078	0.058	0.083	0.120

Figure 4: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with sample size $T = 200$ and the number of joint hypotheses $p = 2$ ("K6", "K12", "K24" and "KVB" are the Near-Unity Fixed-smoothing tests while "K6*", "K12*", "K24*" are the Near-Unity F tests)

8 Appendix

The following lemma will be used in our proofs. It is a slight modification of Lemma 1 in Sun (2013b) and can be proved in the same way. The details are omitted here.

Lemma 1 *Suppose $\omega_T = \xi_{T,M} + \eta_{T,M}$ and ω_T does not depend on M. Assume that there exist $\xi_{\infty,M}^*$ and $\xi_{\infty,\infty}^*$ such that*

 (i) $P(\xi_{T,M} < \xi) - P\left(\xi_{\infty,M}^* < \xi\right) = o(1)$ for each fixed M and each $\xi \in \mathbb{R}$ as $T \to \infty$,

 (ii) $P\left(\xi_{\infty,M}^* < \xi\right) - P\left(\xi_{\infty,\infty}^* < \xi\right) = o(1)$ for each $\xi \in \mathbb{R}$ as $M \to \infty$,

 (iii) the CDF of $\xi_{\infty,\infty}^*$ is continuous on $\mathbb{R}$,

 (iv) For every $\delta > 0$, $\sup_T P\left(|\eta_{T,M}| > \delta\right) \xrightarrow{p} 0$ as $M \to \infty$. Then

$$P(\omega_T < \xi) = P\left(\xi_{\infty,\infty}^* < \xi\right) + o(1) \text{ for each } \xi \in \mathbb{R} \text{ as } T \to \infty.$$

Proof of Theorem 1. Under the null hypothesis, we have

$$\frac{1}{T} \sum_{t=1}^{T} \frac{y_t}{\sqrt{T}} = \frac{1}{T} \sum_{t=1}^{T} \frac{e_t}{\sqrt{T}} \to^d \int_0^1 J_c(r)\, dr \tag{22}$$

by the continuous mapping theorem. Under Assumption 2, it is not hard to show that

$$\frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_{Th}^*\left(\frac{t}{T}, \frac{s}{T}\right)(y_t - \bar{y})(y_s - \bar{y})' = \frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h^*\left(\frac{t}{T}, \frac{s}{T}\right)(y_t - \bar{y})(y_s - \bar{y})' + o_p(1).$$

It remains to show that

$$\frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h^*\left(\frac{t}{T}, \frac{s}{T}\right)(y_t - \bar{y})(y_s - \bar{y})' \to^d \int_0^1 \int_0^1 Q_h^*(r,s) J_{c_m}(r) J_{c_m}'(s)\, dr ds \tag{23}$$

jointly with (22) and $\int_0^1 \int_0^1 Q_h^*(r,s) J_{c_m}(r) J_{c_m}'(s)\, dr ds$ is nonsingular with probability one. We focus on proving the marginal convergence in (23), as the joint convergence results can be proved using the Cramer-Wold device and the nonsingularity of the limiting matrix can be proved easily.

Note that (23) is equivalent to

$$\frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h^*\left(\frac{t}{T}, \frac{s}{T}\right)(e_t - \bar{e})' A (e_s - \bar{e}) \to^d \int_0^1 \int_0^1 Q_h^*(r,s) J_{c_m}'(r) A J_{c_m}(s)\, dr ds \tag{24}$$

for all $p \times p$ symmetric matrix A. In view of the spectral decomposition of $A = \sum \mu_\ell a_\ell a_\ell'$, the above holds if

$$\frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h^*\left(\frac{t}{T}, \frac{s}{T}\right)(e_t - \bar{e})' a a' (e_s - \bar{e}) \to^d \int_0^1 \int_0^1 Q_h^*(r,s) J_{c_m}'(r) a a' J_{c_m}(s)\, dr ds$$

$$\tag{25}$$

for all $p \times 1$ vector a. But

$$\frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h^* \left(\frac{t}{T}, \frac{s}{T} \right) (e_t - \bar{e})' aa' (e_s - \bar{e})$$

$$= \frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} \left[Q_h^* \left(\frac{t}{T}, \frac{s}{T} \right) - \bar{Q}_h^* \left(\cdot, \frac{s}{T} \right) - \bar{Q}_h^* \left(\frac{t}{T}, \cdot \right) + \bar{Q}_h^* \left(\cdot, \cdot \right) \right] e_t' aa' e_s$$

$$= \frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h^* \left(\frac{t}{T}, \frac{s}{T} \right) e_t' aa' e_s + O \left(\frac{1}{T} \right) \frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} |e_t' aa' e_s|$$

$$= \frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h^* \left(\frac{t}{T}, \frac{s}{T} \right) e_t' aa' e_s + O \left(\frac{1}{T} \right) \left(\frac{1}{T} \sum_{t=1}^{T} \frac{1}{\sqrt{T}} \|e_t\| \right)^2 \|a\|^2$$

$$= \frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h^* \left(\frac{t}{T}, \frac{s}{T} \right) e_t' aa' e_s + O_p \left(\frac{1}{T} \right).$$

So it suffices to show that

$$\omega_T := \frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h^* \left(\frac{t}{T}, \frac{s}{T} \right) e_t' aa' e_s \to^d \int_0^1 \int_0^1 Q_h^* (r, s) J_{c_m}' (r) aa' J_{c_m} (s) \, dr ds$$

for all $p \times 1$ vector a.

Given that $\sum_j \lambda_j \Phi_j (r) \Phi_j (s)$ converges uniformly over $(r, s) \in [0, 1]^2$ to $Q_h^* (r, s)$, we have, for any $\varepsilon > 0$, there exists an $M > 0$ such that

$$\left| Q_h^* \left(\frac{t}{T}, \frac{s}{T} \right) - \sum_{j=1}^{M} \lambda_j \Phi_j \left(\frac{t}{T} \right) \Phi_j \left(\frac{s}{T} \right) \right| \leq \varepsilon$$

for all $1 \leq t, s \leq T$ and all T. Now

$$\omega_T = \frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h^* \left(\frac{t}{T}, \frac{s}{T} \right) e_t' aa' e_s$$

$$= \frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} \left[\sum_{j=1}^{M} \lambda_j \Phi_j \left(\frac{t}{T} \right) \Phi_j \left(\frac{s}{T} \right) \right] e_t' aa' e_s$$

$$+ \frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} \left[Q_h^* \left(\frac{t}{T}, \frac{s}{T} \right) - \sum_{j=1}^{M} \lambda_j \Phi_j \left(\frac{t}{T} \right) \Phi_j \left(\frac{s}{T} \right) \right] e_t' aa' e_s$$

$$:= \xi_{T,M} + \eta_{T,M}.$$

We use Lemma 1 to complete the proof. Condition (i) of Lemma 1 holds because for each fixed M, we have

$$\xi_{T,M} = \sum_{j=1}^{M} \lambda_j \left[\frac{1}{T} \sum_{t=1}^{T} \Phi_j \left(\frac{t}{T} \right) \frac{e'_t}{\sqrt{T}} a \right] \left[\frac{1}{T} \sum_{s=1}^{T} \Phi_j \left(\frac{s}{T} \right) \frac{e'_s}{\sqrt{T}} a \right]$$

$$\to^d \sum_{j=1}^{M} \lambda_j \left[\int_0^1 \Phi_j(r) J'_{c_m}(r) a \, dr \right] \left[\int_0^1 \Phi_j(s) J'_{c_m}(s) a' ds \right]$$

$$= \int_0^1 \int_0^1 \sum_{j=1}^{M} \lambda_j \Phi_j(r) \Phi_j(s) J'_{c_m}(r) aa' J_{c_m}(s) \, ds := \xi^*_{\infty,M}.$$

To verify condition (ii) of Lemma 1, we let

$$\xi^*_{\infty,\infty} = \int_0^1 \int_0^1 \left[\sum_{j=1}^{\infty} \lambda_j \Phi_j(r) \Phi_j(s) \right] J'_{c_m}(r) aa' J_{c_m}(s) \, ds$$

$$= \int_0^1 \int_0^1 Q^*_h(r,s) J'_{c_m}(r) aa' J_{c_m}(s) \, ds.$$

Then

$$\xi^*_{\infty,M} - \xi^*_{\infty,\infty} = \int_0^1 \int_0^1 \left[\sum_{j=M+1}^{\infty} \lambda_j \Phi_j(r) \Phi_j(s) \right] J'_{c_m}(r) aa' J_{c_m}(s) \, dr ds.$$

Since $\sum_{j=1}^{M} \lambda_j \Phi_j(r) \Phi_j(s)$ converges to $Q^*_h(r,s)$ uniformly in (r,s) as $M \to \infty$, we have: for any $\varepsilon > 0$, there exist $M^* > 0$ such that $\left| \sum_{j=M+1}^{\infty} \lambda_j \Phi_j(r) \Phi_j(s) \right| < \varepsilon$ for all $M > M^*$. So when M is large enough,

$$\left| \xi^*_{\infty,M} - \xi^*_{\infty,\infty} \right| \leq \varepsilon \int_0^1 \int_0^1 \left| J'_{c_m}(r) aa' J_{c_m}(s) \right| dr ds$$

$$\leq \varepsilon \left(\int_0^1 \left| J'_{c_m}(r) a \right| dr \right)^2 \leq \varepsilon \left(\int_0^1 \| J_{c_m}(r) \| dr \right)^2 \|a\|^2$$

$$\leq \varepsilon \left(\int_0^1 \| J_{c_m}(r) \|^2 dr \right) \|a\| = \varepsilon O_p(1).$$

Since ε is arbitrary, we have $\xi^*_{\infty,M} - \xi^*_{\infty,\infty} = o_p(1)$ as $M \to \infty$. Hence condition (ii) of Lemma 1 holds. It is easy to see that Condition (iii) of Lemma 1 also holds. It remains to verify the last condition in Lemma 1. We have

$$E |\eta_{T,M}| \leq \frac{\varepsilon}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} E \left| e'_t aa' e_s \right| \leq \frac{\varepsilon}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} \left[E \left(e'_t a \right)^2 \right]^{1/2} \left[E \left(a' e_s \right)^2 \right]^{1/2}$$

$$= \frac{\varepsilon}{T^3} \left\{ \sum_{t=1}^{T} \left[E \left(e'_t a \right)^2 \right]^{1/2} \right\}^2.$$

26

In view of

$$\frac{e'_t a}{\sqrt{T}} = \frac{1}{\sqrt{T}} \sum_{j=0}^{t-1} \rho^j_{T,m} u'_{t-j} a + \frac{1}{\sqrt{T}} \rho^t_{T,m} e'_0 a,$$

and defining $\Gamma_u(j_1 - j_2) := E u_{t-j_1} u'_{t-j_2}$, we have

$$
\begin{aligned}
var\left(\frac{e'_t a}{\sqrt{T}}\right) &\le 2var\left(\frac{1}{\sqrt{T}} \sum_{j=0}^{t-1} \rho^j_{T,m} u'_{t-j} a\right) + \frac{2}{T} var(e_0) \\
&= \frac{2}{T} \sum_{j_1=0}^{t-1} \sum_{j_2=0}^{t-1} \rho^{j_1}_{T,m} \rho^{j_2}_{T,m} a' u_{t-j_1} u'_{t-j_2} a + O\left(\frac{1}{T}\right) \\
&\le \frac{2}{T} \sum_{j_1=0}^{t-1} \sum_{j_2=0}^{t-1} |a' \Gamma_u(j_1 - j_2) a| + O\left(\frac{1}{T}\right) \\
&\le \frac{2}{T} \sum_{j_1=0}^{t-1} \sum_{j_2=0}^{t-1} \|\Gamma_u(j_1 - j_2)\| \|a\|^2 + O\left(\frac{1}{T}\right) = O(1)
\end{aligned}
$$

uniformly in T. Hence

$$E|\eta_{T,M}| \le C \frac{\varepsilon}{T^3} \left\{\sum_{t=1}^T T^{1/2}\right\}^2 \le C\varepsilon$$

for some constant C that does not depend on T. This implies that, for every $\delta > 0$, $\sup_T P(|\eta_{T,M}| > \delta) \xrightarrow{p} 0$ as $M \to \infty$.

Given that all conditions in Lemma 1 hold, we have

$$\omega_T := \frac{1}{T^3} \sum_{t=1}^T \sum_{s=1}^T Q^*_h\left(\frac{t}{T}, \frac{s}{T}\right) e'_t a a' e_s \xrightarrow{d} \xi^*_{\infty,\infty} := \int_0^1 \int_0^1 Q^*_h(r,s) J'_{c_m}(r) a a' J_{c_m}(s) \, ds.$$

Consequently,

$$
\begin{aligned}
F_T &= \left(\frac{1}{T} \sum_{t=1}^T \frac{y_t}{\sqrt{T}}\right)' \left[\frac{1}{T^3} \sum_{t=1}^T \sum_{s=1}^T Q^*_h\left(\frac{t}{T}, \frac{s}{T}\right) (y_t - \bar{y})(y_s - \bar{y})'\right]^{-1} \left(\frac{1}{T} \sum_{t=1}^T \frac{y_t}{\sqrt{T}}\right) / p \\
&\xrightarrow{d} \left(\int_0^1 J_c(r) \, dr\right)' \left(\int_0^1 \int_0^1 Q^*_h(t,s) J_c(r) J_c(s) \, dr ds\right)^{-1} \left(\int_0^1 J_c(r) \, dr\right) / p
\end{aligned}
$$

as desired.

The proof for t-statistic t_T is similar and is omitted here. ∎

Proof of Theorem 2. In view of $J_{c_m}(t) = \int_0^t e^{-c_m(t-s)} dW_p(s)$, we have

$$
\begin{aligned}
c_m &\int_0^1 \Phi_j(r) J_{c_m}(r) \, dr \\
&= \int_0^1 \left[c_m \int_0^r \Phi_j(r) e^{-c_m(r-s)} dW_p(s)\right] dr \\
&= \int_0^1 \left[c_m \int_s^1 \Phi_j(r) e^{-c_m(r-s)} dr\right] dW_p(s).
\end{aligned}
$$

But

$$c_m \int_s^1 \Phi_j(r) e^{-c_m(r-s)} dr$$

$$= -\int_s^1 \Phi_j(r) d\left[e^{-c_m(r-s)}\right]$$

$$= -\Phi_j(r) e^{-c_m(r-s)}|_s^1 + \int_s^1 e^{-c_m(r-s)}\Phi_j'(r) dr$$

$$= \Phi_j(s) - \Phi_j(1) e^{-c_m(1-s)} + \int_s^1 e^{-c_m(r-s)}\Phi_j'(r) dr,$$

and so

$$c_m \int_0^1 \Phi_j(r) J_{c_m}(r) dr = \int_0^1 \Phi_j(s) dW_p(s) - \Phi_j(1) \int_0^1 e^{-c_m(1-s)} dW_p(s)$$

$$+ \int_0^1 \left(\int_s^1 e^{-c_m(r-s)}\Phi_j'(r) dr\right) dW_p(s).$$

We now show that the last two terms in the above expression are $o_p(1)$. First, since $E \int_0^1 e^{-c_m(1-s)} dW_p(s) = 0$ and

$$var\left[\int_0^1 e^{-c_m(1-s)} dW_p(s)\right] = \int_0^1 e^{-2c_m(1-s)} ds = \frac{1}{2c_m}\left(1 - e^{-2c_m}\right) \to 0 \text{ as } c_m \to \infty,$$

we have

$$\Phi_j(1) \int_0^1 e^{-c_m(1-s)} dW_p(s) = \Phi_j(1) o_p(1)$$

uniformly over $j = 0, 1, \ldots$

Second, $E \int_0^1 \left(\int_s^1 e^{-c_m(r-s)}\Phi_j'(r) dr\right) dW_p(s) = 0$ and

$$var\left[\int_0^1 \left(\int_s^1 e^{-c_m(r-s)}\Phi_j'(r) dr\right) dW_p(s)\right]$$

$$= \int_0^1 \left(\int_s^1 e^{-c_m(r-s)}\Phi_j'(r) dr\right)^2 ds$$

$$\leq \int_0^1 \left(\int_s^1 e^{-2c_m(r-s)} dr\right)\left(\int_s^1 \left[\Phi_j'(r)\right]^2 dr\right) ds$$

$$\leq \left[\int_0^1 \left(\int_s^1 e^{-2c_m(r-s)} dr\right) ds\right]\left(\int_0^1 \left[\Phi_j'(r)\right]^2 dr\right)$$

$$= \left[\frac{1}{2c_m}\int_0^1 \left(1 - e^{-2c_m(1-s)}\right) ds\right]\left(\int_0^1 \left[\Phi_j'(r)\right]^2 dr\right)$$

$$\leq \frac{1}{2c_m}\left(\int_0^1 \left[\Phi_j'(r)\right]^2 dr\right).$$

This implies that

$$\int_0^1 \left(\int_s^1 e^{-c_m(r-s)}\Phi_j'(r) dr\right) dW_p(s) = \left\|\Phi_j'(r)\right\|_2 o_p(1).$$

28

When $\Phi'_j(r) \neq 0$ for large enough j as in the kernel case, we need to strengthen the above result. In this case,

$$\Phi'_j(r) = \begin{cases} -\frac{1}{2}\pi j \sin \frac{1}{2}\pi jr, & j \text{ is even} \\ \frac{1}{2}\pi(j+1)\cos \frac{1}{2}\pi r(j+1), & j \text{ is odd} \end{cases}$$

When $\Phi'_j(r) = -\pi\ell \sin \pi\ell r$ for $j = 2\ell$, we have

$$\left| \int_s^1 e^{-c_m(r-s)}(\sin(\pi\ell r))\,dr \right|$$

$$\leq \left| \frac{1}{c_m^2 + \pi^2\ell^2} \left[c_m e^{-c_m(1-s)}\sin \pi\ell + \pi\ell e^{-c_m(1-s)}\cos \pi\ell \right] \right|$$

$$+ \left| \frac{1}{c_m^2 + \pi^2\ell^2} \left[c_m \sin \pi\ell s + \pi\ell\cos \pi\ell s \right] \right|$$

$$\leq \frac{2(c_m + \pi\ell)}{c_m^2 + \pi^2\ell^2}$$

and

$$\int_0^1 \left(\int_s^1 e^{-c_m(r-s)}\sin(\pi kr)\,dr \right)^2 ds \leq \frac{4(c_m + \pi\ell)^2}{(c_m^2 + \pi^2\ell^2)^2} \leq \frac{8}{c_m^2 + \pi^2\ell^2}.$$

Hence

$$\int_0^1 \left(\int_s^1 e^{-c_m(r-s)}\Phi'_j(r)\,dr \right)^2 ds \leq \frac{8\pi^2\ell^2}{c_m^2 + \pi^2\ell^2} = \frac{2\pi^2 j^2}{\pi^2 j^2 + 4c_m^2}. \tag{26}$$

Similarly, when $\Phi'_j(r) = \pi\ell\cos \pi r\ell$ for $j = 2\ell - 1$, we have

$$\left| \int_s^1 e^{-c_m(r-s)}(\cos \pi r\ell)\,dr \right|$$

$$\leq \left| \frac{c_m(\cos \pi\ell)e^{-c_m(1-s)} - \pi\ell(\sin \pi\ell)e^{-c_m(1-s)}}{c_m^2 + \pi^2\ell^2} \right|$$

$$+ \left| \frac{c_m(\cos \pi s\ell) - \pi\ell(\sin \pi s\ell)}{c_m^2 + \pi^2\ell^2} \right|$$

$$\leq \frac{2(c_m + \pi\ell)}{c_m^2 + \pi^2\ell^2}.$$

Therefore

$$\int_0^1 \left(\int_s^1 e^{-c_m(r-s)}\Phi'_j(r)\,dr \right)^2 ds \leq \frac{2\pi(j+1)^2}{\pi^2(j+1)^2 + 4c_m^2}. \tag{27}$$

Combining (26) with (27), we have

$$\int_0^1 \left(\int_s^1 e^{-c_m(r-s)}\Phi'_j(r)\,dr \right) dW_p(s) = O_p\left(\sqrt{\frac{\pi^2 j^2}{\pi^2 j^2 + 4c_m^2}} \right).$$

To sum up, we have proved the following

$$c_m \int_0^1 \Phi_j(r) J_{c_m}(r)\,dr = \int_0^1 \Phi_j(s)\,dW_p(s) + \left[|\Phi_j(1)| + \|\Phi'_j(r)\|_2^2 \right] o_p(1),$$

29

which holds for both the series case and the kernel case. In the kernel case, we have proved:

$$c_m \int_0^1 \Phi_j(r) J_{c_m}(r) \, dr = \int_0^1 \Phi_j(s) \, dW_p(s) + |\Phi_j(1)| \, o_p(1) + O_p\left(\sqrt{\frac{\pi^2 j^2}{\pi^2 j^2 + 4 c_m^2}}\right).$$

Both results hold uniformly over $j = 1, 2, \ldots$.

It follows that

$$c_m \int_0^1 \Phi_j(r) J_{c_m}(r) \, dr = \int_0^1 \Phi_j(s) \, dW_p(s) + o_p(1) \text{ as } c_m \to \infty$$

uniformly over $j = 0, 1, \ldots$ for both the series case and the kernel case. Therefore,

$$\int_0^1 \int_0^1 Q_h^*(r,s) \, c_m J_{c_m}(r) \, c_m J'_{c_m}(s) \, dr ds$$

$$= \int_0^1 \int_0^1 \sum_{j=1}^\infty \lambda_j \Phi_j(r) \Phi_j(s) \, c_m J_{c_m}(r) \, c_m J'_{c_m}(s) \, dr ds$$

$$= \sum_{j=1}^\infty \lambda_j \left[c_m \int_0^1 \Phi_j(r) J_{c_m}(r) \, dr \right] \left[c_m \int_0^1 \Phi_j(s) J_{c_m}(s) \, ds \right]'$$

$$= \sum_{j=1}^\infty \lambda_j \left[\int_0^1 \Phi_j(r) \, dW_p(r) \right] \left[\int_0^1 \Phi_j(s) \, dW_p(s) \right]' + \left[\sum_{j=1}^\infty |\lambda_j| \right] o_p(1)$$

$$= \int_0^1 \int_0^1 \left[\sum_{j=1}^\infty \lambda_j \Phi_j(r) \Phi_j(s) \right] dW_p(r) \, dW'_p(s) + o_p(1)$$

$$= \int_0^1 \int_0^1 Q_h(r,s) \, dW_p(r) \, dW'_p(s) + o_p(1).$$

As a consequence,

$$F_\infty(c_m)$$

$$= \left[\int_0^1 \Phi_0(s) \, dW_p(s) + o_p(1) \right]' \times \left[\int_0^1 \int_0^1 Q_h(r,s) \, dW_p(r) \, dW'_p(s) + o_p(1) \right]^{-1}$$

$$\times \left[\int_0^1 \Phi_0(s) \, dW_p(s) + o_p(1) \right] / p$$

$$= F_\infty(\infty) + o_p(1) \text{ as } c_m \to \infty.$$

The proof for $t_\infty(c_m)$ is similar and is omitted here. ∎

30

References

[1] Alexopoulos, C. (2006): "A Comprehensive Reviews of Methods For Simulation Output Analysis." *Proceedings of the 2006 Winter Simulation Conference.*

[2] Alexopoulos, C. (2007): "Statistical Analysis of Simulation Output: State of the Art." *Proceedings of the 2007 Winter Simulation Conference.*

[3] Davidson, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford University Press.

[4] Giraitis, L., and Phillips, P.C.B. (2006): "Uniform limit theory for stationary autoregression." *Journal of Time Series Analysis* 27, 51–60.

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

[6] Kiefer, N.M., T.J. Vogelsang and H. Bunzel (2000): "Simple Robust Testing of Regression Hypotheses." *Econometrica* 68, 695–714.

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

[8] ——— (2002b): "Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation." *Econometrica* 70, 2093–2095.

[9] ——— (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests." *Econometric Theory* 21, 1130–1164.

[10] Park, J. (2003): "Weak Unit Roots." Working paper, Department of Economics, Indiana University.

[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*, forthcoming.

[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. (1987): "Towards a Unified Asymptotic Theory for Autoregression." *Biometrika* 74, 535–547.

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

[16] Phillips, P.C.B. and S. N. Durlauf (1986): "Multiple Regression with Integrated Processes." *Review of Economic Studies* 53, 473–496.

[17] Phillips, P.C.B. and T. Magdalinos (2007): "Limit theory for Moderate Deviations from a Unit Root". *Journal of Econometrics* 136, 115–130.

[18] Phillips, P.C.B., T. Magdalinos and L. Giraitis (2010): "Smoothing Local-to-moderate Unit Root Theory." *Journal of Econometrics* 158, 274–279.

[19] Phillips, P.C.B. and V. Solo (1992): "Asymptotics for Linear Processes." *Annals of Statistics*, 20(2), 971–1001.

[20] Politis, D. (2011): "Higher-order Accurate, Positive Semi-definite Estimation of Large-sample Covariance and Spectral Density Matrices." *Econometric Theory* 27, 703–744.

[21] Sun, Y. (2004): "A Convergent t-statistic in Spurious Regressions." *Econometric Theory* 20, 943–962.

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

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

[24] Sun, Y. (2013b): "Fixed-smoothing Asymptotics in a Two-step GMM Framework." Working paper, Department of Economics, UC San Diego.

[25] Sun, Y. (2014): Comment on "HAC Corrections for Strongly Autocorrelated Time Series" by U.K. Müller, forthcoming in *Journal of Business & Economic Statistics*.

[26] Sun, Y. and M.S. Kim (2013): "Asymptotic F Test in a GMM Framework with Cross Sectional Dependence." *Review of Economics and Statistics,* forthcoming.

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