

# On Testing-Optimal Smoothing Parameter Choice in Robust Multivariate Trend Inference

Yixiao Sun\*  
Department of Economics,  
University of California, San Diego

July 15, 2009

---

\*Email: [yisun@ucsd.edu](mailto:yisun@ucsd.edu). The author 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.

## ABSTRACT

The paper develops a novel testing procedure for hypotheses on deterministic trends in a multivariate trend stationary model. The trends are estimated by the OLS estimator and the long run variance (LRV) matrix is estimated by a multiple window estimator with carefully selected data windows. The multiple window estimator is asymptotically invariant to the model parameters and thus does not suffer from the usual demeaning bias that hurts the performance of conventional kernel LRV estimators. The number of data windows  $K$ , the underlying smoothing parameter, plays a key role in determining the asymptotic properties of the long run variance estimator and the associated semiparametric tests. When  $K$  is fixed, the modified Wald statistic converges to an F-distribution while when  $K$  grows with the sample size, the Wald statistic converges to a Chi-square distribution. We show that the critical values from the fixed- $K$  asymptotics are second order correct under the large- $K$  asymptotics. We propose a novel approach to select  $K$  which minimizes the type II error hence maximizes the power of the test while controlling for the type I error. The new selection rule is fundamentally different from the conventional rule based on the mean square error criterion. A plug-in procedure for implementing the new rule is suggested and simulations show that the new plug-in procedure works remarkably well in finite samples.

*JEL Classification:* C13; C14; C32; C51

*Keywords:* Asymptotic expansion, F-distribution, Hotelling  $T^2$ , long-run variance, multiple window estimator, robust standard error, test-optimal smoothing parameter choice, trend inference, Type I and Type II errors.

# 1 Introduction

Trend regression is one of simple and important regressions in economic and climatic time series analysis. In this paper, we consider a linear trend regression with multiple dependent variables. For example, the dependent variables may consist of DGPs from a number of countries. Vogelsang and Franses (2005) provide more empirical examples. Estimation of the trends is relatively easy as the equation-by-equation OLS estimator is asymptotically as efficient as the system GLS estimator. Hence, for point estimation, there is no need to take error autocorrelation into account in large samples. However, trend inference is subtle as the variance of the OLS trend estimator depends on the long run variance (LRV) of the error process. Since the LRV is proportional to the spectral density of the error process evaluated at zero, many nonparametric spectral density methods can be used to estimate the LRV. In this paper, we focus on multiple window estimators, a very powerful class of spectral estimators proving to be very valuable in the signal processing literature whenever the spectrum of interest is detailed or varies rapidly with a large dynamic range. See Percival and Walden (1993, ch. 7). The multiple window estimators also produce asymptotically valid tests on trend hypotheses that are very easy to use in practice.

The smoothing parameter in the multiple window estimator is the number of data windows or basis functions employed. When the number of basis functions  $K$  is fixed, the LRV estimator is inconsistent and converges 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 assumption that  $K$  is fixed, we can show that the Wald statistic converges to a nonstandard distribution. The fixed- $K$  asymptotics is in the spirit of the fixed- $b$  asymptotics as in Kiefer and Vogelsang (2005). This type of asymptotics captures the randomness of the LRV estimator and tests based on it often have better finite sample size properties than tests based on consistent LRV estimates.

The novelty here is that we design a set of basis functions so that the nonstandard limiting distribution becomes the standard  $F$  distribution. For these basis functions, the LRV estimator is asymptotically invariant to the intercepts and trend parameters. As a result, the LRV estimator does not suffer from the bias arising from the estimation uncertainty of model parameters. This is contrast with the conventional kernel LRV estimators where the estimation uncertainty gives rise to a demeaning bias. This type of bias is of the same order of magnitude as the asymptotic variance, see Sun, Phillips and Jin (2008, SPJ hereafter). Under the asymptotic mean-squared error (MSE) criterion, the squared form of the bias is of smaller order than the variance. Hence the demeaning bias is asymptotically negligible under the MSE criterion. However, for testing problems, this type of bias will have the same order effect on the size and power properties as the variance of the LRV estimator. By selecting the basis functions appropriately, we completely remove this type of bias to the order we care about. This is a desirable property as we generally prefer an estimator with fewer bias terms. Another great advantage of using the new LRV estimator is that the critical values from the fixed- $K$  asymptotics are readily available from statistical tables and econometrics software programs. The computational burden of simulating critical values from nonstandard distributions is completely removed.

While the LRV estimator is inconsistent when  $K$  is finite, it becomes consistent when  $K$  grows with the sample size at a certain rate. The smoothing parameter  $K$  is an important

tuning parameter that determines the asymptotic properties of the LRV estimator. Following the conventional approaches (e.g., Andrews, 1991, and Newey and West, 1987, 1994), Phillips (2005) chooses the smoothing parameter  $K$  to minimize the asymptotic MSE of the LRV estimator. This approach follows what has long been standard practice in the context of spectral estimation (Grenander and Rosenblatt, 1957; Hannan, 1970) where the focus of attention is the spectrum or the LRV. Such a choice of the smoothing parameter is designed to be optimal in the MSE sense for the point estimation of the spectrum or LRV, but is not necessarily best suited for semiparametric testing. Through its effect on the LRV estimator, the smoothing parameter  $K$  affects the type I and type II errors of the associate test. It therefore seems sensible that the choice of  $K$  should take these properties into account.

To develop an optimal choice of  $K$  for semiparametric testing, we first have to decide on which test to use. There are two choices. One is the traditional Wald test which is based on the Wald statistic and uses critical values from a chi-square distribution. The other is the new  $F^*$  test given in this paper, which is based on the modified Wald statistic and uses critical values from an F-distribution. One of the main contributions of the paper is to show that the critical values from the F-distribution are higher order correct under the conventional large  $K$  asymptotics. A direct implication is that the  $F^*$  test has smaller size distortion than the traditional Wald test. On the basis of this theoretical result and the emphasis on the size control in the econometrics literature, we employ and recommend the  $F^*$  test to conduct inference on the trend parameters.

Another main contribution of the paper is to develop an optimal procedure for selecting the smoothing parameter  $K$  that addresses the central concerns of semiparametric testing. For testing problems, we do not care about the LRV estimator per se. Instead, we are interested in the LRV estimator only to use it to construct the  $F^*$  statistic. The ultimate goal of any testing problem is to minimize the type II error hence maximize the power while controlling for the type I error. It is thus desirable to choose the smoothing parameter to achieve this goal. We propose to choose  $K$  to minimize the type II error subject to the constraint that the type I error is bounded. The resulting optimal  $K$  is said to be test-optimal for the given bound. The bound is defined to be  $\kappa\alpha$ , where  $\alpha$  is the nominal type I error and  $\kappa > 1$  is the parameter that captures the user's tolerance on the discrepancy between the nominal and true type I errors. The parameter  $\kappa$  is allowed to be sample size dependent. For a smaller sample size, we may have a higher tolerance while for the larger sample size we may have a lower tolerance. The introduction of the tolerance parameter into the optimal  $K$  selection is a conceptually new idea, which does not seem to appear elsewhere in the literature.

The proposed approach to selecting the test-optimal  $K$  requires asymptotic measurements of type I and type II errors of the  $F^*$  test. These measurements are provided by means of high order asymptotic expansions of the finite sample distribution of the  $F^*$  statistic under the null and local alternative hypotheses. We show that the type I error depends on the nonparametric bias of the LRV estimator. When the nonparametric bias is negative, the type I error is expected to be larger than the nominal type I error. The result is supported by simulations in Vogelsang and Franses (2005) as well as the present paper. The type II error of the  $F^*$  test depends on the local alternative hypothesis through a noncentrality parameter  $\|\tilde{c}\|^2$ , where  $\tilde{c}$  is a vector that characterizes the local departure of the alternative hypothesis from the null. To the first order, the type II error depends on

$\tilde{c}$  only through its squared length  $\|\tilde{c}\|^2$ . So it is reasonable to assume that  $\tilde{c}$  is uniformly distributed on a sphere. This assumption greatly facilitates the higher order expansion under the local alternative hypothesis. In a transformed space, the null hypothesis is a fixed point while the alternative hypothesis is a random point uniformly distributed on the sphere centered at the fixed null. The radius of the sphere is chosen so that the power of the test is 75% under the first order asymptotics. This strategy is similar to that used in the optimal testing literature. In the absence of a uniformly most powerful test, it is often recommended to pick a reasonable point under the alternative and construct an optimal test against this particular point alternative. It is hoped that the resulting test, although not uniformly most powerful, is reasonably close to the power envelope. Here we use the same idea and select the radius of the sphere according to the power requirement. We hope that the smoothing parameter that is optimal for the chosen radius also works well for other points under the alternative hypothesis. This is confirmed by our Monte Carlo study.

There are several important differences between the test-optimal  $K$  and the MSE-optimal  $K$ . First, the test-optimal  $K$  depends on the average bias, a measure that aggregates the bias matrix of the LRV matrix estimator. This is in sharp contrast with the MSE-optimal  $K$  which depends on the squared bias of the LRV estimator. In large samples, the squared bias is of smaller order than the bias itself. So for testing problem, it is more important to reduce the bias of the LRV estimator as compared to the point estimation of the LRV matrix. Second, the expansion rate of the test-optimal  $K$  depends on the rate of expansion of the Lagrange multiplier for the constraint on the type I error, which in turn depends on the rate of the tolerance parameter  $\kappa$  approaching 1. When the Lagrange multiplier is finite, the test-optimal  $K$  grows with the sample size  $T$  at the rate of  $O(T^{2/3})$ . In contrast, the MSE-optimal  $K$  grows with the sample size at the rate of  $O(T^{4/5})$ . As a result, when the Lagrange multiplier is finite, the test-optimal  $K$  diverges to infinity at a slower rate than the conventional MSE-optimal  $K$  given in Phillips (2005) and the current paper. On the other hand, when the Lagrange multiplier grows with the sample size, the test-optimal  $K$  can be bounded or grows faster than the MSE-optimal rate of  $O(T^{4/5})$ . The fixed- $K$  rule can be interpreted as assigning increasingly more weight to the type I error as the same size increases. Finally, the test-optimal  $K$  depends on the hypotheses being tested via the restriction matrix  $R$  in the null hypothesis  $H_0 : R\beta = r$  while the MSE-optimal  $K$  does not depend on  $R$ . Our criterion for  $K$  selection is a testing-focused criterion in that it aims at the testing problem and takes the specific hypothesis into consideration.

The paper that is most closely related to the present paper is SPJ. The asymptotic expansions given here are related to but are more difficult to establish than those for Gaussian location models in SPJ. The reason is that, unlike the simple  $t$  statistic, the Wald statistic can not be separated into a numerator and a denominator. We confront this challenge by using the independence of the length of a standard normal vector from its direction. The method for selecting the test-optimal  $K$  is also closely related. In SPJ, the optimal  $K$  minimizes a loss function that is defined to be a weighted sum of the type I and type II errors. Our procedure can also be cast in this framework with the Lagrange multiplier being the relative weight. The main difference is that our weight is implicitly defined through the tolerance parameter  $\kappa$ . For a given  $\kappa$ , the weight may be different across different data generating processes. In contrast, in the SPJ procedure, the weight is specified a priori and is thus fixed. Both procedures require a user-chosen parameter: the

tolerance parameter or the weight. The tolerance parameter is often easier to choose as it involves only the type I error while the weight is more difficult to choose as it depends on both type I and type II errors. This is an advantage of the new procedure proposed here.

The rest of the paper is organized as follows. Section 2 describes the basic setting and the limiting distribution of the trend estimator. Section 3 motivates the new LRV estimator and establishes its asymptotic properties under the fixed- $K$  and large- $K$  asymptotics. This section also discusses the choice of the basis functions. Section 4 investigates the Wald test under both fixed- $K$  and large- $K$  asymptotics. Section 5 gives a higher order expansion of the finite sample distribution of the modified Wald statistic. On the basis of this expansion, the next section proposes a selection rule for  $K$  that is most suitable for implementation in semiparametric testing. The subsequent section reports simulation evidence on the performance of the new procedure. The last section provides some concluding discussion. Proofs and additional technical results are given in the Appendix.

## 2 The Model and Preliminaries

Consider  $n$  trend-stationary time series denoted by  $(y_{1t}, \dots, y_{nt})'$  with  $t = 1, 2, \dots, T$ . We assume that the data generating process is

$$y_{it} = \alpha_i + \beta_i t + u_{it}, t = 1, 2, \dots, T, i = 1, 2, \dots, n, \quad (1)$$

where  $u_{it}$  is a weakly dependent process with zero mean.

**Assumption 1** Let  $u_t = (u_{1t}, \dots, u_{nt})'$ , we assume that

$$u_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j},$$

where  $\varepsilon_t \sim iid(0, \Sigma)$ ,  $E \|\varepsilon_t\|^v < \infty$  for some  $v \geq 4$ ,

$$\sum_{j=0}^{\infty} j^a \|C_j\| < \infty \text{ for } a > 3, C(1)\Sigma C(1)' > 0$$

and  $\|\cdot\|$  is the matrix Euclidean norm.

Under the above assumption, the process  $u_t$  admits the following BN decomposition

$$u_t = C(1)\varepsilon_t + \tilde{u}_{t-1} - \tilde{u}_t \text{ for } \tilde{u}_t = \sum_{j=0}^{\infty} \tilde{C}_j \varepsilon_{t-j}, \tilde{C}_j = \sum_{s=j+1}^{\infty} C_s, \quad (2)$$

where  $\sum_{j=0}^{\infty} \|\tilde{C}_j\|^2 < \infty$ . Using this decomposition and following Phillips and Solo (1992), we can prove that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t \rightarrow^d \Lambda W_n(r), \text{ as } T \rightarrow \infty, \quad (3)$$

where  $W_n(r)$  is an  $n \times 1$  vector of standard independent Wiener processes and  $\Lambda = [C(1)\Sigma C(1)']^{1/2}$  is the matrix square root of the long run variance matrix  $\Omega$  of  $u_t$  :

$$\Omega = \Lambda\Lambda' = \sum_{j=-\infty}^{\infty} Eu_t u'_{t-j} = C(1)\Sigma C(1)'$$

To represent the OLS estimator of the model parameters, we introduce the following notation:

$$\begin{aligned} y_i &= (y_{i1}, \dots, y_{iT})', Y = (y_1, y_2, \dots, y_n) \\ u_i &= (u_{i1}, \dots, u_{iT})', u = (u_1, \dots, u_n) \\ X_t &= (1, t), X = (X'_1, \dots, X'_T)' \\ \theta &= (\theta_1, \theta_2, \dots, \theta_n) \text{ with } \theta_i = (\alpha_i, \beta_i)'. \end{aligned}$$

The OLS estimator of  $\theta$  is then given by

$$\hat{\theta}_{OLS} = (X'X)^{-1}X'Y.$$

If the errors are second-order stationary, then the OLS estimator is asymptotically equivalent to the GLS estimator. In addition, because (1) is a seemingly unrelated regression (SUR) with the same regressors in each equation, the OLS estimator is equivalent to the SUR estimator, which is the GLS estimator that accounts for contemporaneous correlation across the series. Thus, the simple OLS estimator has some nice optimality properties. Vogelsang and Franses (2005) make the same point.

Let  $D = \text{diag}(T^{-1/2}, T^{-3/2})$ . Then

$$\begin{aligned} DX'XD &= \begin{pmatrix} 1 & \frac{1}{T^2} \sum_{t=1}^T t \\ \frac{1}{T^2} \sum_{t=1}^T t & \frac{1}{T^3} \sum_{t=1}^T t^2 \end{pmatrix} \rightarrow_d \begin{pmatrix} 1 & \int_0^1 r dr \\ \int_0^1 r dr & \int_0^1 r^2 dr \end{pmatrix}, \\ DX'Y &= \begin{pmatrix} \frac{1}{\sqrt{T}} \sum u'_t \\ \frac{1}{T\sqrt{T}} \sum t u'_t \end{pmatrix} \rightarrow_d \begin{pmatrix} \int_0^1 dW'_n(r) \Lambda' \\ \int_0^1 r dW'_n(r) \Lambda' \end{pmatrix}. \end{aligned}$$

Therefore

$$D(\hat{\theta}_{OLS} - \theta) \rightarrow \begin{pmatrix} 1 & \int_0^1 r dr \\ \int_0^1 r dr & \int_0^1 r^2 dr \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 dW'_n(r) \Lambda' \\ \int_0^1 r dW'_n(r) \Lambda' \end{pmatrix} = \begin{pmatrix} 6 \int_0^1 (\frac{2}{3} - r) dW'_n(r) \Lambda' \\ 12 \int_0^1 (r - \frac{1}{2}) dW'_n(r) \Lambda' \end{pmatrix}.$$

So the OLS estimator  $\hat{\beta}_{OLS}$  of  $\beta$  satisfies

$$T^{3/2}(\hat{\beta}_{OLS} - \beta) \rightarrow_d 12\Lambda \int_0^1 \left(r - \frac{1}{2}\right) dW_n(r) =^d N(0, 12\Omega).$$

### 3 Multiple Window Estimator and its Asymptotic Properties

To conduct inference regarding  $\beta$ , we need to first estimate the long run variance matrix  $\Omega$ . Many nonparametric estimation methods are available in both statistics and econometrics literature. In the econometrics literature, the LRV estimators are usually called HAC estimators following Newey and West (1987, 1994) and Andrews (1991) who extend spectral density estimation methods to more general settings.

### 3.1 Motivation of Multiple Window Estimators

To motivate the multiple window estimator, we consider the kernel-based estimator proposed by Phillips, Sun and Jin (2006, 2007, PSJ hereafter). It is given by

$$\hat{\Omega}_{PSJ} = \frac{1}{T} \sum_{r=1}^T \sum_{t=1}^T \hat{u}_t \mathcal{K}_\rho \left( \frac{r-s}{T} \right) \hat{u}'_s,$$

where  $\mathcal{K}_\rho(r-s) = [\mathcal{K}(r-s)]^\rho$  for some kernel function  $\mathcal{K}(\cdot)$ . This estimator is consistent when  $\rho \rightarrow \infty$  at a certain rate. By the Mercer's theorem, we can write

$$\mathcal{K}_\rho(r-s) = \sum_{k=1}^{\infty} \lambda_k \phi_k(r) \phi_k(s), \quad (4)$$

where  $\{\lambda_k\}$  is a sequence of eigenvalues satisfying  $\sum_{k=1}^{\infty} \lambda_k = 1$  and  $\{\phi_k(r)\}$  is an orthonormal sequence of eigenfunctions corresponding to the eigenvalues  $\lambda_k$ . With this representation of  $\mathcal{K}_\rho(r-s)$ , we can write

$$\begin{aligned} \hat{\Omega}_{PSJ} &= \sum_{k=1}^{\infty} \frac{1}{T} \sum_{r=1}^T \sum_{s=1}^T \lambda_k \hat{u}_r \phi_k \left( \frac{r}{T} \right) \phi_k \left( \frac{s}{T} \right) \hat{u}'_s \\ &= \sum_{k=1}^{\infty} \lambda_k \left[ \frac{1}{\sqrt{T}} \sum_{r=1}^T \phi_k \left( \frac{r}{T} \right) \hat{u}_r \right] \left[ \frac{1}{\sqrt{T}} \sum_{s=1}^T \phi_k \left( \frac{s}{T} \right) \hat{u}_s \right]'. \end{aligned} \quad (5)$$

In the above expression,  $\lambda_k$  decays to zero as  $k$  increases. The intuition is that, as  $k$  increases, the eigenfunction  $\phi_k(r)$  becomes more concentrated on high frequency components and we should impose progressively less weight on these components in order to capture the long run properties of the underlying time series. Implicitly, the PSJ estimator employs a soft thresholding method where the weight  $\lambda_k$  approaches to zero but is not equal to zero for any given  $k$ . Instead of soft thresholding, we can also consider the hard thresholding estimator:

$$\hat{\Omega} = \frac{1}{K} \sum_{k=1}^K \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^T \phi_k \left( \frac{t}{T} \right) \hat{u}_t \right] \left[ \frac{1}{\sqrt{T}} \sum_{s=1}^T \phi_k \left( \frac{s}{T} \right) \hat{u}_s \right]' := \frac{1}{K} \sum_{k=1}^K \hat{\Omega}_k, \quad (6)$$

where  $K$  is a positive integer. This estimator truncates the infinite sum in (5) and assigns equal weights to the remaining terms. In other words, the infinite sequence  $(\lambda_1, \dots, \lambda_K, \dots)$  is replaced by  $(1/K, 1/K, \dots, 1/K, 0, \dots)$ . As will be shown below, with approximately chosen  $\phi_k$ , each of the summand  $\hat{\Omega}_k$  is an unbiased estimator of  $\Omega$ . We refer to the LRV estimators of the form  $\hat{\Omega}_k$  as direct LRV estimators so that  $\hat{\Omega}$  is an average of  $K$  direct LRV estimators.

Interestingly, the so-obtained estimator belongs to the class of multiple window estimators, a term we now clarify. For a given stationary and mean zero time series  $x_1, x_2, \dots, x_T$ , a multiple window estimator of its spectral density at frequency  $\omega$  is defined to be

$$\hat{S}(\omega) = \sum_{k=1}^K a_k \left| \sum_{t=1}^T x_t h_t^{(k)} \exp(-\omega i) \right|^2, \quad (7)$$

where  $a_k$  is a constant,  $h^{(k)} = (h_1^{(k)}, h_2^{(k)}, \dots, h_T^{(k)})$  is a sequence of constants called a data window or basis function, and  $K$  is the number of data windows used. In the multiple window spectral analysis of Thomson (1982),  $K$  is normally chosen to be much smaller than the sample size. According to this definition,  $\hat{\Omega}$  is a multiple window estimator at frequency  $\omega = 0$ . The underlying data window is  $\phi_k(\cdot)$  and the number of data windows is  $K$ .

Multiple window estimators have been used in some papers in econometrics. Sun (2006) applies the multiple window estimator to the estimation of realized volatility. The robust long-run variance estimators derived by Müller (2007) also belong to the class of multiple window estimators. Phillips (2005) gives alternative motivation of the multiple window estimator and establishes its asymptotic properties.

### 3.2 Fixed-K Asymptotics

In this subsection, we establish the asymptotic distribution of  $\hat{\Omega}$  under the assumption that  $K$  is fixed. Let  $\hat{u}_t = y_t - \bar{y} - (t - \bar{t})\hat{\beta}$ , then

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \hat{u}_t &= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \left[ u_t - \bar{u} - (t - \bar{t}) (\hat{\beta} - \beta) \right] \\ \rightarrow \Lambda [W_n(r) - rW_n(1)] - \Lambda \left[ \int_0^r \left( s - \frac{1}{2} \right) ds \right] \left[ \int_0^1 \left( s - \frac{1}{2} \right)^2 ds \right]^{-1} \int_0^1 \left( s - \frac{1}{2} \right) dW_n(s) \\ &: = \Lambda V_n(r), \end{aligned}$$

where

$$V_n(r) = W_n(r) - rW_n(1) - 6r(r-1) \int_0^1 \left( t - \frac{1}{2} \right) dW_n(t).$$

Invoking the continuous mapping theorem, we obtain

$$\begin{aligned} \hat{\Omega} &\rightarrow d\Lambda \frac{1}{K} \sum_{k=1}^K \left[ \int_0^1 \phi_k(r) dV_n(r) \right] \left[ \int_0^1 \phi_k(s) dV_n(s) \right]' \Lambda' \\ &= \Lambda \frac{1}{K} \sum_{k=1}^K \left[ \int_0^1 \tilde{\phi}_k(r) dW_n(r) \right] \left[ \int_0^1 \tilde{\phi}_k(s) dW_n(s) \right]' \Lambda' \\ &: = \Lambda \frac{1}{K} \sum_{k=1}^K \xi_k \xi_k' \Lambda', \end{aligned} \tag{8}$$

where

$$\tilde{\phi}_k(r) = \phi_k(r) - \int_0^1 \phi_k(s) ds - 12 \left( \int_0^1 \phi_k(s) \left( s - \frac{1}{2} \right) ds \right) \left( r - \frac{1}{2} \right), \tag{9}$$

is the transformed basis function and

$$\xi_k = \int_0^1 \tilde{\phi}_k(r) dW_n(r).$$

We call the above asymptotics the fixed- $K$  asymptotics. This is similar to the fixed- $b$  asymptotics of Kiefer and Vogelsang (2005).

Common choices of  $\phi_k$  are the sine and cosine trigonometric polynomials. In fact, using a simple Fourier expansion and assuming that  $\mathcal{K}(\cdot)$  is even, we can show that the eigenfunctions in (4) are the sine and cosine functions. A subset of the cosine functions  $\{\phi_k(r) = \sqrt{2} \cos \pi kr, k = 0, 1, \dots\}$  enjoys the desirable property that

$$\tilde{\phi}_k(r) = \phi_k(r) \text{ for } k = 0, 2, 4, \dots$$

So not only  $\{\phi_k(r)\}$  are orthonormal but also are their transforms as defined in (9). Note that the first basis with  $k = 0$  is redundant as  $\sum_{t=1}^T \hat{u}_t = 0$ . We therefore take

$$\phi_k\left(\frac{t}{T}\right) = \sqrt{2} \cos\left(\frac{2\pi kt}{T}\right), \text{ for } k = 1, 2, \dots, K$$

as our data windows or basis functions. Similar to the Hanning window  $(1 - \cos 2\pi t/T)/2$ , the above functions have small side lobes and their Fourier transforms decay to zero rapidly. As a result, the associated LRV estimator has a small bias due to spectral leakage (Priestley (1981, p. 563)). This is an especially desirable feature for hypothesis testing where bias reduction is more important than the point estimation of the LRV.

With the above cosine data windows,  $\xi_k$  is iid  $N(0, \mathbb{I}_n)$ . As a result,  $\sum_{k=1}^K \xi_k \xi_k'$  is a Wishart distribution  $W_{np}(\mathbb{I}_p, K)$ . So  $\hat{\Omega}$  converges to a scaled Wishart distribution. In the scalar case, the limiting distribution reduces to the scaled chi-square distribution. In general, for any conforming constant vector  $z$ ,  $z' \hat{\Omega} z / z' \Omega z$  converges to a chi-square distribution. This result can be used to test hypotheses regarding  $\Omega$ . The resulting test may have better size properties. See PSJ (2006, 2007) and Hashimzade and Vogelsang (2007) for the same point based on conventional kernel estimators. We do not pursue this extension here as our main focus is on the inference for  $\beta$ .

### 3.3 Large-K Asymptotics

While the fixed- $K$  asymptotics may capture the randomness of  $\hat{\Omega}$  very well, it does not reflect the asymptotic bias of the  $\hat{\Omega}$ . In this section, we consider the asymptotic properties of  $\hat{\Omega}$  when both  $K$  and  $T$  go to infinity such that  $K/T \rightarrow 0$ . Our result extends Theorem 1 of Phillips (2005), which is concerned with the HAC estimation with an observed scalar process. Here the error process is an unobserved vector process.

**Theorem 2** *Let Assumption 1 hold. As  $K \rightarrow \infty$  such that  $K/T \rightarrow 0$ , we have*

$$(a) E\hat{\Omega} - \Omega = \frac{K^2}{T^2} B + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right).$$

$$(b) \text{var}\left(\text{vec}(\hat{\Omega})\right) = \frac{1}{K} (\Omega \otimes \Omega) (\mathbb{I}_{n^2} + \mathbb{K}_{nn}) + O\left(\frac{1}{T}\right) \text{ where}$$

$$B = -\frac{2\pi^2}{3} \sum_{h=-\infty}^{\infty} h^2 \Gamma_u(h), \quad \Gamma_u(h) = E u_t u_{t-h}'$$

and  $\mathbb{K}_{nn}$  is the  $n^2 \times n^2$  commutation matrix and  $\mathbb{I}_{n^2}$  is the  $n^2 \times n^2$  identity matrix.

The bias term is different from that given in Theorem 1(i) in Phillips (2005). This is because the basis functions we used is a subset of the basis functions in Phillips (2005). The advantage of dropping  $\{\sqrt{2} \cos \pi(2k-1)r, k = 1, 2, \dots\}$  is that the estimation uncertainty of

$\alpha$  and  $\beta$  does not affect the bias and variance calculation in large samples. More specifically, we show in the proof that  $\hat{\Omega}$  is asymptotically equivalent to

$$\tilde{\Omega} = \frac{1}{K} \sum_{k=1}^K \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^T \phi_k \left( \frac{t}{T} \right) u_t \right] \left[ \frac{1}{\sqrt{T}} \sum_{s=1}^T \phi_k \left( \frac{s}{T} \right) u_s \right]', \quad (10)$$

an estimator that is based on the true but unknown error term  $u_t$ . This result is in sharp contrast to existing results in the HAC estimation literature. For conventional kernel HAC estimators, the estimation uncertainty in model parameters gives rise to a higher order bias term, which is typically the same order of magnitude as the asymptotic variance. See for example SPJ. We have thus provided a novel way to eliminate the effect of the estimation uncertainty of the model parameters on the LRV estimation.

Theorem 2(b) characterizes the asymptotic behavior of the exact variance. This result is different from Theorem 1(ii) Phillips (2005) as the latter provides only the variance of the limiting distribution of  $\hat{\Omega}$ . In terms of moment calculations, our results are stronger than those in Phillips (2005). We can follow Phillips (2005) and show that when  $K = o(n^{4/5})$

$$\sqrt{K} \text{vec}(\hat{\Omega} - \Omega) \rightarrow N[0, (\Omega \otimes \Omega) (\mathbb{I}_{n^2} + \mathbb{K}_{nn})],$$

but our development here does not require this result.

Let

$$MSE(\hat{\Omega}, W) = \text{Evec}(\hat{\Omega} - \Omega)' W \text{vec}(\hat{\Omega} - \Omega)$$

be the mean squared error of  $\text{vec}(\hat{\Omega})$  with weighting matrix  $W$ . It follows from Theorem 2 that, up to smaller order terms:

$$\begin{aligned} MSE(\hat{\Omega}, W) &= \text{tr} \left[ W \text{Evec}(\hat{\Omega} - \Omega) \text{vec}(\hat{\Omega} - \Omega)' \right] \\ &= \text{vec}(B)' W \text{vec}(B)' \frac{K^4}{T^4} + \text{tr} [W (\Omega \otimes \Omega) (\mathbb{I}_{n^2} + \mathbb{K}_{nn})] \frac{1}{K}. \end{aligned}$$

So the MSE optimal  $K$  is given by

$$K = \left( \frac{\text{tr} [W (\Omega \otimes \Omega) (\mathbb{I}_{n^2} + \mathbb{K}_{nn})]}{4 \text{vec}(B)' W \text{vec}(B)} \right)^{1/5} T^{4/5}.$$

This formula is analogous to the conventional MSE optimal formula for bandwidth choice in kernel LRV estimators, e.g. Andrews (1991).

## 4 Robust Inference for the Trend Parameters

The hypotheses of interest in this paper are

$$H_0 : R\beta = r \text{ against } H_1 : R\beta \neq r,$$

where  $R$  is a  $p \times n$  matrix and  $r$  is a  $p \times 1$  vector. The usual Wald statistic  $F_{T,OLS}$  for testing  $H_0$  against  $H_1$  is given by

$$F_{T,OLS} = \left[ RT^{3/2}(\hat{\beta}_{OLS} - \beta) \right]' \left( R12\hat{\Omega}R' \right)^{-1} \left[ RT^{3/2}(\hat{\beta}_{OLS} - \beta) \right].$$

When  $p = 1$ , we can construct the usual t-statistic

$$t_{T,OLS} = \frac{RT^{3/2}(\hat{\beta}_{OLS} - \beta)}{\left(R12\hat{\Omega}R'\right)^{1/2}} = \sqrt{F_{T,OLS}}.$$

#### 4.1 Fixed- $K$ Asymptotics

Under the fixed- $K$  asymptotics and the null hypothesis

$$F_{T,OLS} \rightarrow_d \frac{1}{12K} \left( R\Lambda \int_0^1 \left( r - \frac{1}{2} \right) dW_n(r) \right) \\ \times \left\{ R\Lambda \sum_{k=1}^K \left[ \int_0^1 \tilde{\phi}_k(r) dW_n(r) \right] \left[ \int_0^1 \tilde{\phi}_k(s) dW_n(s) \right] \Lambda' R' \right\}^{-1} \left( R\Lambda \int_0^1 \left( r - \frac{1}{2} \right) dW_n(r) \right)'$$

Let  $R_{p \times n} \Lambda W_n(r) = R^* W_p^*(r)$  for some  $p \times p$  matrix  $R^*$  and  $p$ -dimensional Brownian motion  $W_p^*(r)$ , then for a fixed  $K$ , we have

$$F_{T,OLS} \rightarrow d \left( \frac{1}{\sqrt{12}} \int_0^1 \left( r - \frac{1}{2} \right) dW_p^*(r) \right) \left\{ \frac{1}{K} \sum_{k=1}^K \left[ \int_0^1 \tilde{\phi}_k(r) dW_p^*(r) \right] \left[ \int_0^1 \tilde{\phi}_k(s) dW_p^*(s) \right] \right\}^{-1} \\ \times \left( \frac{1}{\sqrt{12}} \int_0^1 \left( r - \frac{1}{2} \right) dW_p^*(r) \right)' := \eta' \left( \frac{1}{K} \sum_{k=1}^K \xi_k \xi_k' \right)^{-1} \eta,$$

where

$$\eta = \frac{1}{\sqrt{12}} \int_0^1 \left( r - \frac{1}{2} \right) dW_p^*(r) \text{ and } \xi_k = \int_0^1 \tilde{\phi}_k(r) dW_p^*(r).$$

Since

$$\text{cov} \left[ \int_0^1 \left( r - \frac{1}{2} \right) dW_n(r), \int_0^1 \tilde{\phi}_k(r) dW_n(r) \right] = \int_0^1 \left( r - \frac{1}{2} \right) \tilde{\phi}_k(r) dr = 0 \text{ for all } k,$$

$\eta$  and  $\xi_k$  are independent as both are normal random variables. In addition,  $\xi_k \sim iidN(0, \mathbb{I}_p)$  and  $\sum_{k=1}^K \xi_k \xi_k'$  is a Wishart distribution  $W_{np}(\mathbb{I}_p, K)$ . Hence the limiting distribution of  $F_{T,OLS}$  is related to Hotelling's T-square distribution (Hotelling (1931)):

$$F_{T,OLS} \rightarrow_d T^2(p, K),$$

and

$$\frac{(K-p+1)}{pK} F_{T,OLS} \rightarrow_d \frac{K-p+1}{pK} T^2(p, K) \sim F_{p, K-p+1}.$$

In other words,

$$\frac{(K-p+1)}{pK} F_{T,OLS} \rightarrow_d F_{p, K-p+1} := \frac{\chi_p^2/p}{\chi_{K-p+1}^2/(K-p+1)},$$

where  $\chi_p^2$  and  $\chi_{K-p+1}^2$  denote independent  $\chi^2$  random variables. Of course, for the above distribution to be well defined, we need to assume that  $K \geq p$ , a necessary condition to

ensure that  $R\hat{\Omega}R'$  is invertible. In general, we need to assume  $K \geq n$ , a necessary condition for the positive definiteness of  $\hat{\Omega}$ .

When  $p = 1$ , the above result reduces to  $t_T \rightarrow_d t_K$ . That is, the t-statistic converges to the t-distribution with  $K$  degrees of freedom.

We have therefore shown that under the fixed- $K$  asymptotics, the scaled Wald statistic converges weakly to the  $F$  distribution with degrees of freedom  $p$  and  $K - p + 1$  and the t-statistic converges to the t-distribution with degrees of freedom  $K$ . These results are very handy as critical values from the  $F$  distribution or the  $t$  distribution can be easily obtained from statistical tables or standard econometrics packages.

Under the local alternative hypothesis,

$$H_1(\delta^2) : R\beta = r + c / \left( T\sqrt{T} \right) \text{ where } c = (R12\Omega R')^{1/2} \tilde{c} \quad (11)$$

for some  $p \times 1$  vector  $\tilde{c}$ , we have

$$\frac{(K - p + 1)}{pK} F_{T,OLS} \rightarrow_d \frac{(K - p + 1)}{p} (\eta + \tilde{c})' \left( \sum_{k=1}^K \xi_k \xi_k' \right)^{-1} (\eta + \tilde{c}) := F_{p,K-p+1}(\delta^2),$$

a noncentral  $F$  distribution with degrees of freedom  $(p, K - p + 1)$  and noncentrality parameter

$$\delta^2 = (\tilde{c})' \tilde{c} = c' (12R\Omega R')^{-1/2} (12R\Omega R')^{-1/2} c = c' (12R\Omega R')^{-1} c.$$

This result follows from Proposition 8.2 in Bilodeau and Brenner (1999) where  $F_c$  is the canonical F distribution (Bilodeau and Brenner (1999), page 42). Similarly, the t-statistic converges to the noncentral  $t$  distribution with degrees of freedom  $K$  and noncentrality parameter  $\delta = c / (12R\Omega R')^{1/2} = \tilde{c}$ .

The local alternative power depends on  $c$  only through the noncentrality parameter  $\delta^2 = \|\tilde{c}\|^2$ , the squared length of vector  $\tilde{c}$ . The direction of  $\tilde{c}$  does not matter. Hence, for the first order asymptotics given here, it is innocuous to assume that  $\tilde{c}$  is uniformly distributed on the sphere  $S_p(\delta) = \{x \in \mathbb{R}^p : \|x\| = \delta\}$ . It turns out that this assumption greatly simplifies the development of higher order expansions in later sections.

## 4.2 Large-K Asymptotics

When  $K \rightarrow \infty$  such that  $K/T \rightarrow 0$ , the LRV estimator  $\hat{\Omega}$  is consistent. As a consequence

$$F_{T,OLS} \rightarrow \chi_p^2 \text{ under } H_0 \text{ and } F_{T,OLS} \rightarrow \chi_p^2(\delta^2) \text{ under } H_1(\delta^2).$$

When  $p = 1$ , the above result reduces to

$$t_{T,OLS} \rightarrow N(0, 1) \text{ under } H_0 \text{ and } t_{T,OLS} \rightarrow N(\delta, 1) \text{ under } H_1(\delta^2).$$

The above results are of course standard.

To compare the fixed- $K$  asymptotics with the large- $K$  asymptotics, we evaluate the difference in their  $1 - \alpha$  quantiles. Let  $F_{p,K-p+1}^\alpha$  be the  $1 - \alpha$  quantile of the  $F_{p,K-p+1}$  distribution and  $F_{p,\infty}^\alpha$  be the  $1 - \alpha$  quantile of the  $F_{p,\infty} \equiv \chi_p^2/p$  distribution. In other

words,  $pF_{p,\infty}^\alpha \equiv \chi_p^\alpha$  is the  $1 - \alpha$  quantile of the  $\chi_p^2$  distribution. By definition and with a slight abuse of notation,

$$\begin{aligned}
1 - \alpha &= P(F_{p,K-p+1} < F_{p,K-p+1}^\alpha) = P\left(\chi_p^2 < pF_{p,K-p+1}^\alpha \frac{\chi_{K-p+1}^2}{(K-p+1)}\right) \\
&= EG_p\left(pF_{p,K-p+1}^\alpha \frac{\chi_{K-p+1}^2}{(K-p+1)}\right) \\
&= G_p(pF_{p,\infty}^\alpha) + G'_p(pF_{p,\infty}^\alpha) E\left[pF_{p,K-p+1}^\alpha \frac{\chi_{K-p+1}^2}{(K-p+1)} - pF_{p,\infty}^\alpha\right] \\
&\quad + \frac{1}{2}G''_p(pF_{p,\infty}^\alpha) E\left[pF_{p,K-p+1}^\alpha \frac{\chi_{K-p+1}^2}{(K-p+1)} - pF_{p,\infty}^\alpha\right]^2 + o\left(\frac{1}{K^2}\right) \quad (12) \\
&= G_p(pF_{p,\infty}^\alpha) + G'_p(pF_{p,\infty}^\alpha) [pF_{p,K-p+1}^\alpha - pF_{p,\infty}^\alpha] \\
&\quad + \frac{1}{K}G''_p(pF_{p,\infty}^\alpha) (pF_{p,\infty}^\alpha)^2 + o(pF_{p,K-p+1}^\alpha - pF_{p,\infty}^\alpha) + O\left(\frac{1}{K^2}\right).
\end{aligned}$$

Therefore

$$pF_{p,K-p+1}^\alpha = \chi_p^\alpha - \frac{1}{K} \frac{G''_p(\chi_p^\alpha)}{G'_p(\chi_p^\alpha)} (\chi_p^\alpha)^2 + o\left(\frac{1}{K}\right). \quad (13)$$

But

$$-\frac{G''_p(\chi_p^\alpha)}{G'_p(\chi_p^\alpha)} = \frac{1}{2\chi_p^\alpha} (\chi_p^\alpha - p + 2),$$

hence

$$pF_{p,K-p+1}^\alpha = \chi_p^\alpha + \frac{1}{2K} (\chi_p^\alpha - p + 2) \chi_p^\alpha + o\left(\frac{1}{K}\right).$$

Therefore the critical values from the F-distribution are larger than those from the  $\chi^2$  distribution, reflecting the randomness in the denominator of the Wald statistic.

## 5 High Order Expansion of the Finite Sample Distribution

In this section, we consider a high order expansion of the Wald statistic in order to design a procedure to select  $K$ . The procedure is suitable for hypothesis testing or confidence interval construction. We make the simplification assumption that  $u_t$  is normal, which facilitates the derivations. The assumption could be relaxed by taking distributions based (for example) on Gram-Charlier expansions, but at the cost of much greater complexity (see, for example, Phillips (1980), Taniguchi and Puri (1996), Velasco and Robinson (2001)).

Let  $V = \text{var}(\text{vec}(u))$ , then the GLS estimator of  $\text{vec}(\theta)$  is

$$\text{vec}\left(\hat{\theta}_{GLS} - \theta\right) = [(\mathbb{I}_n \otimes X)' V^{-1} (\mathbb{I}_n \otimes X)]^{-1} (\mathbb{I} \otimes X)' V^{-1} \text{vec}(u).$$

Similarly, the OLS estimator can be written as

$$\text{vec}\left(\hat{\theta}_{OLS} - \theta\right) = [(\mathbb{I}_n \otimes X)' (\mathbb{I}_n \otimes X)]^{-1} (\mathbb{I} \otimes X)' \text{vec}(u).$$

So

$$vec\left(\hat{\theta}_{OLS} - \theta\right) = vec\left(\hat{\theta}_{GLS} - \theta\right) + \Delta,$$

where  $\Delta = \left(\Delta'_\alpha, \Delta'_\beta\right)'$  and more explicitly

$$\Delta = \left\{ \left[ (\mathbb{I}_n \otimes X)' (\mathbb{I}_n \otimes X) \right]^{-1} (\mathbb{I} \otimes X)' - \left[ (\mathbb{I}_n \otimes X)' V^{-1} (\mathbb{I}_n \otimes X) \right]^{-1} (\mathbb{I} \otimes X)' V^{-1} \right\} vec(u).$$

It follows from the asymptotic equivalency of  $\hat{\theta}_{OLS}$  and  $\hat{\theta}_{GLS}$  that  $E c' \Delta_\beta \Delta'_\beta c = O(1/T)$  for any vector  $c$ . See Grenander and Rosenblatt (1957).

It is easy to show that

$$E \left[ vec\left(\hat{\theta}_{GLS} - \theta\right) \Delta' \right] = 0.$$

Hence,  $\hat{\theta}_{GLS} - \theta$  and  $\Delta$  are independent. In addition,

$$\hat{u} = \left\{ \mathbb{I}_{nT} - \left[ \mathbb{I}_n \otimes X (X'X)^{-1} X' \right] \right\} vec(u),$$

and thus

$$\begin{aligned} & Evec\left(\hat{\theta}_{GLS} - \theta\right) \hat{u}' \\ &= \left[ (\mathbb{I}_n \otimes X)' V^{-1} (\mathbb{I}_n \otimes X) \right]^{-1} (\mathbb{I}_n \otimes X)' V^{-1} V \left\{ \mathbb{I}_{nT} - \left[ \mathbb{I}_n \otimes X (X'X)^{-1} X' \right] \right\} = 0. \end{aligned}$$

So  $\hat{\theta}_{GLS}$  is independent of both  $\Delta$  and  $\hat{\Omega}$ .

Let  $F_{T,GLS}$  be the Wald statistic based on the GLS estimator:

$$F_{T,GLS} = RT^{3/2} (\hat{\beta}_{GLS} - \beta) \left( R12\hat{\Omega}R' \right)^{-1} RT^{3/2} (\hat{\beta}_{GLS} - \beta).$$

Using the asymptotic equivalence of the OLS and GLS estimators and the above two independence conditions, we can prove the following Lemma.

**Lemma 3** *Let Assumption 1 hold and assume that  $\varepsilon_t \sim iidN(0, \Sigma)$ . Then*

$$(a) P \left( \frac{(K-p+1)}{pK} F_{T,GLS} < z \right) = EG_p \left( pz \frac{K}{K-p+1} \Xi^{-1} \right) + O\left(\frac{1}{T}\right).$$

$$(b) P \left( \frac{(K-p+1)}{pK} F_{T,OLS} < z \right) = P \left( \frac{(K-p+1)}{pK} F_{T,GLS} < z \right) + O\left(\frac{1}{T}\right),$$

where  $G_p$  is the CDF of a  $\chi^2$  random variable with degrees of freedom  $p$  and

$$\begin{aligned} \Xi &= \left[ e'_\beta (R\Omega R')^{1/2} \left( R\hat{\Omega}R' \right)^{-1} (R\Omega R')^{1/2} e_\beta \right], \\ e_\beta &= \frac{(R12\Omega_{T,GLS})^{-1/2} RT^{3/2} (\hat{\beta}_{GLS} - \beta)}{\left\| (R12\Omega_{T,GLS})^{-1/2} RT^{3/2} (\hat{\beta}_{GLS} - \beta) \right\|}. \end{aligned}$$

Lemma 3 shows that the estimation uncertainty of  $\hat{\Omega}$  affects the distribution of the Wald statistic only through the statistic  $\Xi$ . Taking a Taylor expansion, we have

$$\Xi^{-1} = 1 + L + Q + o_p\left(\frac{1}{K}\right),$$

where  $L$  is linear in  $\hat{\Omega} - \Omega$  and  $Q$  is quadratic in  $\hat{\Omega} - \Omega$ . The exact expressions for  $L$  and  $Q$  are not important here but are given in the proof of Theorem 4. Using this stochastic expansion, we obtain

$$\begin{aligned}
& P\left(\frac{(K-p+1)}{pK}F_{T,OLS} < z\right) \\
&= EG_p\left(pz\frac{K}{K-p+1}(1+L+Q)\right) + O\left(\frac{1}{T}\right) + o\left(\frac{1}{K}\right) \\
&= EG_p(pz) + G'_p(pz)pzE\left[\frac{K}{K-p+1}(1+L+Q) - 1\right] \\
&\quad + \frac{1}{2}G''_p(pz)p^2z^2E\left[\frac{K}{K-p+1}(1+L+Q) - 1\right]^2 + O\left(\frac{1}{T}\right) + o\left(\frac{1}{K}\right),
\end{aligned}$$

where the  $O(\cdot)$  term holds uniformly over  $z \in \mathbb{R}^+$ . By developing asymptotic expansions of  $E\left[\frac{K}{K-p+1}(1+L+Q) - 1\right]^m$  for  $m = 1, 2$ , we can establish a higher order expansion of the finite sample distribution for the case where  $K \rightarrow \infty$  such that  $K/T \rightarrow 0$ .

**Theorem 4** *Let Assumption 1 hold and assume that  $\varepsilon_t \sim iidN(0, \Sigma)$ . If  $K \rightarrow \infty$  such that  $K/T \rightarrow 0$ , then*

$$\begin{aligned}
& P\left(\frac{(K-p+1)}{K}F_{T,OLS} < pz\right) \\
&= G_p(pz) + \frac{K^2}{T^2}G'_p(pz)pz\bar{B} + \frac{1}{K}G''_p(pz)p^2z^2 + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right) \quad (14)
\end{aligned}$$

where

$$\bar{B} = \bar{B}(R, B, \Omega) = \frac{\text{tr}\left\{(RBR')(R\Omega R')^{-1}\right\}}{p}.$$

The first term in (14) comes from the standard chi-square approximation of the Wald statistic. The second term captures the nonparametric bias of the LRV estimator while the third term reflects the variance of the LRV estimator. The result is analogous to those obtained by SPJ for Gaussian location models and Sun and Phillips (SP, 2009) for general linear GMM models with stationary data. However, there is an important difference. For conventional kernel estimators as used in SPJ and SP, the asymptotics expansion contains a term that reflects the bias due to the estimation error of the model parameters. Such a term does not appear here because the basis functions we employ are asymptotically orthogonal to the estimation error.

To understand the relationship between the fixed- $K$  and large- $K$  asymptotics, we develop an expansion of the limiting  $F_{p,K-p+1}$  distribution as in (12):

$$P(F_{p,K-p+1} < z) = G_p(pz) + \frac{1}{K}G''_p(pz)p^2z^2 + o\left(\frac{1}{K}\right).$$

Comparing this with Theorem 4, we find that the fixed- $K$  asymptotics captures one of the higher order terms in the high order expansion of the large  $K$  asymptotics. Plugging  $z = F_{p,K-p+1}^\alpha$  into the above equation yields:

$$1 - \alpha = G_p(pF_{p,K-p+1}^\alpha) + \frac{1}{K}G''_p(pF_{p,K-p+1}^\alpha)[pF_{p,K-p+1}^\alpha]^2 + o\left(\frac{1}{K}\right).$$

This implies that

$$\begin{aligned}
& P\left(\frac{(K-p+1)}{pK}F_{T,OLS} < F_{p,K-p+1}^\alpha\right) \\
&= 1 - \alpha + \frac{K^2}{T^2}G'_p(pF_{p,K-p+1}^\alpha)pF_{p,K-p+1}^\alpha\frac{1}{p}\text{tr}\left\{(RBR')(R\Omega R')^{-1}\right\} \\
&\quad + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right). \tag{15}
\end{aligned}$$

Therefore, use of critical values from the  $F_{p,K-p+1}$  distribution removes the variance term  $K^{-1}G''_p(pz)z^2p^2$  in the higher order expansion. The size distortion is then of order  $O(K^2/T^2)$ . In contrast, if the critical value from the conventional  $\chi_p^2$  distribution is used, the size distortion is of order  $O(K^2/T^2) + O(1/K)$ . So when  $K^3/T^2 \rightarrow 0$ , using critical values from the  $F_{p,K-p+1}$  distribution should lead to size improvements. We have thus shown that critical values from the fixed- $K$  asymptotics is second order correct under the large- $K$  asymptotics.

If we use critical values from the conventional  $\chi_p^2$  distribution to construct confidence regions, then the variance term contributes to under-coverage as  $G''_p(pz) < 0$ . This is expected. Use of the  $\chi_p^2$  distribution does not take into account the randomness of the LRV estimator and the critical values from it tend to be smaller than they should be. As a result, the confidence region is smaller, leading to under-coverage. On the other hand, the bias term may contribute to under-coverage or over-coverage. In the scalar case when  $n = 1$ , we may argue that the nonparametric bias is negative for typical economic time series. Hence the bias term is also negative, leading to under-coverage. So in the scalar case, both the bias and variance terms call for a larger critical value. The use of  $F_{p,K-p+1}$  critical values, which is larger than  $\chi_p^2$  critical values, is justified by the demand from both terms. Note that when  $F_{p,K-p+1}^\alpha$  is used, the variance term is removed. Interestingly, the bias term is also reduced as  $G'_p(x)x$  is a decreasing function of  $x$ . Therefore, for scalar economic time series, the use of  $F_{p,K-p+1}^\alpha$ , which is designed to capture the randomness of the LRV estimator, does not have any unintended consequence. To some extent, the use of  $F_{p,K-p+1}^\alpha$  kills two birds with one stone.

There is a caveat to the preceding observation. In multivariate cases, the direction of the bias is not obvious.  $p^{-1}\text{tr}\left\{(RBR')(R\Omega R')^{-1}\right\}$  can be regarded as a measure of the average bias of the LRV estimator. This measure depends both  $B$  and  $R$ . As an example, consider the simple symmetric bivariate MA(1) process:

$$\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \begin{pmatrix} a & b \\ b & a \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

where  $(\varepsilon_{1t}, \varepsilon_{2t})' \sim iid(0, \mathbb{I}_2)$ . It is easy to show that the spectral density matrix is

$$f(\lambda) = \frac{1}{2\pi} \begin{pmatrix} 1 + 2a \cos(\lambda) + a^2 + b^2 & 2b + 2ab \cos \lambda \\ 2b + 2ab \cos \lambda & 1 + 2a \cos(\lambda) + a^2 + b^2 \end{pmatrix}.$$

When  $a > 0$ , each time series has a typical spectral shape at the origin. The LRV and the bias matrix are

$$\Omega = \begin{pmatrix} 1 + 2a + a^2 + b^2 & 2b + 2ab \\ 2b + 2ab & 1 + 2a + a^2 + b^2 \end{pmatrix}, \quad B = -\frac{4\pi^2}{3} \begin{pmatrix} a & ab \\ ab & a \end{pmatrix}.$$

When  $a > 0$ , the bias of  $\hat{\Omega}_{ii}$  is negative. However, if  $R = (1, -1)$ , then

$$\frac{1}{p} \text{tr} \left\{ (RBR') (R\Omega R')^{-1} \right\} = \frac{4\pi^2}{3} \frac{2a(b-1)}{2(a-b+1)^2},$$

which can be positive if  $b > 1$ . We have therefore provided an example that can be regarded as typical for economic time series and yet the first order bias is positive. In this case, the use of the fixed- $K$  asymptotics may have an unintended consequence. In our example, the bias and variance terms in the high order expansion have opposite signs. It may be advantageous to use the conventional  $\chi^2$  critical values so that the two terms can cancel out with each other. This opportunity is lost if we use the fixed- $K$  asymptotics. In fact, using fixed- $K$  critical values removes the variance term at the cost of inflating the bias term. A direct implication is that the fixed- $K$  asymptotics should not be used without due consideration.

Theorem 4 gives an expansion of the distribution of  $K^{-1}(K-p+1)F_{T,OLS}$ . The factor  $K^{-1}(K-p+1)$  is a finite sample correction factor. Without this correction, we can show that, up to smaller order terms

$$P(F_{T,OLS} < \chi_p^\alpha) = G_p(\chi_p^\alpha) + \frac{K^2}{T^2} G'_p(\chi_p^\alpha) \chi_p^\alpha \bar{B} - \frac{1}{K} G'_p(\chi_p^\alpha) \chi_p^\alpha (p-1) + \frac{1}{K} G''_p(\chi_p^\alpha) [\chi_p^\alpha]^2.$$

Comparing this with (14), we find that the above expansion has an additional term  $-K^{-1}G'_p(\chi_p^\alpha) \chi_p^\alpha (p-1)$ . For any given critical value  $\chi_p^\alpha$ , this term is negative and grows with  $p$ , the number of restrictions in the hypothesis. As a result, the error in rejection probability or the error in coverage probability tends to be larger for larger  $p$ . This explains why confidence regions tend to have large under-coverage when the dimension of the problem is high.

In the rest of the paper, we use the finite sample corrected Wald statistic

$$F_{T,OLS}^* = \frac{(K-p+1)}{K} F_{T,OLS}$$

and employ critical value  $pF_{p,K-p+1}^\alpha$  to perform our test. For convenience, we refer to  $F_{T,OLS}^*$  as the  $F^*$  statistic and the test as the  $F^*$  test. The following theorem gives the size distortion and local power of the  $F^*$  test.

**Theorem 5** *Let Assumption 1 hold and assume that  $\varepsilon_t \sim iidN(0, \Sigma)$ . If  $K \rightarrow \infty$  such that  $K/T \rightarrow 0$ , then*

(a) *The size distortion of the  $F^*$  test is*

$$\begin{aligned} & P(F_{T,OLS}^* > pF_{p,K-p+1}^\alpha) - \alpha \\ &= -\frac{K^2 \bar{B}}{T^2} G'_p(\chi_p^\alpha) \chi_p^\alpha + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right). \end{aligned} \quad (16)$$

(b) *Under the local alternative  $H_1(\delta^2) : R\beta = r + (R\Omega R')^{1/2} \tilde{c} / (T\sqrt{T})$  where  $\tilde{c}$  is uniformly distributed on the sphere  $S_p(\delta) = \{x \in \mathbb{R}^p : \|x\| = \delta\}$ , the power of the  $F^*$  test is*

$$\begin{aligned} & P(F_{T,OLS}^* > pF_{p,K-p+1}^\alpha) = 1 - G_{p,\delta^2}(\chi_p^\alpha) - \frac{K^2}{T^2} G'_{p,\delta^2}(\chi_p^\alpha) \chi_p^\alpha \bar{B} \\ & - \frac{1}{K} \mathcal{Q}_{p,\delta^2}(\chi_p^\alpha) (\chi_p^\alpha)^2 + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right), \end{aligned} \quad (17)$$

where  $G_{p,\delta^2}$  is the CDF of the noncentral  $\chi^2$  distribution with degrees of freedom  $p$  and noncentrality parameter  $\delta^2$  and

$$\mathcal{Q}_{p,\delta^2}(z) = G''_{p,\delta^2}(z) - \frac{G''_p(z)}{G'_p(z)} G'_{p,\delta^2}(z) = \sum_{j=0}^{\infty} \frac{(\delta^2/2)^j}{j!} e^{-\delta^2/2} \frac{z^{j+p/2-1} e^{-z/2}}{2^{j+p/2} \Gamma(j+p/2)} \frac{j}{z}.$$

Theorem 5(a) follows from Theorem 4. The uniformity of  $\tilde{c}$  on a sphere enables us to use a similar argument to prove Theorem 5(b). A key point in the proof of Theorem 4 is that  $e_\beta$  is uniformly distributed on the unit sphere  $S_p(1)$ , which follows from the rotation invariance of the multivariate standard normal distribution. The uniformity of  $\tilde{c}$  ensures the same property holds for the corresponding statistic

$$e_{\beta\delta} = \frac{(R12\Omega_{T,GLS})^{-1/2} RT^{3/2}(\hat{\beta}_{GLS} - \beta) + \tilde{c}}{\left\| (R12\Omega_{T,GLS})^{-1/2} RT^{3/2}(\hat{\beta}_{GLS} - \beta) + \tilde{c} \right\|}$$

under the local alternative hypothesis.

The quantity  $\mathcal{Q}_{p,\delta^2}(\chi_p^\alpha)$  reflects the difference in curvatures of the two CDF functions  $G_p(z)$  and  $G_{p,\delta^2}(z)$  at the point  $z = \chi_p^\alpha$ . When we use the second order correct critical value  $pF_{p,K-p+1}^\alpha$ , the variance term is removed under the null. However, due to the difference in curvatures, the variance term remains under the local alternative hypothesis. The  $O(1/K)$  term in Theorem 5(b) captures this effect. Since  $\mathcal{Q}_{p,\delta^2}(z) > 0$  for all  $z > 0$ , this term increases monotonically with  $K$ . According to this term, the value of  $K$  should be chosen as large as possible. This is not surprising. In order to improve the power of the  $F^*$  test, we should minimize the randomness of the LRV estimator, which calls for a large  $K$  value. However, a large  $K$  value may produce large bias which may lead to power loss or size distortion. In the next section, we show that there is an opportunity to select  $K$  to trade off the bias effect and variance effect on the size and power properties.

## 6 Optimal Smoothing Parameter Selection

We have shown that the optimal  $K$  that minimizes the asymptotic mean squared error in LRV estimation has the form  $K = O(T^{4/5})$ . However, there is no reason to expect that such a choice is the most appropriate in statistical testing using nonparametrically scaled statistics. In this section, we provide a novel approach that is most suitable for semiparametric testing.

### 6.1 Optimal K Formula

In view of the asymptotic expansion in (16) and ignoring the higher order terms, we know that the type I error for the  $F^*$  test can be measured by

$$e_I = \alpha - \frac{K^2 \bar{B}}{T^2} G'_p(\chi_p^\alpha) \chi_p^\alpha.$$

Similarly, from (17), the type II error of the  $F^*$  test can be measured by

$$e_{II} = G_{p,\delta^2}(\chi_p^\alpha) + \frac{K^2}{T^2} G'_{p,\delta^2}(\chi_p^\alpha) \chi_p^\alpha \bar{B} + \frac{1}{K} \mathcal{Q}_{p,\delta^2}(\chi_p^\alpha) (\chi_p^\alpha)^2.$$

We choose  $K$  to minimize the type II error while controlling for the type I error. More specifically, we solve

$$\min e_{II}, \text{ s.t. } e_I \leq \kappa\alpha$$

where  $\kappa$  is a constant greater than 1. Ideally, the type I error is less than or equal to the nominal type I error  $\alpha$ . In finite samples, there are always some approximation error and we allow for some discrepancy by introducing the tolerance factor  $\kappa$ . For example, when  $\alpha = 5\%$  and  $\kappa = 1.2$ , we aim to control the type I error such that it is not greater than 6%. Note that  $\kappa$  may depend on the sample size  $T$ . For a larger sample size, we may require  $\kappa$  to take smaller values.

The solution to the minimization problem depends on the sign of  $\bar{B}$ . When  $\bar{B} > 0$ , the constraint  $e_I \leq \kappa\alpha$  is not binding and we have the unconstrained minimization problem:  $\min e_{II}$ . The optimal  $K$  is

$$K_{\text{opt}} = \left( \frac{\mathcal{Q}_{p,\delta^2}(\chi_p^\alpha) \chi_p^\alpha}{2\bar{B}G'_{p,\delta^2}(\chi_p^\alpha)} \right)^{1/3} T^{2/3}. \quad (18)$$

When  $\bar{B} < 0$ , the constraint  $e_I \leq \kappa\alpha$  may be binding and we have to use the Kuhn-Tucker theorem to search for the optimum. Let  $\lambda$  be the Lagrange multiplier, and define

$$\begin{aligned} L(K, \lambda) = & G_{p,\delta^2}(\chi_p^\alpha) + \frac{K^2}{T^2} G'_{p,\delta^2}(\chi_p^\alpha) \chi_p^\alpha \bar{B} + \frac{1}{K} \mathcal{Q}_{p,\delta^2}(\chi_p^\alpha) (\chi_p^\alpha)^2 \\ & + \lambda \left( \left( \alpha - \frac{K^2 \bar{B}}{T^2} G'_{p,\delta^2}(\chi_p^\alpha) \chi_p^\alpha \right) - \kappa\alpha \right). \end{aligned} \quad (19)$$

It is easy to show that at the optimal  $K$ , the constant  $e_I \leq \kappa\alpha$  is indeed binding and  $\lambda > 0$ . Hence, the optimal  $K$  is

$$K_{\text{opt}} = \left( \frac{(\kappa - 1)\alpha}{|\bar{B}| G'_{p,\delta^2}(\chi_p^\alpha) \chi_p^\alpha} \right)^{1/2} T, \quad (20)$$

and the corresponding Lagrange multiplier is

$$\lambda_{\text{opt}} = \frac{G'_{p,\delta^2}(\chi_p^\alpha)}{G'_{p,\delta^2}(\chi_p^\alpha)} + \frac{|\bar{B}|^{1/2} \mathcal{Q}(\chi_p^\alpha) [\chi_p^\alpha]^{5/2} [G'_{p,\delta^2}(\chi_p^\alpha)]^{3/2}}{2[(\kappa - 1)\alpha]^{3/2} T}.$$

Formulae (18) and (20) can be written collectively as

$$K_{\text{opt}} = \left[ \frac{\mathcal{Q}_{p,\delta^2}(\chi_p^\alpha) \chi_p^\alpha}{2\bar{B} [G'_{p,\delta^2}(\chi_p^\alpha) - \lambda_{\text{opt}} G'_{p,\delta^2}(\chi_p^\alpha)]} \right]^{1/3} T^{2/3},$$

where

$$\lambda_{\text{opt}} = \begin{cases} 0, & \text{if } \bar{B} > 0 \\ \frac{G'_{p,\delta^2}(\chi_p^\alpha)}{G'_{p,\delta^2}(\chi_p^\alpha)} + \frac{|\bar{B}|^{1/2} \mathcal{Q}(\chi_p^\alpha) [\chi_p^\alpha]^{5/2} [G'_{p,\delta^2}(\chi_p^\alpha)]^{3/2}}{2[(\kappa - 1)\alpha]^{3/2} T}, & \text{if } \bar{B} < 0 \end{cases} \quad (21)$$

The function  $L(K, \lambda)$  is a weighted sum of the type I and type II errors with weight given by the optimal Lagrange multiplier. When  $\bar{B} > 0$ , the type I error is expected to be capped by the nominal type I error. As a result, the optimal Lagrange multiplier is zero

and we assign all weight to the type II error. This weighting scheme might be justified by the argument that it is worthwhile to take advantage of the extra reduction in the type II errors without inflating the type I error. When  $\bar{B} < 0$ , the type I error is expected to be larger than the nominal type I error. The constraint on the type I error is binding and the Lagrange multiplier is positive. In this case, the loss function is a genuine weighted sum of type I and type II errors. As the tolerance parameter  $\kappa$  decreases toward 1, the weight to the type I error increases.

When the Lagrange multiplier  $\lambda_{\text{opt}}$  is finite, the optimal  $K_{\text{opt}}$  has an expansion rate of  $O(T^{2/3})$ . This rate contrasts with the optimal rate  $O(T^{4/5})$  for minimizing the mean squared error of the corresponding LRV estimator. Thus, the MSE optimal values of  $K$  for LRV estimation are much larger as  $T \rightarrow \infty$  than those which are most suited for statistical testing. On the other hand, when the Lagrange multiplier  $\lambda_{\text{opt}}$  grows with  $T$  such that  $\lambda_{\text{opt}} \sim T^2$ , which holds if  $\kappa - 1 \sim 1/T^2$ , the optimal  $K$  is bounded. Fixed  $K$  rules may then be interpreted as allowing for increasingly smaller deviation from the nominal type I error. This gives us an interpretation of fixed  $K$  rules in terms of the tolerance of the type I error we can allow.

When the Lagrange multiplier  $\lambda_{\text{opt}}$  grows with  $T$  such that  $K_{\text{opt}} = O(T/\sqrt{\lambda_{\text{opt}}})^{2/3}$ , the size distortion is of order  $O(T/\sqrt{\lambda_{\text{opt}}})^{-2/3}$ . It is apparent that the size distortion becomes smaller if a larger weight is assigned to the type I error. In particular, when  $\lambda_{\text{opt}}$  is finite, the size distortion is of order  $O(T^{-2/3})$ , which is larger than  $O(T^{-1})$ , the size distortion for the case  $\lambda_{\text{opt}} \sim T^2$ . It is obvious that the use of  $K_{\text{opt}}$  involves some tradeoff between two elements in the loss function (19). Note that even when  $\lambda_{\text{opt}}$  is finite, the size distortion is smaller than  $O(T^{-2/5})$ , which is the size distortion for the conventional Wald test using the MSE optimal  $K$ , that is,  $K$  is set to be  $O(T^{4/5})$ .

The formula for  $K_{\text{opt}}$  depends on the noncentrality parameter  $\delta^2$ . For practical implementation, we suggest choosing  $\delta^2$  such that the first order power of the test, as measured by  $1 - G_{p,\delta^2}(\chi_p^\alpha)$ , is 75%. That is, we solve  $1 - G_{p,\delta^2}(\chi_p^\alpha) = 75\%$  for a given  $p$  and a given significance level  $\alpha$ . As usual, we consider  $\alpha = 5\%$  and  $10\%$ . The value of  $\delta^2$  can be easily computed using standard statistical programs. For convenience, we provide the value of  $\delta^2$  for some  $\alpha$  and  $p$  combinations in the table below.

Values of $\delta^2$ for different $(\alpha, p)$ combinations										
p	p=1	p=2	p=3	p=4	p=5	p=6	p=7	p=8	p=9	p=10
$\alpha = 5\%$	6.94	8.59	9.76	10.72	11.55	12.29	12.96	13.59	14.17	14.72
$\alpha = 10\%$	5.38	6.77	7.76	8.57	9.27	9.90	10.47	10.99	11.49	11.95

To sum up, when the size distortion is expected to be negative, the expansion rate of the optimal  $K$  is  $O(T^{2/3})$ . When the size distortion is expected to be positive, the optimal  $K$  has an expansion rate that increases with the tolerance on the type I error. The expansion can range from  $O(1)$  when the tolerance is very low to  $O(T)$  when the tolerance is very high.

## 6.2 Data Driven Implementation

The optimal bandwidth in (21) can be written as  $K_{\text{opt}} = K_{\text{opt}}(\bar{B})$ . It involves unknown parameter  $\bar{B}$  which can be estimated nonparametrically (e.g. Newey and West (1994)) or by a standard plug-in procedure based on a simple parametric model like VAR (e.g.

Andrews (1991)). Both methods achieve a valid order of magnitude and the procedure is analogous to conventional data-driven methods for LRV estimation.

We focus the discussion on the plug-in procedure, which involves the following steps. First, we estimate the model using the OLS estimator and compute the residuals  $\{\hat{u}_t\}$ . Second, we specify a multivariate approximating parametric model and fit the model to  $\{\hat{u}_t\}$  by the standard OLS method. Third, we treat the fitted model as if it were the true model for the process  $\{u_t\}$  and compute  $\bar{B}$  as a function of the parameters of the parametric model. Plugging the estimate  $\bar{B}$  into (21) gives the automatic bandwidth  $\hat{K}$ .

As in the case of MSE-optimal bandwidth choice, the automatic bandwidth considered here deviates from the finite sample optimal one due to the error introduced by estimation, the use of approximating parametric models, and the approximation inherent in the asymptotic formula employed. It is hoped that in practical work the deviation is not large so that the resulting test still has super performance. Some simulation evidence reported in the next section supports this argument.

Suppose we use a VAR(1) as the approximating parametric model for  $u_t$ . Let  $\hat{A}$  be the estimated parameter matrix and  $\hat{\Sigma}$  be the estimated innovation covariance matrix, then the plug-in estimates of  $\Omega$  and  $B$  are

$$\hat{\Omega} = (\mathbb{I}_n - \hat{A})^{-1} \hat{\Sigma} (\mathbb{I}_n - \hat{A}')^{-1}, \quad (22)$$

$$\begin{aligned} \hat{B} = & -\frac{2\pi^2}{3} (\mathbb{I}_{d_2} - \hat{A})^{-3} \left( \hat{A} \hat{\Sigma} + \hat{A}^2 \hat{\Sigma} \hat{A}' + \hat{A}^2 \hat{\Sigma} - 6 \hat{A} \hat{\Sigma} \hat{A}' \right. \\ & \left. + \hat{\Sigma} (\hat{A}')^2 + \hat{A} \hat{\Sigma} (\hat{A}')^2 + \hat{\Sigma} \hat{A}' \right) (\mathbb{I}_{d_2} - \hat{A}')^{-3}. \end{aligned} \quad (23)$$

For the plug-in estimates under a general VAR(p) model, we refer to Andrews (1991) for the corresponding formulae. Given the plug-in estimates of  $\Omega$  and  $B$ , the data-driven automatic bandwidth can be computed as

$$\hat{K}_{opt} = \hat{K}_{opt}(\bar{B}(R, \hat{B}, \hat{\Omega})). \quad (24)$$

It should be pointed out that the computational cost involved in this automatic bandwidth is the same as that of the conventional plug-in bandwidth.

## 7 Simulation Evidence

This section provides some simulation evidence on the finite sample performance of the  $F^*$  test based on the plug-in procedure that minimizes the type II error while controlling for the type I error.

As in Vogelsang and Franses (2005), we set  $n = 6$ . The error follows either a VAR(1) or VMA(1) process:

$$\begin{aligned} u_t &= Au_{t-1} + \sqrt{1 - \rho^2} \varepsilon_t \\ u_t &= A\varepsilon_{t-1} + \sqrt{1 - \rho^2} \varepsilon_t \end{aligned}$$

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

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

Table 1: Type I error of different tests for VAR(1) error with  $T = 100, \kappa = 1.1$

	<b>New</b>	<b>MSE</b>	<b>Hybrid</b>	<b>VF</b>		<b>New</b>	<b>MSE</b>	<b>Hybrid</b>	<b>VF</b>
	$p = 1$					$p = 2$			
$\rho=0$	0.0516	0.0580	0.0484	0.0525		0.0508	0.0685	0.0439	0.0438
$\rho=0.25$	0.0830	0.0857	0.0621	0.0602		0.0943	0.1128	0.0607	0.0553
$\rho=0.50$	0.0959	0.1096	0.0670	0.0707		0.1003	0.1692	0.0670	0.0712
$\rho=0.75$	0.1043	0.1595	0.0908	0.0936		0.1170	0.2609	0.0893	0.1111
	$p = 3$					$p = 6$			
$\rho=0$	0.0494	0.0856	0.0460	0.0505		0.0480	0.1677	0.0492	0.0482
$\rho=0.25$	0.1085	0.1590	0.0643	0.0644		0.1526	0.3675	0.0712	0.0808
$\rho=0.50$	0.1181	0.2604	0.0692	0.0898		0.1594	0.5807	0.0697	0.1429
$\rho=0.75$	0.1539	0.4172	0.1142	0.1625		0.2575	0.7916	0.1958	0.3291

where  $\mathbb{J}_n$  is a matrix of ones. The parameter  $\mu$  determines the degree of dependence among the time series considered. When  $\mu = 0$ , the six series are uncorrelated with each other. When  $\mu = 1$ , the six series have the same pairwise correlation coefficient 0.5. The variance-covariance matrix of  $u_t$  is normalized so that the variance of each series  $u_{it}$  is equal to one for all values of  $|\rho| < 1$ . For the VAR(1) process,  $\Omega = (1 - \rho^2) (\mathbb{I} - A)^{-1} \Sigma (\mathbb{I} - A')^{-1}$ . For the VMA(1) process  $\Omega = (1 - \rho^2) \left( \mathbb{I} + A/\sqrt{1 - \rho^2} \right) \Sigma \left( A/\sqrt{1 - \rho^2} \right)'$ .

For the model parameters, we take  $\rho = 0, 0.25, 0.50, 0.75$  and set  $\mu = 0$  and 1. We set the intercepts and slopes to zero as the tests we consider are invariant to those parameters. For each test, we consider two significance levels  $\alpha = 5\%$  and  $\alpha = 10\%$ , two different choices of the tolerance parameter:  $\kappa = 1.1$  and 1.2, and three different sample sizes  $T = 100, 200, 500$ .

As in Vogelsang and Franses (2005), we consider the following null hypotheses:

$$\begin{aligned}
 H_{01} &: \beta_1 = 0, \\
 H_{02} &: \beta_1 = \beta_2 = 0, \\
 H_{03} &: \beta_1 = \beta_2 = \beta_3 = 0, \\
 H_{04} &: \beta_1 = \beta_2 = \dots = \beta_6 = 0,
 \end{aligned}$$

where  $p = 1, 2, 3, 6$ , respectively. The corresponding matrix  $R$  is the first  $p$  rows of the identity matrix  $\mathbb{I}_6$ . To explore the finite sample size of the tests, we generate data under these null hypotheses. To compare the power of the tests, we generate data under the local alternative hypothesis  $H_1(\delta^2)$ .

We examine the finite sample performance of three different testing methods. The first one is the new  $F^*$  test which is based on the modified Wald-statistic and critical values from the F-distribution and uses the test-optimal  $K$ . The second one is the conventional Wald test, which is based on the unmodified Wald statistic and critical values from  $\chi^2$  distribution and uses the MSE-optimal  $K$ . The last one is the test proposed by Vogelsang and Franses (2005) which is based on the Bartlett kernel LRV estimator with bandwidth equal to the sample size and uses the nonstandard asymptotic theory. The three methods are referred as ‘New’, ‘MSE’, and ‘VF’ respectively in the tables and figures below.

Table 1 gives the type I error of the three testing methods for the VAR(1) error with sample size  $T = 100$ , tolerance parameter  $\kappa = 1.1$  and  $\mu = 0$ . The table also includes

Table 2: Type I error for various tests based on VMA(1) error with  $T = 100, \kappa = 1.1$

	<b>New</b>	<b>MSE</b>	<b>Hybrid</b>	<b>VF</b>		<b>New</b>	<b>MSE</b>	<b>Hybrid</b>	<b>VF</b>
	$p = 1$					$p = 2$			
$\rho=0$	0.0516	0.0580	0.0484	0.0525		0.0508	0.0685	0.0439	0.0438
$\rho=0.25$	0.0705	0.0746	0.0556	0.0560		0.0766	0.0968	0.0526	0.0516
$\rho=0.50$	0.0649	0.0836	0.0558	0.0577		0.0663	0.1208	0.0502	0.0546
$\rho=0.75$	0.0624	0.0866	0.0547	0.0581		0.0623	0.1336	0.0488	0.0544
	$p = 3$					$p = 6$			
$\rho=0$	0.0494	0.0856	0.0460	0.0505		0.0480	0.1677	0.0492	0.0482
$\rho=0.25$	0.0862	0.1346	0.0544	0.0590		0.1192	0.3133	0.0615	0.0685
$\rho=0.50$	0.0804	0.1750	0.0541	0.0627		0.0954	0.4569	0.0551	0.0779
$\rho=0.75$	0.0771	0.1986	0.0531	0.0628		0.0856	0.4919	0.0480	0.0796

a hybrid procedure which employs the MSE-optimal  $K$  and critical values from the  $F$ -distribution. The only difference between the conventional method and the hybrid method lies in the critical value used. The significance level is 5% which is also the nominal type I error. Several patterns emerge. First, as it is clear from the table, the conventional method has a large size distortion. The size distortion increases with both the error dependence and the number of restrictions being tested. This result is consistent with our theoretical analysis. The size distortion can be very severe. For example, when  $\rho = 0.75$  and  $p = 6$ , the empirical type I error of the test is 0.7916, which is far from 0.05, the nominal type I error. Using the  $F$  critical values eliminates the distortion to a great extent. This is especially true when the size distortion is large. Intuitively, larger size distortion occurs when  $K$  is smaller so that the LRV estimator has a larger variation. This is the scenario where the difference between the  $F$  critical values and  $\chi^2$  critical values is larger. Second, the size distortion of the new method and the VF method is substantially smaller than the conventional method. This is because both tests employ asymptotic approximations that capture the estimation uncertainty of the LRV estimator. The smaller size distortion of the new method is also consistent with that of the hybrid method as both are based on  $F$ -approximations. Third, compared with the VF method, the new method has similar size distortion. Since the bandwidth is set equal to the sample size, the VF method is designed to achieve the smallest possible size distortion. Given this observation, we can conclude that the new method succeeds in controlling the type I error.

Table 2 presents the simulated type I errors for the VMA(1) error process. The qualitative observations for the VAR(1) error remain valid. In fact, these qualitative observations hold for other parameter configurations such as different sample sizes and different values of  $\mu$ . All else being equal, the size distortion of the new method for  $\kappa = 1.2$  is slightly larger than that for  $\kappa = 1.1$ . This is expected as we have a higher tolerance for the type I error when the value of  $\kappa$  is larger.

Figures 1-4 present the finite sample power under the VAR(1) error for different values of  $p$ . 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. Two observations can be drawn from these figures. First, the power of the new test is always higher than the conventional Wald test.

The difference is largest when the autocorrelation coefficient is in the medium range around  $\rho = 0.5$ . This result demonstrates the advantage of using the criterion that focuses on the testing problem at hand. Consistent with the asymptotics result, the focused criterion has superior performance in finite samples than the MSE criterion which is not suitable for hypothesis testing. Second, the new test has higher power than the VF test in most cases except when the error dependence is very high and the number of restrictions is large. When the error dependence is low, the selected  $K$  value is relatively large and the variance of the associated LRV estimator is small. In contrast, the LRV estimator used in the VF test is inconsistent and hence has a large variance. As a result, the new test is more powerful than the VF test. On the other hand, when the error dependence is very large, the selected  $K$  values are very small. In this case, both the VF test and the new test employ a LRV estimator with large variance. The VF test can be more powerful in this scenario.

Figures 5-8 present the power curves under the VMA(1) error. The figures reinforce and strengthen the observations for the VAR(1) error. It is clear now that the new test is more powerful than both the conventional Wald test and the VF test. This is true for all parameter combinations considered.

In simulations not reported here, we have considered VAR(1) and VMA(1) errors with negative values of  $\rho$  and hypotheses of the form  $\beta_1 = \beta_2 = \dots = \beta_{j_0}$  for some  $j_0 \geq 2$ . For some of these configurations,  $\bar{B} > 0$ . Regardless of the sign of  $\bar{B}$ , in terms of the type I error and size adjusted power, the performance of the new  $F^*$  test is at least as good as the conventional Wald test and often much better. It also dominates the VF test in most scenarios considered. Details are available upon request.

## 8 Conclusion

The paper proposes a novel approach to multivariate trend inference in the presence of nonparametric autocorrelation. The inference procedure is based on a novel LRV estimator. As a multiple window estimator, the LRV estimator is asymptotically invariant to the intercept and trend parameters. This property releases us from worrying about the estimation uncertainty of those parameters. Another advantage of the LRV estimator is that the associated (modified) Wald statistic converges to a standard distribution regardless of the asymptotic specification of the smoothing parameter. This property releases practitioners from the computation burden of simulating nonstandard critical values. We propose a new method to select the smoothing parameter in the new LRV estimator. The optimal smoothing parameter is selected to minimize the type II error hence maximize the power of the test while controlling for the type I error. The idea is in the spirit of the Neyman-Pearson Lemma. Monte Carlo simulations show that our inference procedure enjoys superior performance in finite samples.

There are many extensions to the current paper. One possibility is to consider near-integrated error processes and select basis functions so that the nice properties of the new LRV estimator and associated tests can be maintained. It is also desirable to consider general polynomial trends although linear trend is most common in empirical applications. Further, the idea of optimal smoothing choice and the inference procedure given here may be used in general linear and non-linear models with stationary and unit root nonstationary time series.



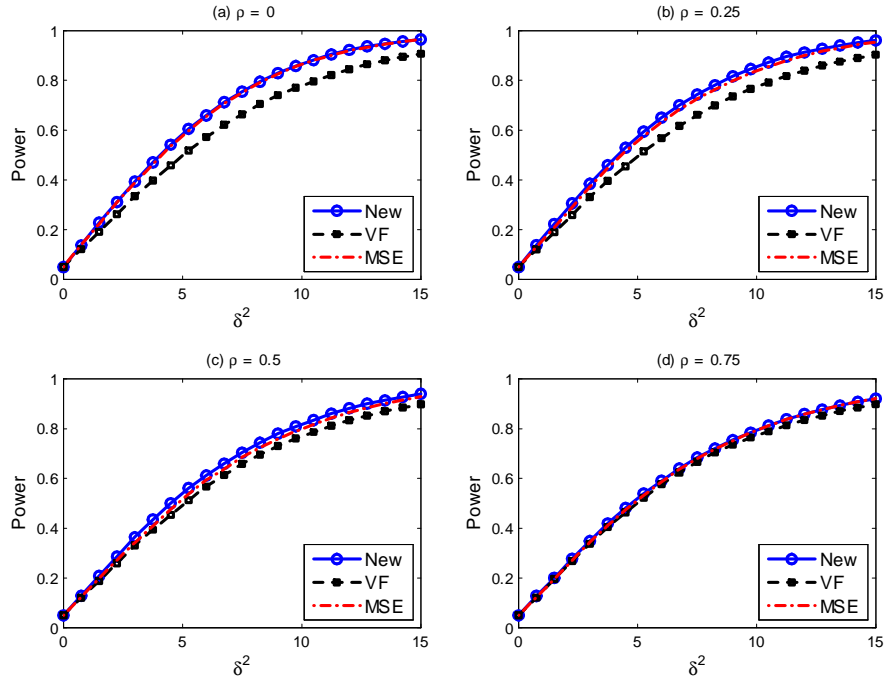


Figure 1: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with  $T = 100, \kappa = 1.1$  and  $p = 1$ .

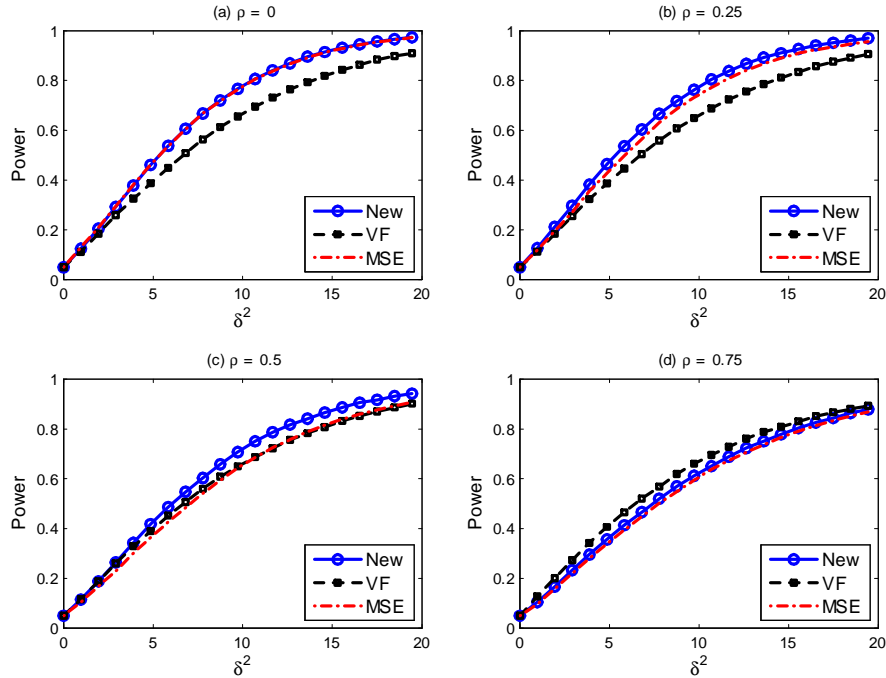


Figure 2: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with  $T = 100, \kappa = 1.1$  and  $p = 2$ .

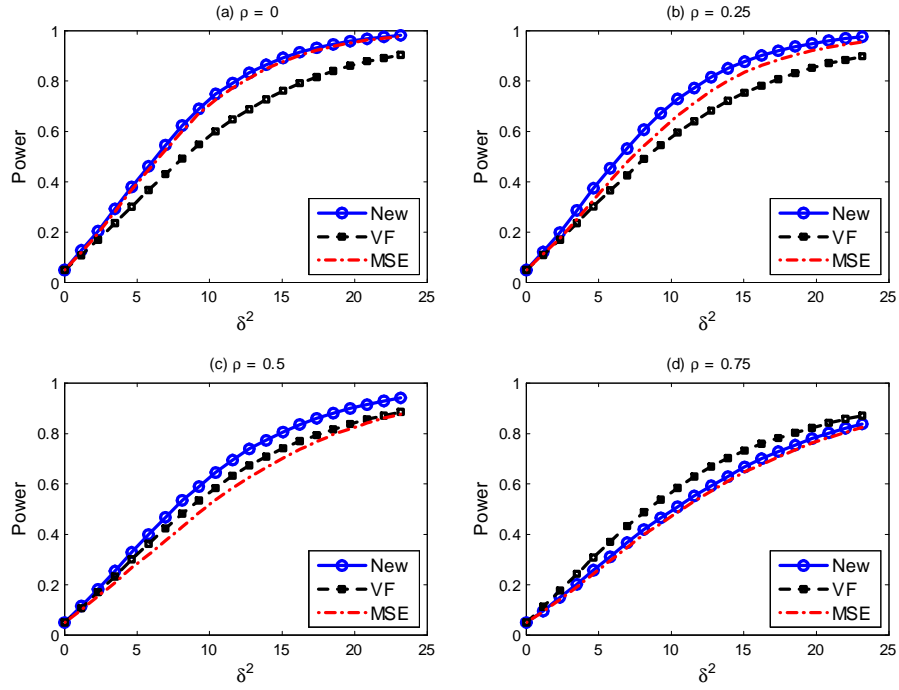


Figure 3: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with  $T = 100$ ,  $\kappa = 1.1$  and  $p = 3$ .

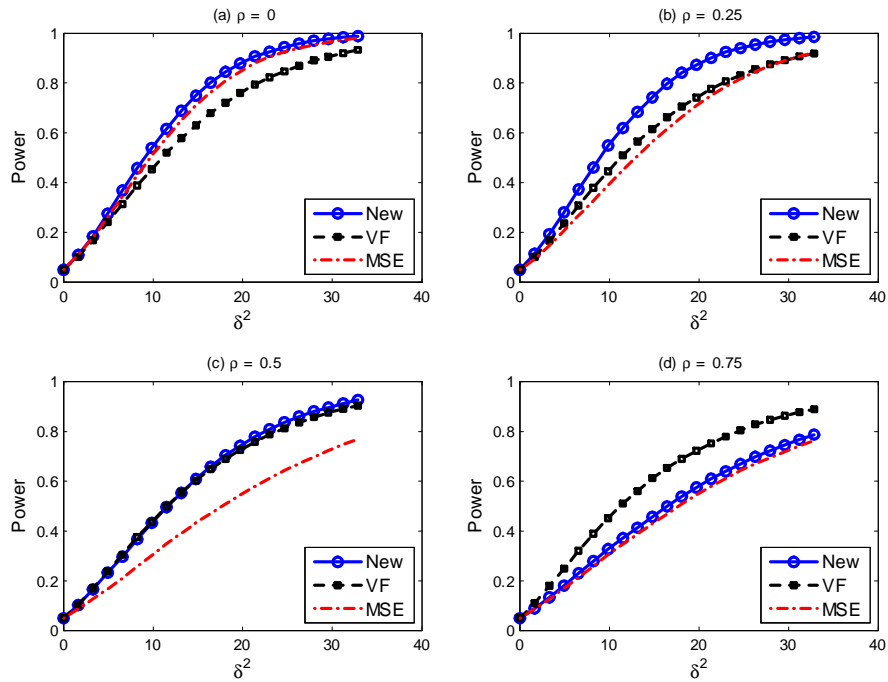


Figure 4: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with  $T = 100$ ,  $\kappa = 1.1$  and  $p = 6$ .

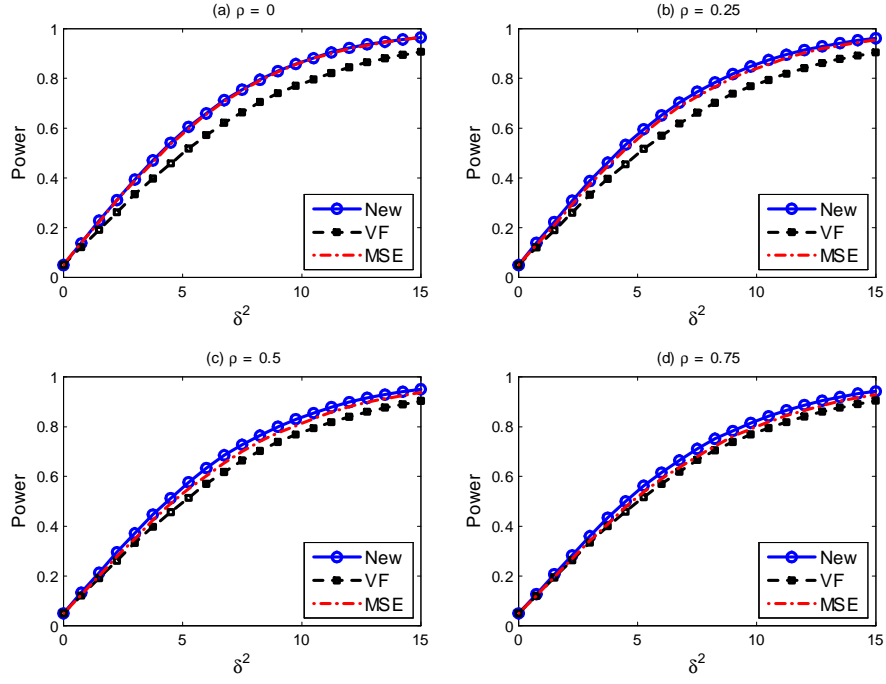


Figure 5: Size-adjusted Power of Different Testing Procedures for VMA(1) Error with  $T = 100, \kappa = 1.1$  and  $p = 1$ .

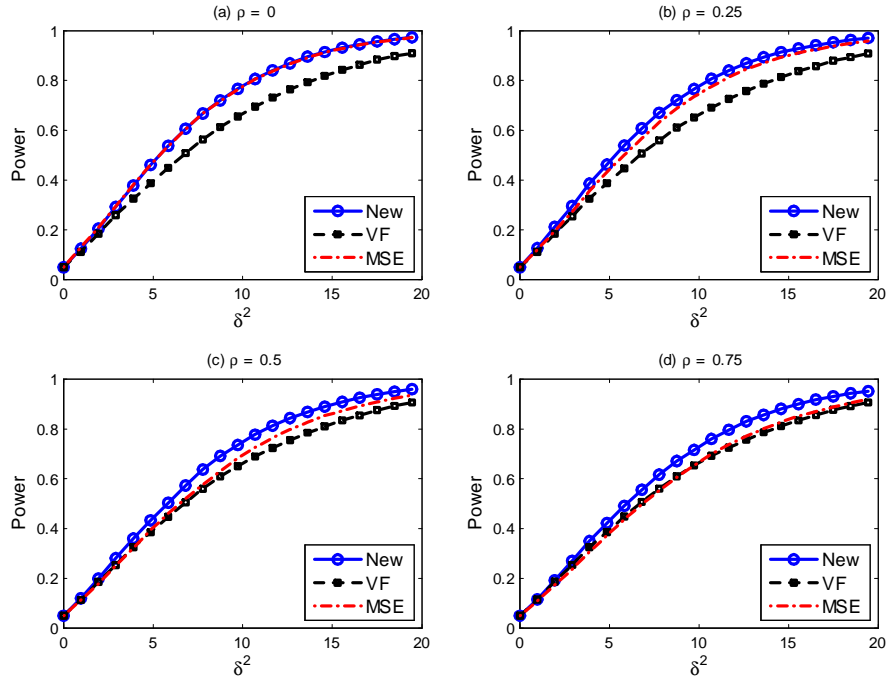


Figure 6: Size-adjusted Power of Different Testing Procedures for VMA(1) Error with  $T = 100, \kappa = 1.1$  and  $p = 2$ .

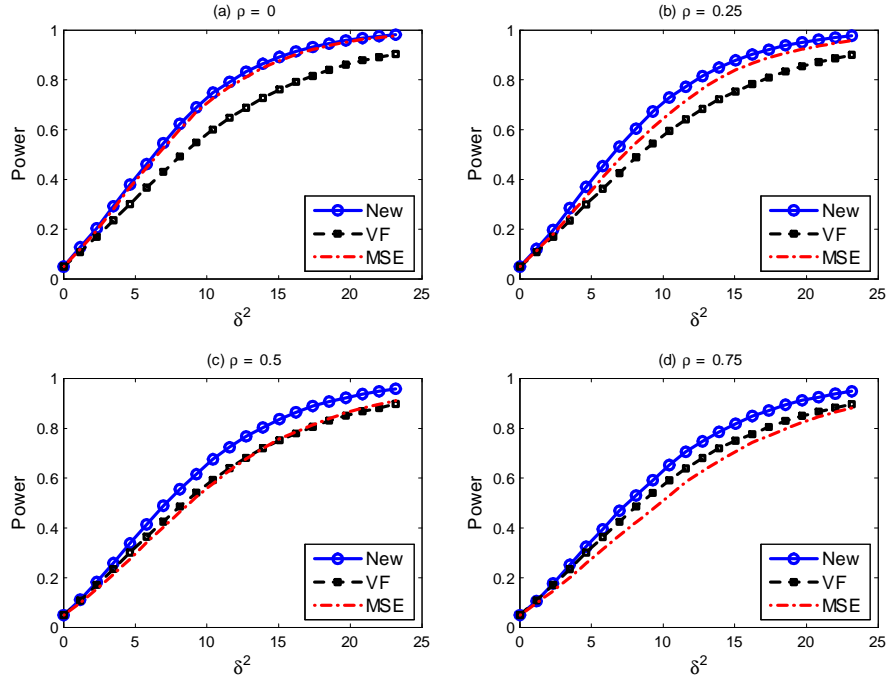


Figure 7: Size-adjusted Power of Different Testing Procedures for VMA(1) Error with  $T = 100, \kappa = 1.1$  and  $p = 3$ .

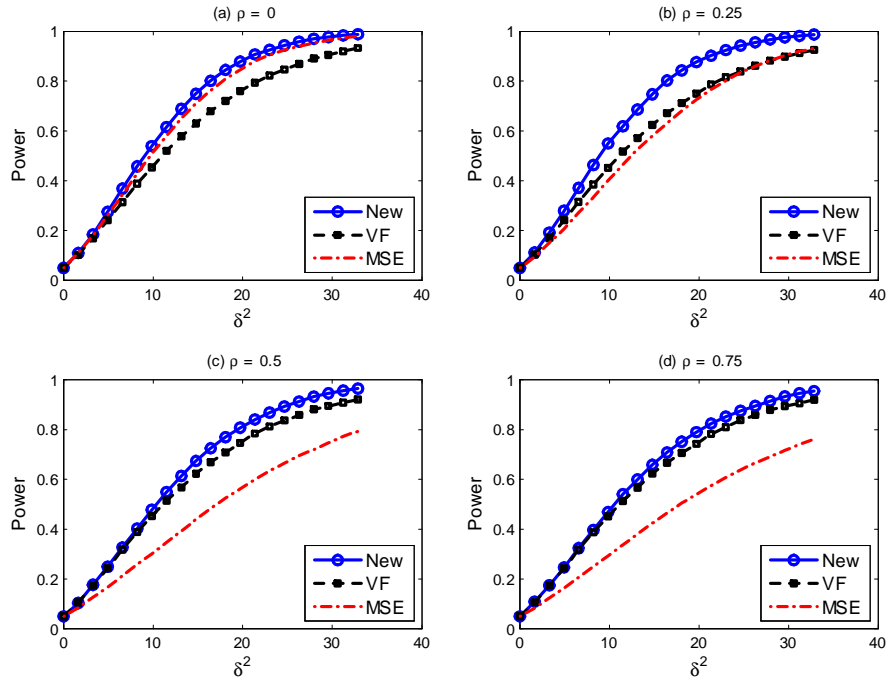


Figure 8: Size-adjusted Power of Different Testing Procedures for VMA(1) Error with  $T = 100, \kappa = 1.1$  and  $p = 2$ .

## 9 Appendix of Proofs

**Proof of Theorem 2. Part (a).** Let  $\bar{x}$  be the mean of the sequence  $x_1, \dots, x_T$ , then

$$\begin{aligned}
\frac{1}{\sqrt{T}} \sum_{t=1}^T \phi_k \left( \frac{t}{T} \right) \hat{u}_t &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \phi_k \left( \frac{t}{T} \right) \left[ u_t - \bar{u} - (t - \bar{t}) (\hat{\beta} - \beta) \right] \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^T \phi_k \left( \frac{t}{T} \right) u_t - \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{1}{T} \sum_{s=1}^T \phi_k \left( \frac{s}{T} \right) u_t \\
&\quad - \frac{1}{\sqrt{T}} \sum_{t=1}^T \left[ \sum_{s=1}^T \phi_k \left( \frac{s}{T} \right) (s - \bar{s}) \right] \left[ \sum_{s=1}^T (s - \bar{s})^2 \right]^{-1} (t - \bar{t}) u_t \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^T \left\{ \phi_k \left( \frac{t}{T} \right) - \frac{1}{T} \sum_{s=1}^T \phi_k \left( \frac{s}{T} \right) - \left[ \frac{\sum_{s=1}^T \phi_k \left( \frac{s}{T} \right) (s - \bar{s})}{\sum_{s=1}^T (s - \bar{s})^2} \right] (t - \bar{t}) \right\} u_t \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^T \check{\phi}_k \left( \frac{t}{T} \right) u_t,
\end{aligned}$$

where

$$\check{\phi}_k \left( \frac{t}{T} \right) = \phi_k \left( \frac{t}{T} \right) - \frac{1}{T} \sum_{s=1}^T \phi_k \left( \frac{s}{T} \right) - \left[ \frac{\sum_{s=1}^T \phi_k \left( \frac{s}{T} \right) (s - \bar{s})}{\sum_{s=1}^T (s - \bar{s})^2} \right] (t - \bar{t}).$$

Some calculations show that

$$\frac{1}{T} \sum_{s=1}^T \phi_k \left( \frac{s}{T} \right) = 0, \quad \sum_{s=1}^T \phi_k \left( \frac{s}{T} \right) (s - \bar{s}) = \frac{T}{2}, \quad \sum_{s=1}^T (s - \bar{s})^2 = \frac{1}{12} T (T^2 - 1).$$

Hence,

$$\begin{aligned}
\check{\phi}_k \left( \frac{t}{T} \right) &= \phi_k \left( \frac{t}{T} \right) - \frac{6}{(T^2 - 1)} \left( t - \frac{T}{2} - \frac{1}{2} \right) \\
&= \phi_k \left( \frac{t}{T} \right) - \frac{6T}{(T^2 - 1)} \left( \frac{t}{T} - \frac{1}{2} - \frac{1}{2T} \right) \\
&= \phi_k \left( \frac{t}{T} \right) + O \left( \frac{1}{T} \right).
\end{aligned}$$

Define

$$\tilde{\Omega} = \frac{1}{K} \sum_{k=1}^K \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^T \phi_k \left( \frac{t}{T} \right) u_t \right] \left[ \frac{1}{\sqrt{T}} \sum_{s=1}^T \phi_k \left( \frac{s}{T} \right) u_s \right]'. \quad (25)$$

It is easy to see that the bias and variance of  $\tilde{\Omega}$  are the same as those of  $\hat{\Omega}$  up to order  $O(1/T)$ . Therefore it suffices to compute the bias and variance of  $\tilde{\Omega}$ .

It follows from equation (22) in Phillips (2005) that

$$E\tilde{\Omega} - \Omega = \sum_{h=-L_T}^{L_T} \frac{1}{K} \sum_{k=1}^K \frac{1}{T} \sum_{1 \leq t, t+h \leq T} \phi_k \left( \frac{t}{T} \right) \left[ \phi_k \left( \frac{t+h}{T} \right) - \phi_k \left( \frac{t}{T} \right) \right] \Gamma_u(h) + o \left( \frac{K^2}{T^2} \right) + O \left( \frac{1}{T} \right),$$

where  $\Gamma_u(h) = Eu_t u'_{t-h}$  and  $L_T$  satisfies

$$\frac{T}{L_T^{3/2}} + \frac{L_T K}{T} = o(1). \quad (26)$$

Next consider

$$\begin{aligned} & \frac{1}{T} \sum_{1 \leq t, t+h \leq T} \phi_k \left( \frac{t}{T} \right) \left[ \phi_k \left( \frac{t+h}{T} \right) - \phi_k \left( \frac{t}{T} \right) \right] \\ &= \frac{2}{T} \sum_{1 \leq t, t+h \leq T} \cos \frac{2\pi kt}{T} \left[ \cos \frac{2\pi k(t+h)}{T} - \cos \frac{2\pi kt}{T} \right] \\ &= \frac{1}{T} \sum_{1 \leq t, t+h \leq T} \left\{ \cos \left( \frac{2\pi k(2t+h)}{T} \right) + \cos \left( \frac{2\pi kh}{T} \right) - \left( \cos \left( \frac{4\pi kt}{T} \right) + 1 \right) \right\} \\ &= \left( 1 - \frac{|h|}{T} \right) \left( \cos \left( \frac{2\pi kh}{T} \right) - 1 \right) \\ & \quad + \frac{1}{T} \sum_{1 \leq t, t+h \leq T} \left( \cos \left( \frac{4\pi k(t+\frac{1}{2}h)}{T} \right) - \cos \left( \frac{4\pi kt}{T} \right) \right). \end{aligned} \quad (27)$$

$$(28)$$

Taking the first term of (28), averaging over  $k$  and using the fact that  $\|h\| \leq L_n$  and  $L_n$  satisfies (26), we get

$$\begin{aligned} \frac{1}{K} \sum_{k=1}^K \left( \cos \left( \frac{2\pi kh}{T} \right) - 1 \right) &= -\frac{1}{K} \sum_{k=1}^K \frac{1}{2} \left( \frac{2\pi kh}{T} \right)^2 (1 + o(1)) \\ &= -2\pi^2 h^2 \frac{K^2}{T^2} \frac{1}{K} \sum_{k=1}^K \left( \frac{k}{K} \right)^2 (1 + o(1)) \\ &= \frac{-2\pi^2 h^2 K^2}{3 T^2} (1 + o(1)). \end{aligned}$$

Approximating the sums by integrals, we can show that

$$\frac{1}{K} \sum_{k=1}^K \frac{1}{T} \sum_{1 \leq t, t+h \leq T} \left[ \cos \left( \frac{4\pi k(t+\frac{1}{2}h)}{T} \right) - \cos \left( \frac{4\pi kt}{T} \right) \right] = o \left( \frac{K^2 L_n^2}{T^2} \right).$$

Hence

$$\frac{1}{K} \sum_{k=1}^K \frac{1}{T} \sum_{1 \leq t, t+h \leq T} \phi_k \left( \frac{t}{T} \right) \left[ \phi_k \left( \frac{t+h}{T} \right) - \phi_k \left( \frac{t}{T} \right) \right] = -\frac{2\pi^2 h^2 K^2}{3 T^2} (1 + o(1)) + O \left( \frac{1}{T} \right).$$

So

$$\begin{aligned} E\tilde{\Omega} - \Omega &= -\frac{2\pi^2 K^2}{3 T^2} \sum_{h=-L_T}^{L_T} \left( 1 - \frac{|h|}{T} \right) h^2 \Gamma_u(h) + o \left( \frac{K^2}{T^2} \right) + O \left( \frac{1}{T} \right) \\ &= -\frac{2\pi^2 K^2}{3 T^2} \sum_{h=-\infty}^{\infty} h^2 \Gamma_u(h) + o \left( \frac{K^2}{T^2} \right) + O \left( \frac{1}{T} \right) \end{aligned}$$

as desired.

**Part (b).** Using the device in Phillips and Solo (1992), we have the BN decomposition

$$u_t = C(1)\varepsilon_t + \tilde{u}_{t-1} - \tilde{u}_t \text{ for } \tilde{u}_t = \sum_{j=0}^{\infty} \tilde{C}_j \varepsilon_{t-j}, \quad \tilde{C}_j = \sum_{s=j+1}^{\infty} C_s.$$

Plugging this into the definition of  $\tilde{\Omega}$  yields

$$\tilde{\Omega} = R_1 + R_2 + R_2' + R_3,$$

where

$$\begin{aligned} R_1 &= \frac{1}{K} \sum_{k=1}^K \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^T \phi_k \left( \frac{t}{T} \right) C(1) \varepsilon_t \right] \left[ \frac{1}{\sqrt{T}} \sum_{\tau=1}^T \phi_k \left( \frac{\tau}{T} \right) C(1) \varepsilon_\tau \right]', \\ R_2 &= \frac{1}{K} \sum_{k=1}^K \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^T \phi_k \left( \frac{t}{T} \right) C(1) \varepsilon_t \right] \left[ \frac{1}{\sqrt{T}} \sum_{\tau=1}^T \phi_k \left( \frac{\tau}{T} \right) (\tilde{u}_{\tau-1} - \tilde{u}_\tau) \right]', \\ R_3 &= \frac{1}{K} \sum_{k=1}^K \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^T \phi_k \left( \frac{t}{T} \right) (\tilde{u}_{t-1} - \tilde{u}_t) \right] \left[ \frac{1}{\sqrt{T}} \sum_{\tau=1}^T \phi_k \left( \frac{\tau}{T} \right) (\tilde{u}_{\tau-1} - \tilde{u}_\tau) \right]'. \end{aligned}$$

We proceed to show that  $Etr(\text{vec}(R_2)\text{vec}(R_2)') = O(1/T)$ . Note that

$$\frac{1}{\sqrt{T}} \sum_{\tau=1}^T \phi_k \left( \frac{\tau}{T} \right) (\tilde{u}_{\tau-1} - \tilde{u}_\tau) = \frac{1}{\sqrt{T}} \sum_{\tau=1}^{T-1} \left[ \phi_k \left( \frac{\tau+1}{T} \right) - \phi_k \left( \frac{\tau}{T} \right) \right] \tilde{u}_\tau + \frac{1}{\sqrt{T}} (\tilde{u}_0 - \tilde{u}_T),$$

we have

$$R_2 = R_2^{(1)} + R_2^{(2)} + R_2^{(3)},$$

where

$$\begin{aligned} R_2^{(1)} &= \frac{1}{K} \sum_{k=1}^K \frac{1}{T} \sum_{t=1}^T \phi_k \left( \frac{t}{T} \right) C(1) \varepsilon_t \tilde{u}_0', \\ R_2^{(2)} &= -\frac{1}{K} \sum_{k=1}^K \frac{1}{T} \sum_{t=1}^T \phi_k \left( \frac{t}{T} \right) C(1) \varepsilon_t \tilde{u}_T', \\ R_2^{(3)} &= \frac{1}{K} \sum_{k=1}^K \frac{1}{T} \sum_{t=1}^T \sum_{\tau=1}^{T-1} \phi_k \left( \frac{t}{T} \right) \left[ \phi_k \left( \frac{\tau+1}{T} \right) - \phi_k \left( \frac{\tau}{T} \right) \right] C(1) \varepsilon_t \tilde{u}_\tau'. \end{aligned}$$

Let

$$\bar{\phi} \left( \frac{t}{T} \right) = \frac{1}{K} \sum_{k=1}^K \phi_k \left( \frac{t}{T} \right) = O(1),$$

then  $Etr \left( \text{vec} \left( R_2^{(1)} \right) \text{vec} \left( R_2^{(1)} \right)' \right)$  is

$$\begin{aligned}
& \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \bar{\phi} \left( \frac{t}{T} \right) \bar{\phi} \left( \frac{s}{T} \right) Etr \left( \text{vec} \left( C(1) \varepsilon_t \tilde{u}'_0 \right) \text{vec} \left( C(1) \varepsilon_s \tilde{u}'_0 \right)' \right) \\
&= \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \bar{\phi} \left( \frac{t}{T} \right) \bar{\phi} \left( \frac{s}{T} \right) Etr \left( (\tilde{u}_0 \otimes C(1) \varepsilon_t) (\tilde{u}_0 \otimes C(1) \varepsilon_s)' \right) \\
&= \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \bar{\phi} \left( \frac{t}{T} \right) \bar{\phi} \left( \frac{s}{T} \right) Etr \left( (\tilde{u}_0 \otimes C(1) \varepsilon_t) (\tilde{u}'_0 \otimes \varepsilon'_s C'(1)) \right) \\
&= \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \bar{\phi} \left( \frac{t}{T} \right) \bar{\phi} \left( \frac{s}{T} \right) Etr \left( \tilde{u}_0 \tilde{u}'_0 \otimes C(1) \varepsilon_t \varepsilon'_s C'(1) \right) \\
&= \frac{1}{T^2} \sum_{t=1}^T \left[ \bar{\phi} \left( \frac{t}{T} \right) \right]^2 E \|\tilde{u}_0\|^2 \|C(1)\|^2 = O\left(\frac{1}{T}\right),
\end{aligned}$$

where the first equality follows from the fact that for  $n \times 1$  vectors  $A$  and  $B$ ,  $\text{vec}(AB') = B \otimes A$ , the third and fourth equalities follow from the rules that  $(A \otimes B)(C \otimes D) = AC \otimes BD$  and  $\text{tr}(C \otimes D) = \text{tr}(C)\text{tr}(D)$  respectively.

Similarly,  $Etr \left( \text{vec} \left( R_2^{(2)} \right) \text{vec} \left( R_2^{(2)} \right)' \right)$  is equal to

$$\begin{aligned}
& \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \bar{\phi} \left( \frac{t}{T} \right) \bar{\phi} \left( \frac{s}{T} \right) Etr \left( \tilde{u}_T \tilde{u}'_T \otimes C(1) \varepsilon_t \varepsilon'_s C'(1) \right) \\
&= \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \bar{\phi} \left( \frac{t}{T} \right) \bar{\phi} \left( \frac{s}{T} \right) Etr \left( \tilde{C}_p \varepsilon_{T-p} \varepsilon'_{T-q} \tilde{C}'_q \otimes C(1) \varepsilon_t \varepsilon'_s C'(1) \right) \\
&= \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{p=0}^{T-1} \sum_{q=0}^{T-1} \bar{\phi} \left( \frac{t}{T} \right) \bar{\phi} \left( \frac{s}{T} \right) Etr \left( \tilde{C}_p \varepsilon_{T-p} \varepsilon'_{T-q} \tilde{C}'_q \otimes C(1) \varepsilon_t \varepsilon'_s C'(1) \right) \\
&\quad + \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{p=T}^{\infty} \sum_{q=T}^{\infty} \bar{\phi} \left( \frac{t}{T} \right) \bar{\phi} \left( \frac{s}{T} \right) Etr \left( \tilde{C}_p \varepsilon_{T-p} \varepsilon'_{T-q} \tilde{C}'_q \otimes C(1) \varepsilon_t \varepsilon'_s C'(1) \right) \\
&= \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{p=1}^T \sum_{q=1}^T \bar{\phi} \left( \frac{t}{T} \right) \bar{\phi} \left( \frac{s}{T} \right) Etr \left( \tilde{C}_{T-p} \varepsilon_p \varepsilon'_q \tilde{C}'_{T-q} \otimes C(1) \varepsilon_t \varepsilon'_s C'(1) \right)
\end{aligned} \tag{29}$$

$$\begin{aligned}
& + \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \bar{\phi} \left( \frac{t}{T} \right) \bar{\phi} \left( \frac{s}{T} \right) Etr \left( \tilde{C}_{p-T} \varepsilon_{-p} \varepsilon'_{-q} \tilde{C}'_{q-T} \otimes C(1) \varepsilon_t \varepsilon'_s C'(1) \right).
\end{aligned} \tag{30}$$

It is easy to show that (30) is

$$\begin{aligned}
& \frac{1}{T^2} \sum_{t=1}^T \sum_{p=0}^{\infty} \left[ \bar{\phi}\left(\frac{t}{T}\right) \right]^2 \operatorname{tr} \left( \tilde{C}_{p-T} \tilde{C}'_{p-T} \otimes C(1)C'(1) \right) \\
& \leq \frac{1}{T^2} \sum_{t=1}^T \sum_{p=0}^{\infty} \left[ \bar{\phi}\left(\frac{t}{T}\right) \right]^2 \operatorname{tr} \left( \left( \tilde{C}_{p-T} \tilde{C}'_{p-T} \right) \operatorname{tr} \left( C(1)C'(1) \right) \right) \\
& \leq \frac{1}{T^2} \sum_{t=1}^T \sum_{p=0}^{\infty} \left[ \bar{\phi}\left(\frac{t}{T}\right) \right]^2 \left\| \tilde{C}_{p-T} \right\|^2 \|C(1)\|^2 \\
& \leq \frac{1}{T^2} \sum_{t=1}^T \left[ \bar{\phi}\left(\frac{t}{T}\right) \right]^2 \left( \sum_{p=0}^{\infty} \left\| \tilde{C}_p \right\|^2 \right) \|C(1)\|^2 = O(T^{-1}).
\end{aligned}$$

We now consider (29). Note that

$$\begin{aligned}
& \tilde{C}_{T-p} \varepsilon_p \varepsilon'_q \tilde{C}'_{T-q} \otimes C(1) \varepsilon_t \varepsilon'_s C'(1) \\
& = \left( \tilde{C}_{T-p} \otimes C(1) \right) \left( \varepsilon_p \varepsilon'_q \otimes \varepsilon_t \varepsilon'_s \right) \left( \tilde{C}'_{T-q} \otimes C'(1) \right),
\end{aligned}$$

Some tedious calculations show that  $E \left( \varepsilon_p \varepsilon'_q \otimes \varepsilon_t \varepsilon'_s \right)$  is

$$\begin{cases} \mathbb{I}_{n^2}, & \text{if } p = q \neq t = s \\ \operatorname{vec}(\mathbb{I}_n) \operatorname{vec}(\mathbb{I}_n)', & \text{if } p = t \neq q = s \\ \mathbb{K}_{nn}, & \text{if } p = s \neq q = t \\ \mathbb{I}_{n^2} + \operatorname{vec}(\mathbb{I}_n) \operatorname{vec}(\mathbb{I}_n)' + \mathbb{K}_{nn} + (\nu^4 - 3) \sum_{l=1}^n e_{ll} \otimes e_{ll}, & \text{if } p = s = q = t \end{cases}$$

where  $\nu^4 = E \varepsilon_{ii}^4$  and  $e_{ll}$  is the  $n \times n$  matrix where the  $(l, l)^{th}$  element is one and the other elements are zeros. Then, (29) is

$$\begin{aligned}
& \frac{1}{T^2} \sum_{t=1}^T \sum_{p=1}^T \left[ \bar{\phi}\left(\frac{t}{T}\right) \right]^2 \operatorname{tr} \left( \left( \tilde{C}_{T-p} \otimes C(1) \right) \left( \tilde{C}'_{T-p} \otimes C'(1) \right) \right) \\
& + \frac{1}{T^2} \sum_{p=1}^T \sum_{q=1}^T \bar{\phi}\left(\frac{p}{T}\right) \bar{\phi}\left(\frac{q}{T}\right) \operatorname{tr} \left( \left( \tilde{C}_{T-p} \otimes C(1) \right) \operatorname{vec}(\mathbb{I}_n) \operatorname{vec}(\mathbb{I}_n)' \left( \tilde{C}'_{T-q} \otimes C'(1) \right) \right) \\
& + \frac{1}{T^2} \sum_{p=1}^T \sum_{q=1}^T \bar{\phi}\left(\frac{p}{T}\right) \bar{\phi}\left(\frac{q}{T}\right) \operatorname{tr} \left( \left( \tilde{C}_{T-p} \otimes C(1) \right) \mathbb{K}_{nn} \left( \tilde{C}'_{T-q} \otimes C'(1) \right) \right) \\
& + (\nu^4 - 3) \frac{1}{T^2} \sum_{p=1}^T \left[ \bar{\phi}\left(\frac{p}{T}\right) \right]^2 \operatorname{tr} \left( \left( \tilde{C}_{T-p} \otimes C(1) \right) \left( \sum_{i=1}^n e_{ii} \otimes e_{ii} \right) \left( \tilde{C}'_{T-p} \otimes C'(1) \right) \right) \\
& : = I_1 + I_2 + I_3 + I_4, \text{ say.}
\end{aligned}$$

We now consider the above four terms one by one. For  $I_1$ , we have

$$\begin{aligned}
& \frac{1}{T^2} \sum_{t=1}^T \sum_{p=1}^T \left[ \bar{\phi}\left(\frac{t}{T}\right) \right]^2 Etr \left( \left( \tilde{C}_{T-p} \tilde{C}'_{T-p} \right) \otimes (C(1)C'(1)) \right) \\
&= \frac{1}{T^2} \sum_{t=1}^T \sum_{p=1}^T \left[ \bar{\phi}\left(\frac{t}{T}\right) \right]^2 \left( tr \left( \tilde{C}_{T-p} \tilde{C}'_{T-p} \right) tr \left( C(1)C'(1) \right) \right) \\
&\leq \frac{1}{T^2} \sum_{t=1}^T \left[ \bar{\phi}\left(\frac{t}{T}\right) \right]^2 \sum_{p=1}^{\infty} \left\| \tilde{C}_p \right\|^2 \|C(1)\|^2 = O(1/T),
\end{aligned}$$

For  $I_2$ , we have, using  $tr(AB) = \text{vec}(A)' \text{vec}(B)$  and  $tr(CD) \leq \|C\| \|D\|$  for square matrices  $C$  and  $D$ ,

$$\begin{aligned}
& \frac{1}{T^2} \sum_{p=1}^T \sum_{q=1}^T \bar{\phi}\left(\frac{p}{T}\right) \bar{\phi}\left(\frac{q}{T}\right) Etr \left( \text{vec}(\tilde{C}_{T-p} C'(1)) \text{vec}(\tilde{C}_{T-q} C'(1))' \right) \\
&= \frac{1}{T^2} \sum_{p=1}^T \sum_{q=1}^T \bar{\phi}\left(\frac{p}{T}\right) \bar{\phi}\left(\frac{q}{T}\right) tr \left( \tilde{C}'_{T-p} \tilde{C}_{T-q} C'(1) C(1) \right) \\
&\leq \frac{1}{T^2} \sum_{p=1}^T \sum_{q=1}^T \left| \bar{\phi}\left(\frac{p}{T}\right) \bar{\phi}\left(\frac{q}{T}\right) \right| \left\| \tilde{C}'_{T-p} \tilde{C}_{T-q} \right\| \|C(1)\|^2 \\
&\leq \frac{1}{T^2} \left( \sum_{p=1}^T \left| \bar{\phi}\left(\frac{p}{T}\right) \right| \left\| \tilde{C}_{T-p} \right\|^2 \right) \|C(1)\|^2 = O(1/T^2).
\end{aligned}$$

For  $I_3$ , we have, using  $tr((C \otimes D) \mathbb{K}_{nn}) \leq \|C\| \|D\|$  for square matrices  $C$  and  $D$ ,

$$\begin{aligned}
I_3 &= \frac{1}{T^2} \sum_{p=1}^T \sum_{q=1}^T \bar{\phi}\left(\frac{p}{T}\right) \bar{\phi}\left(\frac{q}{T}\right) tr \left( \left( \tilde{C}_{T-p} \otimes C(1) \right) \left( \tilde{C}'_{T-q} \otimes C'(1) \right) \mathbb{K}_{nn} \right) \\
&= \frac{1}{T^2} \sum_{p=1}^T \sum_{q=1}^T \bar{\phi}\left(\frac{p}{T}\right) \bar{\phi}\left(\frac{q}{T}\right) tr \left( \left( \tilde{C}_{T-p} \tilde{C}'_{T-q} \right) \otimes (C(1)C'(1)) \mathbb{K}_{nn} \right) \\
&\leq \frac{1}{T^2} \sum_{p=1}^T \sum_{q=1}^T \left| \bar{\phi}\left(\frac{p}{T}\right) \bar{\phi}\left(\frac{q}{T}\right) \right| \left( \left\| \tilde{C}_{T-p} \tilde{C}'_{T-q} \right\| \|C(1)\|^2 \right) \\
&= O(1/T^2).
\end{aligned}$$

Finally,  $I_4 = 0$  if  $\nu^4 = 3$ . Otherwise,  $I_4$ , divided by  $(\nu^4 - 3)$ , is

$$\begin{aligned}
& \frac{1}{T^2} \sum_{p=1}^T \sum_{i=1}^n \left[ \bar{\phi}\left(\frac{p}{T}\right) \right]^2 \text{tr} \left( \left( \tilde{C}_{T-p} \otimes C(1) \right) (e_{ii} \otimes e_{ii}) \left( \tilde{C}'_{T-p} \otimes C'(1) \right) \right) \\
&= \frac{1}{T^2} \sum_{p=1}^T \sum_{i=1}^n \left[ \bar{\phi}\left(\frac{p}{T}\right) \right]^2 \text{tr} \left( \tilde{C}_{T-p} e_{ii} \tilde{C}'_{T-p} \right) \otimes (C(1) e_{ii} C'(1)) \\
&= \frac{1}{T^2} \sum_{p=1}^T \sum_{i=1}^n \left[ \bar{\phi}\left(\frac{p}{T}\right) \right]^2 \left\| \tilde{C}_{T-p} e_{ii} \right\|^2 \|C(1) e_{ii}\|^2 \\
&\leq \frac{n}{T^2} \sum_{p=1}^T \left[ \bar{\phi}\left(\frac{p}{T}\right) \right]^2 \left\| \tilde{C}_{T-p} \right\|^2 \|C(1)\|^2 = O(1/T^2),
\end{aligned}$$

where we have used  $\|e_{ii}\| = 1$ .

Combining the above results yields  $E \text{tr} \left( \text{vec} \left( R_2^{(2)} \right) \text{vec} \left( R_2^{(2)} \right)' \right) = O(1/T)$ .

Let

$$\Delta(t, \tau) = \frac{1}{K} \sum_{k=1}^K \phi_k \left( \frac{t}{T} \right) \left[ \phi_k \left( \frac{\tau+1}{T} \right) - \phi_k \left( \frac{\tau}{T} \right) \right],$$

then  $E \text{tr} \left( \text{vec} \left( R_2^{(3)} \right) \text{vec} \left( R_2^{(3)} \right)' \right)$  can be written as

$$\begin{aligned}
& E \text{tr} \frac{1}{T^2} \sum_{t=1}^T \sum_{\tau=1}^{T-1} \sum_{p=1}^T \sum_{q=1}^{T-1} \Delta(t, \tau) \Delta(p, q) \text{vec} \left( C(1) \varepsilon_t \tilde{u}'_{\tau} \right) \text{vec} \left( C'(1) \varepsilon_p \tilde{u}'_q \right)' \\
&= E \text{tr} \frac{1}{T^2} \sum_{t=1}^T \sum_{\tau=1}^{T-1} \sum_{p=1}^T \sum_{q=1}^{T-1} \Delta(t, \tau) \Delta(p, q) \left( \tilde{u}_{\tau} \tilde{u}'_q \otimes C(1) \varepsilon_t \varepsilon'_p C'(1) \right) \\
&= E \text{tr} \frac{1}{T^2} \sum_{t=1}^T \sum_{\tau=1}^{T-1} \sum_{p=1}^T \sum_{q=1}^{T-1} \Delta(t, \tau) \Delta(p, q) \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \tilde{C}_k \varepsilon_{\tau-k} \varepsilon'_{q-j} \tilde{C}'_j \otimes C(1) \varepsilon_t \varepsilon'_p C'(1) \\
&= E \text{tr} \frac{1}{T^2} \sum_{t=1}^T \sum_{\tau=1}^{T-1} \sum_{p=1}^T \sum_{q=1}^{T-1} \Delta(t, \tau) \Delta(p, q) \sum_{k=0}^{\tau-1} \sum_{j=0}^{q-1} \tilde{C}_k \varepsilon_{\tau-k} \varepsilon'_{q-j} \tilde{C}'_j \otimes C(1) \varepsilon_t \varepsilon'_p C'(1) \\
&\quad + E \text{tr} \frac{1}{T^2} \sum_{t=1}^T \sum_{\tau=1}^{T-1} \sum_{p=1}^T \sum_{q=1}^{T-1} \Delta(t, \tau) \Delta(p, q) \sum_{k=\tau}^{\infty} \sum_{j=q}^{\infty} \tilde{C}_k \varepsilon_{\tau-k} \varepsilon'_{q-j} \tilde{C}'_j \otimes C(1) \varepsilon_t \varepsilon'_p C'(1) \\
&= E \text{tr} \frac{1}{T^2} \sum_{t=1}^T \sum_{\tau=1}^{T-1} \sum_{p=1}^T \sum_{q=1}^{T-1} \Delta(t, \tau) \Delta(p, q) \sum_{k=1}^{\tau} \sum_{j=1}^q \tilde{C}_{\tau-k} \varepsilon_k \varepsilon'_j \tilde{C}'_{q-j} \otimes C(1) \varepsilon_t \varepsilon'_p C'(1)
\end{aligned} \tag{31}$$

$$\begin{aligned}
& + E \text{tr} \frac{1}{T^2} \sum_{t=1}^T \sum_{\tau=1}^{T-1} \sum_{p=1}^T \sum_{q=1}^{T-1} \Delta(t, \tau) \Delta(p, q) \sum_{k=\tau}^{\infty} \sum_{j=q}^{\infty} \tilde{C}_k \varepsilon_{\tau-k} \varepsilon'_{q-j} \tilde{C}'_j \otimes C(1) \varepsilon_t \varepsilon'_p C'(1)
\end{aligned} \tag{32}$$

It is easy to see that (32) is bounded by

$$\begin{aligned}
& Etr \frac{1}{T^2} \sum_{t=1}^T \sum_{\tau=1}^{T-1} \sum_{q=1}^{T-1} |\Delta(t, \tau) \Delta(t, q)| \sum_{k=0}^{\infty} \left( \tilde{C}_{\tau+k} \tilde{C}'_{q+k} \right) \otimes C(1) C'(1) \\
& \leq \frac{1}{T^2} \sum_{t=1}^T \sum_{\tau=1}^{T-1} \sum_{q=1}^{T-1} |\Delta(t, \tau) \Delta(t, q)| \left( \sum_{k=0}^{\infty} \left\| \tilde{C}_{\tau+k} \right\|^2 \right) \|C(1)\|^2 \\
& \leq \frac{1}{T^2} \sum_{t=1}^T O(1) = O\left(\frac{1}{T}\right),
\end{aligned}$$

using

$$\sum_{q=1}^{T-1} \Delta(t, q) = \left( \frac{1}{K} \sum_{k=1}^K \left| \phi_k \left( \frac{t}{T} \right) \right| \right) \left( \sum_{q=1}^{T-1} \left| \phi_k \left( \frac{\tau+1}{T} \right) - \phi_k \left( \frac{\tau}{T} \right) \right| \right) = O(1).$$

With similar calculations, we can show that (31) is  $O(1/T)$ . We have therefore proved  $Etr(\text{vec}(R_2) \text{vec}(R_2)') = O(1/T)$ . Similarly, we can prove  $Etr(\text{vec}(R_3) \text{vec}(R_3)') = O(1/T)$ . Details are omitted.

Combining the above calculations, we have

$$\text{var}(\text{vec}(\tilde{\Omega})) = \text{var}(\text{vec}(R_1)) + O\left(\frac{1}{T}\right).$$

But

$$\begin{aligned}
\text{var}(\text{vec}(R_1)) &= \frac{1}{T^2} \sum_{t=1}^T \sum_{\tau=1}^T \sum_{p=1}^T \sum_{q=1}^T \left[ \frac{1}{K} \sum_{k=1}^K \phi_k \left( \frac{t}{T} \right) \phi_k \left( \frac{\tau}{T} \right) \right] \left[ \frac{1}{K} \sum_{k=1}^K \phi_k \left( \frac{p}{T} \right) \phi_k \left( \frac{q}{T} \right) \right] \\
&\quad \times E \text{vec} [C(1) \varepsilon_t \varepsilon_\tau' C(1)'] \text{vec} [C(1) \varepsilon_p \varepsilon_q' C(1)']' \\
&= \frac{1}{T^2} \sum_{t=1}^T \sum_{\tau=1}^T \sum_{p=1}^T \sum_{q=1}^T \left[ \frac{1}{K} \sum_{k=1}^K \phi_k \left( \frac{t}{T} \right) \phi_k \left( \frac{\tau}{T} \right) \right] \left[ \frac{1}{K} \sum_{k=1}^K \phi_k \left( \frac{p}{T} \right) \phi_k \left( \frac{q}{T} \right) \right] \\
&\quad \times E [C(1) \varepsilon_\tau \otimes C(1) \varepsilon_t] [\varepsilon_q' C(1) \otimes \varepsilon_p' C(1)] \\
&= (\Omega \otimes \Omega) (\mathbb{I}_{n^2} + \mathbb{K}_{nm}) \frac{1}{T^2} \sum_{t=1}^T \sum_{\tau=1}^T \left[ \frac{1}{K} \sum_{k=1}^K \phi_k \left( \frac{t}{T} \right) \phi_k \left( \frac{\tau}{T} \right) \right]^2 + O\left(\frac{1}{T}\right),
\end{aligned}$$

and

$$\begin{aligned}
& \frac{1}{T^2} \sum_{t=1}^T \sum_{\tau=1}^T \left[ \frac{1}{K} \sum_{k=1}^K \phi_k \left( \frac{t}{T} \right) \phi_k \left( \frac{\tau}{T} \right) \right]^2 \\
&= \frac{1}{T^2} \sum_{t=1}^T \sum_{\tau=1}^T \frac{1}{K} \sum_{k=1}^K \phi_k \left( \frac{t}{T} \right) \phi_k \left( \frac{\tau}{T} \right) \frac{1}{K} \sum_{\ell=1}^K \phi_\ell \left( \frac{t}{T} \right) \phi_\ell \left( \frac{\tau}{T} \right) \\
&= \frac{1}{K^2} \sum_{k=1}^K \left[ \frac{1}{T} \sum_{t=1}^T \phi_k \left( \frac{t}{T} \right) \phi_\ell \left( \frac{t}{T} \right) \right]^2 \\
&= \frac{1}{K^2} \sum_{k=1}^K \left( \frac{1}{T} \sum_{t=1}^T \left[ \phi_k \left( \frac{t}{T} \right) \right]^2 \right)^2 = \frac{1}{K} \left( 1 + O\left(\frac{1}{T}\right) \right).
\end{aligned}$$

Therefore

$$\begin{aligned} \text{var} \left( \text{vec} \left( \tilde{\Omega} \right) \right) &= \text{var} \left( \text{vec} \left( R_1 \right) \right) + O \left( \frac{1}{T} \right) \\ &= \frac{1}{K} \left( \Omega \otimes \Omega \right) \left( \mathbb{I}_{n^2} + \mathbb{K}_{nn} \right) + O \left( \frac{1}{T} \right) \end{aligned}$$

as stated. ■

**Proof of Lemma 3. Part (a).** We write the statistic  $F_{T,GLS}$  as

$$\begin{aligned} F_{T,GLS} &= \left[ RT^{3/2}(\hat{\beta}_{GLS} - \beta) \right]' \left( 12R\Omega_{T,GLS}R' \right)^{-1/2} \left( R12\Omega_{T,GLS}R' \right)^{1/2} \left( R12\hat{\Omega}R' \right)^{-1} \\ &\quad \times \left( 12R\Omega_{T,GLS}R' \right)^{1/2} \left( R12\Omega_{T,GLS}R' \right)^{-1/2} \left[ RT^{3/2}(\hat{\beta}_{GLS} - \beta) \right] \\ &= \left\| \left( R12\Omega_{T,GLS}R' \right)^{-1/2} \left[ RT^{3/2}(\hat{\beta}_{GLS} - \beta) \right] \right\|^2 \\ &\quad \times e'_\beta \left( R12\Omega_{T,GLS}R' \right)^{1/2} \left( R12\hat{\Omega}R' \right)^{-1} \left( 12R\Omega_{T,GLS}R' \right)^{1/2} e_\beta \\ &: = \Theta \Xi + O_p \left( \frac{1}{T} \right), \end{aligned}$$

where

$$\Theta = \left\| \left( R12\Omega_{T,GLS}R' \right)^{-1/2} \left[ RT^{3/2}(\hat{\beta}_{GLS} - \beta) \right] \right\|^2,$$

and

$$\Xi = \left[ e'_\beta \left( R\Omega R' \right)^{1/2} \left( R\hat{\Omega}R' \right)^{-1} \left( R\Omega R' \right)^{1/2} e_\beta \right].$$

Here we have used

$$\Omega_{T,GLS} = \frac{1}{12} \text{var} \left[ T^{3/2} \left( \hat{\beta}_{GLS} - \beta \right) \right] = \Omega \left( 1 + O \left( \frac{1}{T} \right) \right).$$

Note that  $\Theta$  is independent of  $\Xi$  because (i)  $(\hat{\beta}_{GLS} - \beta)$  is independent of  $\hat{\Omega}$ . (ii)  $\Theta$  is the squared length of a standard normal vector and  $e_\beta$  is the direction of this vector. The length is independent of the direction. Hence

$$\begin{aligned} P \left[ \frac{(K-p+1)}{pK} F_{T,GLS} < z \right] &= P \left\{ \frac{(K-p+1)}{pK} [\Theta \Xi] < z \right\} + O \left( \frac{1}{T} \right) \\ &= EG_p \left( pz \frac{K}{K-p+1} \Xi^{-1} \right) + O \left( \frac{1}{T} \right) \end{aligned}$$

as stated.

**Part (b).** Let

$$\begin{aligned} \zeta_{1T} &= 2RT^{3/2}(\Delta_\beta)' \left( R12\hat{\Omega}R' \right)^{-1} \left( R12\Omega_{T,GLS}R' \right)^{1/2} e_\beta \\ \zeta_{2T} &= RT^{3/2}(\Delta_\beta)' \left( R12\hat{\Omega}R' \right)^{-1} RT^{3/2}\Delta_\beta \end{aligned}$$

and  $\zeta_T = \sqrt{\Theta}\zeta_{1T} + \zeta_{2T}$ . Then  $F_{T,OLS} = F_{T,GLS} + \zeta_T$ . Note that  $\Theta$  is independent of  $\zeta_{1T}, \zeta_{2T}$  and  $\Xi$ , we have

$$\begin{aligned}
& P \left[ \frac{(K-p+1)}{pK} F_{T,OLS} < z \right] = P \left[ \frac{(K-p+1)}{pK} (F_{T,GLS} + \zeta_T) < z \right] \\
& = P \left\{ \frac{(K-p+1)}{pK} \left[ \Theta \Xi + \sqrt{\Theta} \zeta_{1T} + \zeta_{2T} + O_p \left( \frac{1}{T} \right) \right] < z \right\} \\
& = P \left\{ \frac{(K-p+1)}{pK} \left[ \Theta \Xi + \sqrt{\Theta} \zeta_{1T} + \zeta_{2T} \right] < z \right\} + O \left( \frac{1}{T} \right) \\
& : = EF(\zeta_{1T}, \zeta_{2T}, \Xi) + O \left( \frac{1}{T} \right),
\end{aligned}$$

where

$$F(a, b, c) = P \left\{ \frac{(K-p+1)}{pK} \left[ \Theta c + \sqrt{\Theta} a + b \right] < z \right\}.$$

But

$$\begin{aligned}
EF(\zeta_{1T}, \zeta_{2T}, \Xi) & = EF(0, 0, \Xi) + EF'_1(0, 0, \Xi) \zeta_{1T} + O(E\zeta_{1T}^2) + O(E|\zeta_{1T}\zeta_{2T}|) + O(E\zeta_{2T}) \\
& = EF(0, 0, \Xi) + O(E\zeta_{1T}^2) + O(E|\zeta_{1T}\zeta_{2T}|) + O(E\zeta_{2T}) \\
& = EF(0, 0, \Xi) + O \left( \frac{1}{T} \right).
\end{aligned}$$

Here we have used (i)

$$EF'_1(0, 0, \Xi) \zeta_{1T} = E(E[F'_1(0, 0, \Xi) \zeta_{1T} | e_\beta]) = 0$$

because conditional on  $e_\beta$ ,  $\Xi$  is an even function of  $u$  and  $\zeta_{1T}$  is an odd function of  $u$ ; (ii)  $O(E\zeta_{1T}^2) = O(1/T)$  and  $O(E\zeta_{2T}) = O(1/T)$  which follows from  $var(c'\Delta_\beta\Delta'_\beta c) = O(1/T)$  for any constant  $c$ .

We have therefore shown that

$$\begin{aligned}
P \left[ \frac{(K-p+1)}{pK} F_{T,OLS} < z \right] & = EF(0, 0, \Xi) + O \left( \frac{1}{T} \right) \\
& = P \left\{ \frac{(K-p+1)}{pK} [\Theta \Xi] < z \right\} + O \left( \frac{1}{T} \right) \\
& = P \left[ \frac{(K-p+1)}{pK} F_{G,OLS} < z \right] + O \left( \frac{1}{T} \right)
\end{aligned}$$

as desired. ■

**Proof of Theorem 4.** We write  $\Xi = \Xi(\hat{\Omega})$  and proceed to take a Taylor expansion of  $\Xi(\hat{\Omega})$  around  $\Xi(\Omega) = 1$ . To this end, we first compute the derivatives of  $\Xi(\hat{\Omega})$  with respect

to  $\hat{\Omega}$  :

$$\begin{aligned}
& d\Xi^{-1}(\hat{\Omega}) \\
&= -\Xi^{-2}d\Xi(\hat{\Omega}) \\
&= -\Xi^{-2}e'_1H(R\Omega R')^{1/2}\left[d(R\hat{\Omega}R')^{-1}\right](R\Omega R')^{1/2}H'e_1 \\
&= \Xi^{-2}e'_1H(R\Omega R')^{1/2}\underbrace{(R\hat{\Omega}R')^{-1}}_{1\times n}R(d\hat{\Omega})\underbrace{R'(R\hat{\Omega}R')^{-1}}_{n\times 1}(R\Omega R')^{1/2}H'e_1 \\
&= \Xi^{-2}\left(\left[e'_1H(R\Omega R')^{1/2}(R\hat{\Omega}R')^{-1}R\right]\otimes\left[e'_1H(R\Omega R')^{1/2}(R\hat{\Omega}R')^{-1}R\right]\right)dvec(\hat{\Omega}).
\end{aligned}$$

Hence

$$\frac{\partial\Xi^{-1}(\hat{\Omega})}{\partial\left[vec(\hat{\Omega})\right]'} = \Xi^{-2}\times\left(\left[e'_1H(R\Omega R')^{1/2}(R\hat{\Omega}R')^{-1}R\right]\otimes\left[e'_1H(R\Omega R')^{1/2}(R\hat{\Omega}R')^{-1}R\right]\right).$$

Evaluating the above derivative at  $\Omega$  yields:

$$\frac{\partial\Xi^{-1}(\Omega)}{\partial\left[vec(\hat{\Omega})\right]'} = \left(\left[e'_1H(R\Omega R')^{-1/2}R\right]\otimes\left[e'_1H(R\Omega R')^{-1/2}R\right]\right).$$

Next, we compute the second order derivative:

$$d\frac{\partial\Xi^{-1}(\hat{\Omega})}{\partial\left[vec(\hat{\Omega})\right]'} = D_1 + D_2 + D_3,$$

where

$$\begin{aligned}
D_1 &= -2\Xi^{-3}d\Xi\left(\left[e'_\beta(R\Omega R')^{1/2}(R\hat{\Omega}R')^{-1}R\right]\otimes\left[e'_\beta(R\Omega R')^{1/2}(R\hat{\Omega}R')^{-1}R\right]\right), \\
D_2 &= \Xi^{-2}\left[e'_\beta(R\Omega R')^{1/2}d(R\hat{\Omega}R')^{-1}R\right]\otimes\left[e'_\beta(R\Omega R')^{1/2}(R\hat{\Omega}R')^{-1}R\right], \\
D_3 &= \Xi^{-2}\left[e'_\beta(R\Omega R')^{1/2}(R\hat{\Omega}R')^{-1}R\right]\otimes\left[e'_\beta(R\Omega R')^{1/2}d(R\hat{\Omega}R')^{-1}R\right] = D_2\mathbb{K}_{nn}.
\end{aligned}$$

Furthermore

$$\begin{aligned}
& D_1 \\
&= 2\Xi^{-3}dvec(\hat{\Omega})'\left(\left[e'_\beta(R\Omega R')^{1/2}(R\hat{\Omega}R')^{-1}R\right]\otimes\left[e'_\beta(R\Omega R')^{1/2}(R\hat{\Omega}R')^{-1}R\right]\right)' \\
&\quad\times\left(\left[e'_\beta(R\Omega R')^{1/2}(R\hat{\Omega}R')^{-1}R\right]\otimes\left[e'_\beta(R\Omega R')^{1/2}(R\hat{\Omega}R')^{-1}R\right]\right),
\end{aligned}$$

$$\begin{aligned}
& D_2 \\
&= -\Xi^{-2} e'_\beta (R\Omega R')^{1/2} (R\hat{\Omega}R')^{-1} R (d\hat{\Omega}) R' (R\hat{\Omega}R')^{-1} R \otimes \left[ e'_\beta (R\Omega R')^{1/2} (R\hat{\Omega}R')^{-1} R \right] \\
&= -\Xi^{-2} \left[ R' (R\hat{\Omega}R')^{-1} R (d\hat{\Omega}) \left[ e'_\beta (R\Omega R')^{1/2} (R\hat{\Omega}R')^{-1} R \right]' \right]' \\
&\quad \otimes \left[ e'_\beta (R\Omega R')^{1/2} (R\hat{\Omega}R')^{-1} R \right] \\
&= -\Xi^{-2} \left[ \left( e'_\beta (R\Omega R')^{1/2} (R\hat{\Omega}R')^{-1} R \otimes R' (R\hat{\Omega}R')^{-1} R \right) \text{vec} (d\hat{\Omega}) \right]' \\
&\quad \otimes \left[ e'_\beta (R\Omega R')^{1/2} (R\hat{\Omega}R')^{-1} R \right] \\
&= -\text{dvec} (\hat{\Omega})' \Xi^{-2} \left[ e'_\beta (R\Omega R')^{1/2} (R\hat{\Omega}R')^{-1} R \right]' \otimes R' (R\hat{\Omega}R')^{-1} R \\
&\quad \otimes \left[ e'_\beta (R\Omega R')^{1/2} (R\hat{\Omega}R')^{-1} R \right].
\end{aligned}$$

Therefore

$$\begin{aligned}
& \frac{\partial^2 \Xi (\hat{\Omega})}{\partial \text{vec} (\hat{\Omega}) \partial \text{vec} (\hat{\Omega})'} \\
&= 2\Xi^{-3} R' (R\Omega R')^{-1/2} (H' e_1 e_1' H) (R\Omega R')^{-1/2} R \otimes R' (R\Omega R')^{-1/2} (H' e_1 e_1' H) (R\Omega R')^{-1/2} R \\
&\quad - \Xi^{-2} \left[ R' (R\Omega R')^{-1/2} H' e_1 \right] \otimes R' (R\Omega R')^{-1} R \otimes \left[ e_1' H (R\Omega R')^{-1/2} R \right] \\
&\quad - \Xi^{-2} \left[ R' (R\Omega R')^{-1/2} H' e_1 \right] \otimes R' (R\Omega R')^{-1} R \otimes \left[ e_1' H (R\Omega R')^{-1/2} R \right] \\
&= 2\Xi^{-3} R' (R\Omega R')^{-1/2} (H' e_1 e_1' H) (R\Omega R')^{-1/2} R \otimes R' (R\Omega R')^{-1/2} (H' e_1 e_1' H) (R\Omega R')^{-1/2} R \\
&\quad - \Xi^{-2} \left[ R' (R\Omega R')^{-1/2} H' e_1 \right] \otimes R' (R\Omega R')^{-1} R \otimes \left[ e_1' H (R\Omega R')^{-1/2} R \right] (\mathbb{I}_{n^2} + \mathbb{K}_{nn}).
\end{aligned}$$

Evaluating the above derivative at  $\Omega$  yields:

$$\begin{aligned}
& \frac{\partial^2 \Xi (\Omega)}{\partial \text{vec} (\hat{\Omega}) \partial \text{vec} (\hat{\Omega})'} \\
&= 2R' (R\Omega R')^{-1/2} (e_\beta e'_\beta) (R\Omega R')^{-1/2} R \otimes R' (R\Omega R')^{-1/2} (e_\beta e'_\beta) (R\Omega R')^{-1/2} R \\
&\quad - \left[ R' (R\Omega R')^{-1/2} e_\beta \right] \otimes R' (R\Omega R')^{-1} R \otimes \left[ e'_\beta (R\Omega R')^{-1/2} R \right] (\mathbb{I}_{n^2} + \mathbb{K}_{nn}) \\
&= 2R' (R\Omega R')^{-1/2} (e_\beta e'_\beta) (R\Omega R')^{-1/2} R \otimes R' (R\Omega R')^{-1/2} (e_\beta e'_\beta) (R\Omega R')^{-1/2} R \\
&\quad - \left[ R' (R\Omega R')^{-1/2} e_\beta e'_\beta (R\Omega R')^{-1/2} R \otimes R' (R\Omega R')^{-1} R \right] \mathbb{K}_{nn} (\mathbb{I}_{n^2} + \mathbb{K}_{nn}) \\
&: = J_1 + J_2
\end{aligned}$$

Now

$$\left[ \Xi (\hat{\Omega}) \right]^{-1} = 1 + L + Q + \text{remainder} \tag{33}$$

where

$$L = \frac{\partial \Xi^{-1}(\Omega)}{\partial \left[ \text{vec} \left( \hat{\Omega} \right) \right]'} \text{vec} \left( \hat{\Omega} - \Omega \right), Q = \frac{1}{2} \text{vec} \left( \hat{\Omega} - \Omega \right)' (J_1 + J_2) \text{vec} \left( \hat{\Omega} - \Omega \right).$$

We proceed to compute the expected values of  $L$  and  $Q$ . As a by-product, we obtain the order of the remainder term.

For notational simplicity, we let  $X = (X_1, \dots, X_p)' = e_\beta \in \mathbb{R}^p$ . It is easy to see that  $X$  is a random vector uniformly distributed on the surface of the  $p$ -dimensional sphere with center 0 and radius 1. It follows from Khokhlov (2006) that the density of  $X_1$ , the first element of  $X$ , is

$$f_{X_1}(x) = \frac{\Gamma(p/2)}{\sqrt{\pi} \Gamma((p-1)/2)} (1-x^2)^{\frac{p-3}{2}}, x \in [-1, 1],$$

where  $\Gamma(\cdot)$  is the gamma function. Therefore

$$EX_1^2 = \int_{-1}^1 x^2 f_{X_1}(x) dx = \frac{1}{2} \frac{\Gamma(\frac{1}{2}p)}{\Gamma(\frac{1}{2}p+1)} = \frac{1}{p}, \quad (34)$$

$$EX_1^4 = \int_{-1}^1 x^4 f_{X_1}(x) dx = \frac{3}{4} \frac{\Gamma(\frac{1}{2}p)}{\Gamma(\frac{1}{2}p+2)} = \frac{3}{p(p+2)}. \quad (35)$$

By definition,  $E(\sum_{i=1}^p X_i^2)^2 = 1$ . Using the permutational symmetry of the distribution of  $X$ , we have

$$pEX_1^4 + p(p-1)EX_1^2X_2^2 = 1,$$

which implies that

$$EX_1^2X_2^2 = \frac{1}{p(p+2)}. \quad (36)$$

Using (34) and  $EX_1X_2 = 0$ , we have

$$\begin{aligned} EL &= E \left( \left[ e'_\beta (R\Omega R')^{-1/2} R \right] \otimes \left[ e'_\beta (R\Omega R')^{-1/2} R \right] \right) \text{vec} \left( \hat{\Omega} - \Omega \right) \\ &= E e'_\beta (R\Omega R')^{-1/2} R \left( \hat{\Omega} - \Omega \right) R' (R\Omega R')^{-1/2} e_\beta \\ &= \frac{K^2}{T^2} E e'_\beta (R\Omega R')^{-1/2} (RBR') (R\Omega R')^{-1/2} e_\beta (1 + o(1)) \\ &= \frac{K^2}{T^2} E \text{tr} \left( (R\Omega R')^{-1/2} (RBR') (R\Omega R')^{-1/2} e_\beta e'_\beta \right) (1 + o(1)) \\ &= \frac{K^2}{T^2} \text{tr} \left[ (R\Omega R')^{-1/2} (RBR') (R\Omega R')^{-1/2} \right] \frac{1}{p}. \end{aligned}$$

To compute  $E(Q)$ , we note that  $Q$  consists of two terms. The first term is

$$\begin{aligned}
& \frac{1}{2} E \text{vec}(\hat{\Omega} - \Omega)' J_1 \text{vec}(\hat{\Omega} - \Omega) \\
&= E \text{vec}(\hat{\Omega} - \Omega)' R' (R\Omega R')^{-1/2} (e_\beta e'_\beta) (R\Omega R')^{-1/2} R \\
&\quad \otimes R' (R\Omega R')^{-1/2} (e_\beta e'_\beta) (R\Omega R')^{-1/2} R \text{vec}(\hat{\Omega} - \Omega) \\
&= E \left[ e'_\beta (R\Omega R')^{-1/2} R (\hat{\Omega} - \Omega) R' (R\Omega R')^{-1/2} e_\beta \right]^2 \\
&= E \left( \sum_{i,j} \sum_{m,\ell} A_{ij} A_{\ell m} X_i X_j X_\ell X_m | A \right),
\end{aligned}$$

where  $A = (A_{ij}) = (R\Omega R')^{-1/2} R(\hat{\Omega} - \Omega) R' (R\Omega R')^{-1/2}$ . Plugging in (35) and (36), we have

$$\begin{aligned}
& \frac{1}{2} E \text{vec}(\hat{\Omega} - \Omega)' J_1 \text{vec}(\hat{\Omega} - \Omega) \\
&= \sum_i A_{ii}^2 X_i^4 + E \sum_{i \neq m} A_{ii} A_{mm} E X_i^2 X_m^2 + 2E \sum_{i \neq j} A_{ij} A_{ij} E X_i^2 X_j^2 \\
&= E \sum_i A_{ii}^2 \frac{3}{p(p+2)} + E \sum_{i \neq m} A_{ii} A_{mm} \frac{1}{p(p+2)} + 2E \sum_{i \neq j} A_{ij}^2 \frac{1}{p(p+2)} \\
&= \frac{3}{p(p+2)} E \sum_i A_{ii}^2 + \frac{1}{p(p+2)} E \sum_{i \neq m} (A_{ii} A_{mm} + 2A_{im}^2) \\
&= \frac{1}{p(p+2)} E \sum_{i,m} (A_{ii} A_{mm} + 2A_{im}^2) = \frac{1}{p(p+2)} E \left( 2\text{tr}(AA) + [\text{tr}(A)]^2 \right).
\end{aligned}$$

Now

$$\begin{aligned}
E \text{tr}(AA) &= E \text{vec}(A)' \text{vec}(A) \\
&= E \left\{ \left[ \left( R' (R\Omega R')^{-1/2} \right)' \otimes R' (R\Omega R')^{-1/2} \right] \text{vec}(\hat{\Omega} - \Omega) \right\}' \\
&\quad \times \left[ \left( R' (R\Omega R')^{-1/2} \right)' \otimes R' (R\Omega R')^{-1/2} \right] \text{vec}(\hat{\Omega} - \Omega) \\
&= \frac{1}{K} \text{tr} \left\{ \left[ \left( R' (R\Omega R')^{-1/2} \right)' \otimes R' (R\Omega R')^{-1/2} \right] (\Omega \otimes \Omega) (\mathbb{I}_n + \mathbb{K}_{nn}) \right. \\
&\quad \left. \times \left[ R' (R\Omega R')^{-1/2} \otimes \left( R' (R\Omega R')^{-1/2} \right)' \right] \right\} + o\left(\frac{K^2}{T^2}\right) \\
&= \frac{1}{K} \text{tr} \left\{ \left( \left[ R' (R\Omega R')^{-1} R \right] \otimes \left[ R' (R\Omega R')^{-1} R \right] \right) (\Omega \otimes \Omega) (\mathbb{I}_{n^2} + \mathbb{K}_{nn}) \right\} + o\left(\frac{K^2}{T^2}\right) \\
&= \frac{1}{K} (p^2 + p) + o\left(\frac{K^2}{T^2}\right),
\end{aligned}$$

and

$$\begin{aligned}
E [\text{tr}(A)]^2 &= E \left[ \text{tr} \left( (R\Omega R')^{-1/2} R(\hat{\Omega} - \Omega)R' (R\Omega R')^{-1/2} \right) \right]^2 \\
&= E \left[ \text{tr} \left( (\hat{\Omega} - \Omega)R' (R\Omega R')^{-1} R \right) \right]^2 \\
&= E \left[ \text{vec} \left( R' (R\Omega R')^{-1} R \right) \right]' \left( \text{vec}(\hat{\Omega} - \Omega) \right)' \left( \text{vec}(\hat{\Omega} - \Omega) \right) \text{vec} \left( R' (R\Omega R')^{-1} R \right) \\
&= E \left[ \text{vec} \left( R' (R\Omega R')^{-1} R \right) \right]' (\Omega \otimes \Omega) \text{vec} \left( R' (R\Omega R')^{-1} R \right) + o \left( \frac{K^2}{T^2} \right) \\
&= \frac{2}{K} \left[ \text{vec} \left( R' (R\Omega R')^{-1} R \right) \right]' \text{vec} \left( \Omega R' (R\Omega R')^{-1} R \Omega \right) + o \left( \frac{K^2}{T^2} \right) \\
&= \frac{2}{K} \text{tr} \left( R' (R\Omega R')^{-1} R \Omega R' (R\Omega R')^{-1} R \Omega \right) + o \left( \frac{K^2}{T^2} \right) \\
&= \frac{2p}{K} + o \left( \frac{K^2}{T^2} \right).
\end{aligned}$$

Therefore

$$\begin{aligned}
\frac{1}{2} E \text{vec}(\hat{\Omega} - \Omega)' J_1 \text{vec}(\hat{\Omega} - \Omega) &= \frac{1}{p(p+2)} \left( \frac{2}{K} (p^2 + p) + \frac{2p}{K} \right) + o \left( \frac{K^2}{T^2} \right) \\
&= \frac{2}{K} + o \left( \frac{K^2}{T^2} \right).
\end{aligned}$$

It is easy to show that

$$\frac{1}{2} E \text{vec}(\hat{\Omega} - \Omega)' J_2 \text{vec}(\hat{\Omega} - \Omega) = \frac{1}{K} \frac{1}{p} (p^2 + p) + o \left( \frac{K^2}{T^2} \right).$$

Hence

$$EQ = \frac{2}{K} - \frac{1}{K} \frac{1}{p} (p^2 + p) + o \left( \frac{K^2}{T^2} \right) = -\frac{1}{K} (p-1) + o \left( \frac{K^2}{T^2} \right),$$

and

$$\left[ \Xi(\hat{\Omega}) \right]^{-1} = 1 + L + Q + o_p \left( \frac{1}{K} \right). \quad (37)$$

Using the above expansion, we have

$$\begin{aligned}
&P \left( \frac{(K-p+1)}{pK} F_{T,GLS} < z \right) = P \left( \frac{(K-p+1)}{pK} \Theta < z \Xi^{-1} \right) \\
&= P \left( \frac{(K-p+1)}{pK} \Theta < z(1+L+Q) \right) + o \left( \frac{1}{K} \right) \\
&= EG_p \left( pz \frac{K}{K-p+1} (1+L+Q) \right) + o \left( \frac{1}{K} \right) \\
&= G_p \left( pz \frac{K}{K-p+1} \right) + G'_p \left( pz \frac{K}{K-p+1} \right) pz \frac{K}{K-p+1} E(L+Q) \\
&\quad + \frac{1}{2} EG''_p \left( pz \frac{K}{K-p+1} \right) p^2 z^2 \left( \frac{K}{K-p+1} \right)^2 E(L+Q)^2 + o \left( \frac{1}{K} \right) \\
&= G_p \left( pz \frac{K}{K-p+1} \right) + G'_p(pz) (pz) E(L+Q) + \frac{1}{2} EG''_p(pz) p^2 z^2 (EL^2) + o \left( \frac{1}{K} \right).
\end{aligned}$$

But

$$\begin{aligned}
EL^2 &= \frac{1}{K} E \left( \left[ e'_\beta (R\Omega R')^{-1/2} R \right] \otimes \left[ e'_\beta (R\Omega R')^{-1/2} R \right] (\Omega \otimes \Omega) (\mathbb{I}_{n^2} + \mathbb{K}_{nn}) \right) \\
&\quad \times \left[ e'_\beta (R\Omega R')^{-1/2} R \right]' \otimes \left[ e'_\beta (R\Omega R')^{-1/2} R \right]' + o\left(\frac{1}{K}\right) \\
&= \frac{2}{K} E [e'_\beta e_\beta]^2 = \frac{2}{K} \frac{1}{p(p+2)} (p^2 + 2p) + o\left(\frac{1}{K}\right) = \frac{2}{K} + o\left(\frac{1}{K}\right),
\end{aligned}$$

so

$$\begin{aligned}
&P \left( \frac{(K-p+1)}{pK} F_{T,GLS} < z \right) \\
&= G_p \left( pz \frac{K}{K-p+1} \right) + \frac{K^2}{T^2} G'_p(pz) \text{ztr} \left[ (RBR') (R\Omega R')^{-1} \right] \\
&\quad - \frac{1}{K} G'_p(pz) zp(p-1) + \frac{2}{2K} G''_p(pz) p^2 z^2 + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) \\
&= G_p \left( pz \frac{K}{K-p+1} \right) + \frac{K^2}{T^2} G'_p(pz) \text{ztr} \left\{ (RBR') (R\Omega R')^{-1} \right\} \\
&\quad + \frac{1}{K} [G''_p(pz) z^2 p^2 - G'_p(pz) z (p^2 - 1)] + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) \\
&= G_p(pz) + \frac{K^2}{T^2} G'_p(pz) \text{ztr} \left\{ (RBR') (R\Omega R')^{-1} \right\} + \frac{1}{K} G''_p(pz) z^2 p^2 + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right)
\end{aligned}$$

as desired. ■

**Proof of Theorem 5. Part (a).** It follows from Theorem 4 that

$$\begin{aligned}
&P (F_{T,OLS}^* > pF_{p,K-p+1}^\alpha) - \alpha \\
&= -\frac{K^2}{T^2} G'_p(pF_{p,K-p+1}^\alpha) pF_{p,K-p+1}^\alpha \bar{B} + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right). \quad (38)
\end{aligned}$$

But

$$pF_{p,K-p+1}^\alpha = \chi_p^\alpha + o(1),$$

hence

$$\begin{aligned}
&P (F_{T,OLS}^* > pF_{p,K-p+1}^\alpha) - \alpha \\
&= -\frac{K^2 \bar{B}}{T^2} G'_p(\chi_p^\alpha) \chi_p^\alpha + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right).
\end{aligned}$$

Part (ii). The  $F_{T,GLS}$  statistic can be written as

$$\begin{aligned}
F_{T,GLS} &= \left[ RT^{3/2}(\hat{\beta}_{GLS} - \beta) + (R12\Omega R')^{1/2} \tilde{c} \right]' \times (R12\hat{\Omega} R')^{-1} \\
&\quad \times \left[ RT^{3/2}(\hat{\beta}_{GLS} - \beta) + (R12\Omega R')^{1/2} \tilde{c} \right] \\
&= \left[ (R12\Omega_{T,GLS})^{-1/2} RT^{3/2}(\hat{\beta}_{GLS} - \beta) + \tilde{c} \right]' \\
&\quad \times (R12\Omega_{T,GLS})^{1/2} (R12\hat{\Omega} R')^{-1} (R12\Omega_{T,GLS})^{1/2} \\
&\quad \times \left[ (R12\Omega_{T,GLS})^{-1/2} RT^{3/2}(\hat{\beta}_{GLS} - \beta) + \tilde{c} \right] + O\left(\frac{1}{T}\right),
\end{aligned}$$

where by assumption  $\|\tilde{c}\|^2 = \delta^2$ . Let

$$e_{\beta\delta} = \frac{(R12\Omega_{T,GLS})^{-1/2} RT^{3/2}(\hat{\beta}_{GLS} - \beta) + \tilde{c}}{\left\| (R12\Omega_{T,GLS})^{-1/2} RT^{3/2}(\hat{\beta}_{GLS} - \beta) + \tilde{c} \right\|},$$

then

$$F_{T,GLS} = \Theta_\delta \Xi_\delta + O_p\left(\frac{1}{T}\right),$$

where

$$\begin{aligned} \Theta_\delta &= \left\| (R12\Omega_{T,GLS})^{-1/2} RT^{3/2}(\hat{\beta}_{GLS} - \beta) + \tilde{c} \right\|^2, \\ \Xi_\delta &= e'_{\beta\delta} (R12\Omega)^{1/2} \left( R12\hat{\Omega}R' \right)^{-1} (R12\Omega)^{1/2} e_{\beta\delta}, \end{aligned}$$

and  $\Theta_\delta$  is independent of  $\Xi_\delta$ . In addition,  $\Theta_\delta \sim \chi_p^2(\delta^2)$  and  $e_{\beta\delta}$  is uniformly distributed on the unit sphere  $S_p(1)$ . Using the same calculation as in the proof of Theorem 4, we have,

$$\begin{aligned} & P \left[ \frac{(K-p+1)}{pK} F_{T,GLS} < pF_{p,K-p+1}^\alpha \mid H_1(\delta^2) \right] \\ &= EG_{p,\delta^2} \left( pF_{p,K-p+1}^\alpha \frac{K}{K-p+1} \Xi_\delta \right) + O\left(\frac{1}{T}\right) \\ &= G_{p,\delta^2}(pF_{p,K-p+1}^\alpha) + G'_{p,\delta^2}(pF_{p,K-p+1}^\alpha) pF_{p,K-p+1}^\alpha E \left[ \frac{K}{K-p+1} \Xi_\delta - 1 \right] \\ &\quad + \frac{1}{2} G''_{p,\delta^2}(pF_{p,K-p+1}^\alpha) (pF_{p,K-p+1}^\alpha)^2 E \left[ \frac{K}{K-p+1} \Xi_\delta - 1 \right]^2 + O\left(\frac{1}{T}\right) + o\left(\frac{1}{K}\right) \\ &= G_{p,\delta^2}(pF_{p,K-p+1}^\alpha) + G'_{p,\delta^2}(pF_{p,K-p+1}^\alpha) pF_{p,K-p+1}^\alpha \frac{K^2}{T^2} \bar{B} \\ &\quad + \frac{1}{2} G''_{p,\delta^2}(pF_{p,K-p+1}^\alpha) (pF_{p,K-p+1}^\alpha)^2 \frac{2}{K} + o\left(\frac{K^2}{T^2}\right) + o\left(\frac{1}{K}\right) + O\left(\frac{1}{T}\right). \end{aligned}$$

Plugging

$$pF_{p,K-p+1}^\alpha = \chi_p^\alpha - \frac{1}{K} \frac{G''_p(\chi_p^\alpha)}{G'_p(\chi_p^\alpha)} (\chi_p^\alpha)^2 + o\left(\frac{1}{K}\right),$$

we have

$$\begin{aligned} & P \left[ \frac{(K-p+1)}{pK} F_{T,GLS} < pF_{p,K-p+1}^\alpha \mid H_1(\delta^2) \right] \\ &= G_{p,\delta^2}(\chi_p^\alpha) + \frac{K^2}{T^2} G'_{p,\delta^2}(\chi_p^\alpha) \chi_p^\alpha \bar{B} + \frac{1}{K} \mathcal{Q}_{p,\delta^2}(\chi_p^\alpha) (\chi_p^\alpha)^2 \\ &\quad + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right), \end{aligned}$$

where

$$\mathcal{Q}_{p,\delta^2}(\chi_p^\alpha) = G''_{p,\delta^2}(\chi_p^\alpha) - \frac{G''_p(\chi_p^\alpha)}{G'_p(\chi_p^\alpha)} G'_{p,\delta^2}(\chi_p^\alpha).$$

Now it is known that

$$G'_{p\delta^2}(z) = \sum_{j=0}^{\infty} \frac{(\delta^2/2)^j}{j!} e^{-\delta^2/2} \frac{1}{2^{j+p/2}\Gamma(j+p/2)} z^{j+p/2-1} e^{-z/2}, \quad (39)$$

and thus

$$\begin{aligned} G''_{p\delta^2}(z) &= \sum_{j=0}^{\infty} \frac{(\delta^2/2)^j}{j!} e^{-\delta^2/2} \frac{1}{2^{j+p/2}\Gamma(j+p/2)} z^{j+p/2-1} e^{-z/2} \left[ \frac{2j+p-z-2}{2z} \right] \\ &= -\frac{1}{2} G'_{p\delta}(z) \left( \frac{z+2-p}{z} \right) + \sum_{j=0}^{\infty} \frac{(\delta^2/2)^j}{j!} e^{-\delta^2/2} \frac{1}{2^{j+p/2}\Gamma(j+p/2)} \frac{j}{z} \\ &= -\frac{1}{2} G'_{p\delta}(z) \left( \frac{z+2-p}{z} \right) + \sum_{j=0}^{\infty} \frac{(\delta^2/2)^j}{j!} e^{-\delta^2/2} \frac{z^{j+p/2-1} e^{-z/2}}{2^{j+p/2}\Gamma(j+p/2)} \frac{j}{z} \\ &= G'_{p\delta^2}(z) \frac{G''_p(z)}{G'_p(z)} + \sum_{j=0}^{\infty} \frac{(\delta^2/2)^j}{j!} e^{-\delta^2/2} \frac{z^{j+p/2-1} e^{-z/2}}{2^{j+p/2}\Gamma(j+p/2)} \frac{j}{z}. \end{aligned}$$

Hence

$$\mathcal{Q}_{p,\delta^2}(z) = \sum_{j=0}^{\infty} \frac{(\delta^2/2)^j}{j!} e^{-\delta^2/2} \frac{z^{j+p/2-1} e^{-z/2}}{2^{j+p/2}\Gamma(j+p/2)} \frac{j}{z},$$

completing the proof of the theorem. ■

## References

- [1] Andrews, D. W. K. (1991): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59, 817–854.
- [2] Bilodeau, M. and D. Brenner (1999): *Theory of Multivariate Statistics*, Springer.
- [3] Grenander, U. and M. Rosenblatt (1957): *Statistical Analysis of Stationary Time Series*. New York: Wiley.
- [4] Hashimzade, N. and Vogelsang, T. J. (2007): “Fixed-b Asymptotic Approximation of the Sampling Behavior of Nonparametric Spectral Density Estimators,” *Journal of Time Series Analysis*, 29(1), 142–162.
- [5] Hannan, E. J. (1970): *Multiple Time Series*, New York, Wiley.
- [6] Hotelling H. (1931): “The Generalization of Student’s Ratio,” *Annals of Mathematical Statistics*, 2, 360–378.
- [7] Khokhlov, V.I. (2006): “The Uniform Distribution on a Sphere in  $\mathbb{R}^s$ : Properties of Projections,” *Theory of Probability and Applications*, 50(3), 386–399.
- [8] Kiefer, N. M., T. J. Vogelsang and H. Bunzel (2000): “Simple Robust Testing of Regression Hypotheses,” *Econometrica*, 68, 695–714.

- [9] Kiefer, N. M. and T. J. Vogelsang (2002a): “Heteroskedasticity-autocorrelation Robust Testing Using Bandwidth Equal to Sample Size,” *Econometric Theory*, 18, 1350–1366.
- [10] ——— (2002b): “Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation,” *Econometrica*, 70, 2093–2095.
- [11] ——— (2005): “A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests,” *Econometric Theory*, 21, 1130-1164.
- [12] Müller, U.K. (2007): “A Theory of Robust Long-Run Variance Estimation,” *Journal of Econometrics*, 141(2), 1331-1352
- [13] Newey, W. K. and K. D. West (1987): “A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703–708.
- [14] ——— (1994): “Automatic Lag Selection in Covariance Estimation,” *Review of Economic Studies*, 61, 631–654.
- [15] Phillips, P. C. B. (1980): “Finite Sample Theory and the Distributions of Alternative Estimators of the Marginal Propensity to Consume,” *Review of Economic Studies*, 47, 183–224.
- [16] Phillips, P. C. B. (2005): “HAC Estimation by Automated Regression,” *Econometric Theory*, 21, 116–142.
- [17] Phillips, P. C. B. and V. Solo (1992): “Asymptotics for Linear Processes,” *Annals of Statistics*, 20, 971–1001
- [18] Phillips, P. C. B., Y. Sun and S. Jin (2006): “Spectral Density Estimation and Robust Hypothesis Testing Using Steep Origin Kernels Without Truncation,” *International Economic Review*, 21, 837–894.
- [19] ——— (2007): “Long-Run Variance Estimation and Robust Regression Using Sharp Origin Kernels with No Truncation,” *Journal of Statistical Planning and Inference*, 137, 985–1023.
- [20] Percival D. and A. T. Walden (1993): *Spectral Analysis For Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge University Press.
- [21] Priestley, M. B. (1981): *Spectral Analysis and Time Series*, New York: Academic Press.
- [22] Sun, Y. (2006): “Best Quadratic Unbiased Estimators of Integrated Variance in the Presence of Market Microstructure Noise,” Working paper, Department of Economics, UC San Diego.
- [23] Sun, Y., P. C. B. Phillips, and S. Jin (2008): “Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing,” *Econometrica*, 76, 175–194.
- [24] Sun, Y. and P. C. B. Phillips (2009): “Bandwidth Choice for Interval Estimation in GMM Regression,” Working paper, Department of Economics, UC San Diego.

- [25] Taniguchi, M and M. L. Puri (1996): “Valid Edgeworth Expansions of M-estimators in Regression Models with Weakly Dependent Residuals,” *Econometric Theory*, 12, 331–346.
- [26] Thomson, D. J. (1982): “Spectrum Estimation and Harmonic Analysis,” *IEEE Proceedings*, 70, 1055–1096.
- [27] Velasco, C. and P. M. Robinson (2001): “Edgeworth Expansions for Spectral Density Estimates and Studentized Sample Mean,” *Econometric Theory*, 17, 497–539.
- [28] Vogelsang, T. J. and P. H. Franses (2005): “Testing for Common Deterministic Trend Slopes”, *Journal of Econometrics*, 126, 1–24