

Robust Trend Inference with Series Variance Estimator and Testing-optimal Smoothing Parameter

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

*Email: yisun@ucsd.edu. The author gratefully acknowledges partial research support from NSF under Grant No. SES-0752443. He thanks two anonymous referees for helpful comments and suggestion. 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 series type estimator with carefully selected basis functions. Regardless of whether the number of basis functions K is fixed or grows with the sample size, the Wald statistic converges to a standard distribution. It is shown that critical values from the fixed- K asymptotics are second order correct under the large- K asymptotics. A new practical approach is proposed to select K that addresses the central concern of hypothesis testing: the selected smoothing parameter is testing-optimal in that it minimizes the type II error while controlling for the type I error. Simulations indicate that the new test is as accurate in size as the nonstandard test of Vogelsang and Franses (2005) and as powerful as the corresponding Wald test based on the large- K asymptotics. The new test therefore combines the advantages of the nonstandard test and the standard Wald test while avoiding their main disadvantages (power loss and size distortion, respectively).

JEL Classification: C13; C14; C32; C51

Keywords: Asymptotic expansion, F-distribution, Hotelling's T^2 distribution, long-run variance, robust standard error, series method, testing-optimal smoothing parameter choice, trend inference, Type I and Type II errors.

1 Introduction

Trend regression is one of simple and yet 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 GDPs 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. Commonly used methods are mostly kernel-based. In this paper, we consider estimating the LRV using nonparametric series methods. The resulting series LRV estimator is the sample variance of regression coefficients in a nonparametric series regression.

The smoothing parameter in the series LRV estimator is the number of basis functions employed. When the number of basis functions K is fixed, the LRV estimator is inconsistent and converges to a scaled Wishart distribution. The underlying scale cancels out in the limiting distribution the asymptotic variance of the trend estimator. Hence, when K is fixed, the Wald statistic converges to a pivotal nonstandard distribution. The fixed- K asymptotics is in the spirit of the fixed- b asymptotics as in Kiefer and Vogelsang (2002a, 2002b, 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 those based on consistent LRV estimates. Jansson (2004) and Sun, Phillips and Jin (2008, SPJ hereafter) provide some theoretical justifications for the non-standard asymptotic theory.

We design a set of basis functions so that the fixed- K asymptotic distribution becomes the standard F distribution. For these basis functions, the series LRV estimator is asymptotically invariant to the intercepts and trend parameters. As a result, it does not suffer from the bias arising from the estimation uncertainty of model parameters. This is in contrast with the conventional kernel LRV estimators where the estimation uncertainty gives rise to a demeaning bias. See, for example, Hannan (1957) and Ng and Perron (1994). 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, especially in hypothesis testing. Another advantage of using the new series LRV estimator lies in its convenience in practical use as critical values from the fixed- K asymptotics are readily available from statistical tables and software packages.

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. Such a choice of the smoothing parameter is designed to be optimal in the MSE sense for the point estimation of the 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 associated test. It is thus 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. We can employ the traditional Wald test, which is based the Wald statistic and uses a chi-square distribution as the reference distribution. Alternatively, we can employ the new F^* test given in this paper, which is based on a modified Wald statistic and uses an F-distribution as the reference distribution. We find that critical values from the F-distribution are higher order correct under the conventional large- K asymptotics. A direct implication is that the F^* test generally 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 the F^* test to conduct inference on the trend parameters.

One of the main contributions of the paper is to develop an optimal procedure for selecting the smoothing parameter K that addresses the central concern of semiparametric testing. The ultimate goal of any testing problem is to achieve smaller type I and type II errors. However, these two types of errors often move in opposite directions. We can control one type of error while trying to minimize the other type of error. In this paper, 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 testing-optimal for the given bound. The bound is defined to be $\kappa\alpha$, where α 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 proposed approach to selecting the testing-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. In a transformed space, the null hypothesis is a fixed point while the alternative hypothesis we consider 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.

The testing-optimal K that maximizes the local asymptotic power while preserving size in large samples is fundamentally different from the MSE-optimal K . The testing-optimal K depends on the sign of the nonparametric bias, the hypothesis under consideration and the permitted tolerance for the type I error while the MSE-optimal K does not. When the permitted tolerance becomes sufficiently small, the testing-optimal K is of smaller order than the MSE-optimal K . 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 where robust inference for the mean of a scalar time series is considered. In SPJ, the optimal smoothing parameter 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

for the constrained minimization problem as the relative weight. The main difference is that our weight is implicitly defined through the tolerance parameter κ . For a given κ , 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 same procedure is used in Sun, Phillips, and Jin (2010) for robust mean inference with exponentiated kernels.

The series LRV estimator has been considered in the literature under different names. It belongs to the class of multi-window or multi-taper estimators (Thomson, 1982) and the class of filter-bank estimators (Stoica and Moses, 2005, ch. 5). In the simulation and signal processing literature, the weighted area estimator of Foley and Goldman (1999) is a series LRV estimator with particular basis functions. In econometrics, Phillips (2005) embeds this estimator in a framework of automated regression. Müller (2007) motivates it from the perspective of robust LRV estimation. The fixed- K type of asymptotics has some precursors in the literature. Foley and Goldman (1999) approximate the distribution of their autocorrelation robust t-statistic by a t distribution. As we show later, the t-distribution belongs to the class of fixed- K asymptotic distributions. For some basis functions, the working paper version of Müller (2007) contains the fixed- K asymptotics and F approximation. However, the basis functions considered here are different from the existing literature. They do not constitute a complete basis system, and are designed to eliminate the demeaning effect and the detrending effect at the same time. To the best of my knowledge, the paper is the first to explore the relationship between the fixed- K asymptotics and the conventional large- K asymptotics in trend estimation and inference. It is also the first to propose a testing-optimal smoothing parameter choice in this setting.

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 series LRV estimator and establishes its asymptotic properties under the fixed- K and large- K asymptotics. Section 4 investigates the limiting distribution of the Wald statistic 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, Section 6 proposes a selection rule for K that is most suitable for implementation in semiparametric testing. The next section reports simulation evidence on the performance of the new procedure. The last section provides some concluding discussion. Proofs 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

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

where u_{it} is a weakly dependent process with zero mean. Our focus of interest is on the inference about the trend parameters $\{\beta_i\}$.

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 $v \geq 4$,

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

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

Under the above assumption, the process u_t admits the following BN (Beveridge and Nelson, 1981) decomposition

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

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 Ω of u_t :

$$\Omega = \Lambda \Lambda' = \sum_{j=-\infty}^{\infty} E u_t u_{t-j}' = C(1)\Sigma C(1)'$$

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 θ 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, for u_t defined in Assumption 1,

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

Therefore

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

So the OLS estimator $\hat{\beta}_{OLS}$ of β 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 Series LRV Estimator and its Asymptotic Properties

To conduct inference regarding β , we need to first estimate the LRV matrix Ω . In the next subsection, we motivate the series LRV estimator we use in this paper.

3.1 Motivation of Series LRV Estimator

Consider the kernel-based estimator proposed by Phillips, Sun and Jin (2006, 2007, PSJ hereafter):

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

where $\mathcal{K}_\rho(r-s) = [\mathcal{K}(r-s)]^\rho$ for some second-order kernel function $\mathcal{K}(\cdot)$. This estimator is consistent when $\rho \rightarrow \infty$ at a certain rate. Assume that $\mathcal{K}(\cdot)$ is even, continuous and positive semidefinite. By Mercer's theorem (Mercer, 1909), 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 and $\{\phi_k(r)\}$ is an orthonormal sequence of eigenfunctions corresponding to the eigenvalues λ_k . It can be shown that

$$\sum_{k=1}^{\infty} \lambda_k = 1 \text{ and } \sum_{k=1}^{\infty} \lambda_k^2 = O\left(\frac{1}{\sqrt{\rho}}\right) \text{ as } \rho \rightarrow \infty. \quad (5)$$

With this representation of $\mathcal{K}_\rho(r-s)$, we can write

$$\begin{aligned} \hat{\Omega}_{PSJ} &= \sum_{k=1}^{\infty} \lambda_k \left[\frac{1}{\sqrt{T}} \sum_{r=1}^T \phi_k \left(\frac{r}{T} \right) \hat{u}_r \right] \left[\frac{1}{\sqrt{T}} \sum_{s=1}^T \phi_k \left(\frac{s}{T} \right) \hat{u}_s \right]' \\ &: = \sum_{k=1}^{\infty} \lambda_k \hat{\Omega}_k \end{aligned} \quad (6)$$

where

$$\hat{\Omega}_k = \hat{\Lambda}_k \hat{\Lambda}'_k \text{ and } \hat{\Lambda}_k = \frac{1}{\sqrt{T}} \sum_{t=1}^T \phi_k \left(\frac{t}{T} \right) \hat{u}_t.$$

In the above expression, λ_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 (low frequency) properties of the underlying time series. In addition, for each k , $\lambda_k \rightarrow 0$ as $\rho \rightarrow \infty$. Implicitly, the PSJ estimator employs a soft thresholding method where the weight λ_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:

$$\hat{\Omega} = \frac{1}{K} \sum_{k=1}^K \hat{\Omega}_k \quad (7)$$

where K is a positive integer. This estimator truncates the infinite sum in (6) and assigns equal weights to the remaining terms. In other words, the infinite sequence $(\lambda_1, \dots, \lambda_K, \dots)$ is replaced by $(1/K, 1/K, \dots, 1/K, 0, \dots)$. For this sequence, $\sum_{k=1}^{\infty} \lambda_k = 1$ and $\sum_{k=1}^{\infty} \lambda_k^2 = 1/K$. Comparing the squared sum with that in (5), we can see that K plays the role of $\sqrt{\rho}$ in the PSJ estimator. This can also be seen by comparing the asymptotic biases and the asymptotic variances of these two estimators.

As will be shown below, with appropriately chosen ϕ_k , each of the summand $\hat{\Omega}_k$ is an asymptotically unbiased estimator of Ω . 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.

Note that $\hat{\Lambda}_k$ is approximately the regression coefficient obtained by regressing the time series \hat{u}_t on the regressor $\phi_k(t/T)/\sqrt{T}$. $\hat{\Omega}_k$ is part of ‘the total sum of squares’ $\sum_{t=1}^T \hat{u}_t \hat{u}'_t$ that is explained by the basis function $\phi_k(\cdot)/\sqrt{T}$. This explained sum of squares may be regarded as another ways of thinking about the long run variance matrix—the contributions to the variation of \hat{u}_t that are due to low frequency variations in the series.

To obtain more flexible estimators of the form $\hat{\Omega}$, we can use any orthogonal basis functions to construct $\hat{\Omega}$. For example, we may use polynomial basis functions. In addition, the basis functions do not have to be a complete basis system. In fact, we use incomplete basis functions below in order to remove the demeaning and detrending effects. We have thus obtained a general class of LRV estimators. For convenience, we refer to them as series LRV estimators as they are based on nonparametric series regressions.

The series LRV estimator has different interpretations. First, it can be regarded as a multiple-window estimator with window function $\phi_k(t/T)/\sqrt{T}$, see Thomson (1982) and Percival and Walden (1993). In the econometrics literature, 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. In a different context and for a different model, Müller (2007) has established the fixed- K asymptotics given in Section 3.2. Phillips (2005) gives an alternative motivation of the multiple-window estimator and establishes its asymptotic properties. Second, when $\phi_k(1-x) = \phi_k(x)$, we can write $\hat{\Lambda}_k = 1/\sqrt{T} \sum_{\tau=0}^{T-1} \phi_k(1 - \frac{\tau}{T}) \hat{u}_{T-\tau}$, which can be regarded as output from applying a linear filter to the residual process \hat{u}_t . The transfer function of the linear filter is

$$H_k(\omega) = \frac{1}{\sqrt{T}} \sum_{\tau=0}^{T-1} \phi_k(1 - \frac{\tau}{T}) \exp(i\tau\omega).$$

To capture the long run behavior of the process, we require that $H_k(\omega)$ be concentrated around the origin. That is, $H_k(\omega)$ resembles a band pass filter that passes low frequencies within a certain range and rejects (attenuates) frequencies outside that range. Hence, $\hat{\Omega}_k$ can also be regarded as a filter-bank estimator and $\hat{\Omega}$ is a simple average of these filter-bank estimators. Finally, $\hat{\Omega}$ can be regarded as the sample variance of regression coefficients $\{\hat{\Lambda}_k, k = 1, 2, \dots, K\}$. By construction, it is automatically positive semidefinite, a desirable property for practical use.

Many series LRV estimators can be obtained by choosing different basis functions. However, in nonparametric series estimation, it is a conventional wisdom that the choice of basis functions is often less important than the choice of the smoothing parameter. For this reason, we employ the basis functions that are most convenient for practical use and focus on the problem of selecting the smoothing parameter K .

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

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

Using summation and integration by parts and invoking the continuous mapping theorem, we obtain, for fixed K and under Assumption 1:

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

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

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

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 ϕ_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 basis functions, ζ_k is iid $N(0, \mathbb{I}_n)$. As a result, $\sum_{k=1}^K \zeta_k \zeta_k'$ is a Wishart distribution $\mathbb{W}_n(\mathbb{I}_n, K)$. So $\hat{\Omega}$ converges to a scaled Wishart distribution. In the scalar case, the limiting distribution reduces to the scaled chi-square distribution χ_K^2/K . In general, for any conforming constant vector z , $z' \hat{\Omega} z / z' \Omega z$ converges in distribution to χ_K^2/K . This result can be used to test hypotheses regarding Ω . 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 β .

3.3 Large- K Asymptotics

While the fixed- K asymptotics may capture the randomness of $\hat{\Omega}$ very well, it does not reflect the usual nonparametric bias or Parzen bias of $\hat{\Omega}$. In this section, we consider the asymptotic properties of $\hat{\Omega}$ when both K and T go to infinity such that $K/T \rightarrow 0$.

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

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

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

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

\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.

Theorem 2 extends Theorem 1 of Phillips (2005), which is applicable only to scalar time series with known mean. The bias term here is different from that given in Theorem 1(i) in Phillips (2005). This is because the basis functions we used are 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 θ 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. The higher order bias is not captured in the first-order conventional asymptotic theory, although it is reflected in the nonstandard fixed-b asymptotics. 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. We note in passing that the estimation uncertainty may also be eliminated using recursive OLS residuals.

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).

Let

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

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

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

So the MSE optimal K is given by

$$K = \left(\frac{tr [W (\Omega \otimes \Omega) (\mathbb{I}_{n^2} + \mathbb{K}_{nn})]}{4vec(B)' W vec(B)} \right)^{1/5} T^{4/5}. \quad (11)$$

This approach to optimal K choice is the same as that for bandwidth choice in kernel LRV estimators. See, for example, Andrews (1991).

4 Autocorrelation Robust Inference for 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

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

4.1 Fixed- K Asymptotics

Under the fixed- K asymptotics and the null hypothesis

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

It turns out the scaling factor in the asymptotic distribution of $\hat{\Omega}$ cancels out with that in the asymptotic distribution of $T^{3/2}(\hat{\beta}_{OLS} - \beta)$. To see this, we represent the distribution $R_{p \times n} \Lambda W_n(r)$ by $R^* W_p^*(r)$ 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

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

So the limiting distribution of $F_{T,OLS}$ does not depend on Λ and is pivotal.

Since

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

η and ξ_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 $\mathbb{W}_p(\mathbb{I}_p, K)$. Hence the limiting distribution of $F_{T,OLS}$ is Hotelling's T-square distribution (Hotelling (1931)):

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

Since for $K \geq p$,

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

we have

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

where χ_p^2 and χ_{K-p+1}^2 denote independent χ^2 random variables.

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. These fixed- K asymptotic results can also be proved directly using standard techniques from multivariate statistical analysis.

We have therefore shown that under the fixed- K asymptotics, the scaled Wald statistic converges weakly to the F distribution with degrees of freedom $(p, K - p + 1)$ 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,

$$H_1(\delta^2) : R\beta = r + c / (T\sqrt{T}) \quad \text{where } c = (R12\Omega R')^{1/2} \tilde{c} \quad (12)$$

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

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

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 the notation 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

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

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 χ_p^2 distribution. By definition and with a

slight abuse of notation, we have, as $K \rightarrow \infty$,

$$\begin{aligned}
1 - \alpha &= P(F_{p,K-p+1} < F_{p,K-p+1}^\alpha) = P\left(\chi_p^2 < pF_{p,K-p+1}^\alpha \frac{\chi_{K-p+1}^2}{(K-p+1)}\right) \\
&= EG_p\left(pF_{p,K-p+1}^\alpha \frac{\chi_{K-p+1}^2}{(K-p+1)}\right) \\
&= G_p(pF_{p,\infty}^\alpha) + G'_p(pF_{p,\infty}^\alpha) E\left[pF_{p,K-p+1}^\alpha \frac{\chi_{K-p+1}^2}{(K-p+1)} - pF_{p,\infty}^\alpha\right] \\
&\quad + \frac{1}{2}G''_p(pF_{p,\infty}^\alpha) E\left[pF_{p,K-p+1}^\alpha \frac{\chi_{K-p+1}^2}{(K-p+1)} - pF_{p,\infty}^\alpha\right]^2 + o\left(\frac{1}{K^2}\right) \quad (13) \\
&= G_p(pF_{p,\infty}^\alpha) + G'_p(pF_{p,\infty}^\alpha) [pF_{p,K-p+1}^\alpha - pF_{p,\infty}^\alpha] \\
&\quad + \frac{1}{K}G''_p(pF_{p,\infty}^\alpha) (pF_{p,\infty}^\alpha)^2 + o(pF_{p,K-p+1}^\alpha - pF_{p,\infty}^\alpha) + o\left(\frac{1}{K}\right).
\end{aligned}$$

Therefore

$$pF_{p,K-p+1}^\alpha = \chi_p^\alpha - \frac{1}{K} \frac{G''_p(\chi_p^\alpha)}{G'_p(\chi_p^\alpha)} (\chi_p^\alpha)^2 + o\left(\frac{1}{K}\right) \text{ as } K \rightarrow \infty. \quad (14)$$

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 χ^2 -distribution, reflecting the randomness in the denominator of the Wald statistic. Up to the order $o(1/K)$, the correction term $(\chi_p^\alpha - p + 2) \chi_p^\alpha / (2K)$ increases with p and decreases with K . So when K is small or p is large, the difference between the F and χ^2 approximations may be large.

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 testing-optimal procedure to select K . We make the simplification assumption that u_t is normal, which facilitates the derivations. The assumption could be relaxed but at the cost of much greater complexity, see for example, Sun and Phillips (2009).

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

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

Similarly, the OLS estimator satisfies

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

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

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

It follows from the asymptotic equivalence of $\hat{\theta}_{OLS}$ and $\hat{\theta}_{GLS}$ that $Ec'\Delta_\beta\Delta'_\beta c = O(1/T)$ 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 Δ 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 Δ 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 for $K \geq p$,*

- (a) $P\left(\frac{(K-p+1)}{K}F_{T,GLS} < z\right) = EG_p\left(z\frac{K}{K-p+1}\Xi^{-1}\right) + O\left(\frac{1}{T}\right),$
 - (b) $P\left(\frac{(K-p+1)}{K}F_{T,OLS} < z\right) = P\left(\frac{(K-p+1)}{K}F_{T,GLS} < z\right) + O\left(\frac{1}{T}\right),$
- where G_p is the CDF of a χ^2 random variable with degrees of freedom p ,

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

and $\Omega_{T,GLS} = var\left[T^{3/2}\left(\hat{\beta}_{GLS} - \beta\right)\right] / 12$.

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

$$\Xi^{-1} = 1 + L + Q + o_p\left(\frac{1}{K} + \frac{K^2}{T^2}\right) + O_p\left(\frac{1}{T}\right),$$

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. Plugging this stochastic expansion into Lemma 3, we obtain a higher order expansion of the finite sample distribution of $F_{T,OLS}$ 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*

$$\begin{aligned} & P\left(\frac{(K-p+1)}{K}F_{T,OLS} < z\right) \\ = & G_p(z) + \frac{K^2}{T^2}G'_p(z)z\bar{B} + \frac{1}{K}G''_p(z)z^2 + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right) \end{aligned} \quad (15)$$

where

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

The first term in (15) 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 regressors.

To understand the relationship between the fixed- K and large- K asymptotics, we develop an expansion of the limiting $pF_{p,K-p+1}$ distribution as in (13):

$$P(pF_{p,K-p+1} < z) = G_p(z) + \frac{1}{K}G''_p(z)z^2 + o\left(\frac{1}{K}\right), \text{ as } K \rightarrow \infty.$$

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 $z = pF_{p,K-p+1}^\alpha$ 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

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

Therefore, use of critical value $pF_{p,K-p+1}^\alpha$ removes the variance term $K^{-1}G''_p(z)z^2$ 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 χ_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 value $pF_{p,K-p+1}^\alpha$ 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.

The fixed- K asymptotic distribution of $F_{T,OLS}$ is $K(K-p+1)^{-1} pF_{p,K-p+1}^\alpha$ while its first-order large- K asymptotic distribution is χ_p^2 . When K is fixed, the two distributions are different. Hence, the large- K asymptotic approximation is not even first-order valid under the fixed- K asymptotics.

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 (15), 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 χ_p^α , 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 conventional 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. $F_{T,OLS}^*$ can be viewed as the standard Wald statistic but using the following estimator for $R\Omega R'$:

$$\widehat{R\Omega R'} = \frac{1}{K-p+1} \sum_{k=1}^K R\hat{\Omega}_k R'.$$

So the finite sample correction factor $(K-p+1)/K$ can be viewed as a degree-of-freedom adjustment.

The following theorem gives the type I and type II errors 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 type I error of the F^* test is*

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

(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 type II error of the F^**

test is

$$\begin{aligned}
P(F_{T,OLS}^* < pF_{p,K-p+1}^\alpha | H_1(\delta^2)) &= G_{p,\delta^2}(\chi_p^\alpha) + \frac{K^2 \bar{B}}{T^2} G'_{p,\delta^2}(\chi_p^\alpha) \chi_p^\alpha \\
&+ \frac{1}{K} \mathcal{Q}_{p,\delta^2}(\chi_p^\alpha) (\chi_p^\alpha)^2 + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right),
\end{aligned} \tag{18}$$

where $G_{p,\delta^2}(\cdot)$ and $G'_{p,\delta^2}(\cdot)$ are the CDF and pdf of the noncentral χ^2 distribution with degrees of freedom p and noncentrality parameter δ^2 and

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

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_β 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

$$e_{\beta\delta} = \frac{(R12\Omega_{T,GLS}R')^{-1/2} RT^{3/2}(\hat{\beta}_{GLS} - \beta) + \tilde{c}}{\left\| (R12\Omega_{T,GLS}R')^{-1/2} RT^{3/2}(\hat{\beta}_{GLS} - \beta) + \tilde{c} \right\|}$$

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

In this section, we provide a novel approach to smoothing parameter selection that is most suitable for semiparametric testing.

6.1 Optimal K Formula

In view of the asymptotic expansion in (17) and ignoring the higher order terms, we can approximate the type I error of the F^* test by

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

Similarly, from (18), the type II error of the F^* test can be approximated by

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

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

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

where κ is a constant greater than 1. Ideally, the type I error is less than or equal to the nominal type I error α . In finite samples, there are always some approximation error and we allow for some discrepancy by introducing the tolerance factor κ . 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%. We may allow κ to depend on the sample size T . For a larger sample size, we may require κ to take smaller values.

Note that both the type I and type II errors depend on the asymptotic bias of the estimator $R\hat{\Omega}R'$ through \bar{B} , the relative bias of estimating the variance of $RT^{3/2}(\hat{\beta}_{OLS} - \beta)$. Our testing-oriented criterion is in sharp contrast with the MSE criterion, which depends on a quadratic form of the asymptotic bias of $\hat{\Omega}$. In large samples, the quadratic form is of smaller order than the bias itself. So for testing problems, it is more important to reduce the bias of the LRV estimator as compared to the point estimation of the LRV matrix. In addition, the quadratic form is invariant to sign of B . The MSE-optimal K is the same for B and $-B$. In contrast, for the testing-optimal K , the sign of B (hence that of \bar{B}) is of vital importance as shown below.

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

$$K_{\text{opt}} = \left(\frac{\delta^2 G'_{(p+2),\delta^2}(\chi_p^\alpha)}{4\bar{B}G'_{p,\delta^2}(\chi_p^\alpha)} \right)^{1/3} T^{2/3}. \quad (19)$$

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 λ be the Lagrange multiplier, and define

$$\begin{aligned} L(K, \lambda) &= G_{p,\delta^2}(\chi_p^\alpha) + \frac{K^2 \bar{B}}{T^2} G'_{p,\delta^2}(\chi_p^\alpha) \chi_p^\alpha + \frac{1}{K} \frac{\delta^2}{2} G'_{(p+2),\delta^2}(\chi_p^\alpha) \chi_p^\alpha \\ &+ \lambda \left(\left(\alpha - \frac{K^2 \bar{B}}{T^2} G'_{p,\delta^2}(\chi_p^\alpha) \chi_p^\alpha \right) - \kappa \alpha \right). \end{aligned} \quad (20)$$

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

$$K_{\text{opt}} = \left(\frac{(\kappa - 1) \alpha}{|\bar{B}| G'_{p,\delta^2}(\chi_p^\alpha) \chi_p^\alpha} \right)^{1/2} T, \quad (21)$$

and the corresponding Lagrange multiplier is

$$\lambda_{\text{opt}} = \frac{G'_{p,\delta^2}(\chi_p^\alpha)}{G'_{p,\delta^2}(\chi_p^\alpha)} + \frac{|\bar{B}|^{1/2} \delta^2 G'_{(p+2),\delta^2}(\chi_p^\alpha) [\chi_p^\alpha]^{3/2} [G'_{p,\delta^2}(\chi_p^\alpha)]^{1/2}}{4[(\kappa - 1) \alpha]^{3/2} T}.$$

Formulae (19) and (21) can be written collectively as

$$K_{\text{opt}} = \left[\frac{\delta^2 G'_{(p+2), \delta^2}(\chi_p^\alpha)}{4\bar{B} [G'_{p, \delta^2}(\chi_p^\alpha) - \lambda_{\text{opt}} G'_p(\chi_p^\alpha)]} \right]^{1/3} T^{2/3},$$

where

$$\lambda_{\text{opt}} = \begin{cases} 0, & \text{if } \bar{B} > 0 \\ \frac{G'_{p, \delta^2}(\chi_p^\alpha)}{G'_p(\chi_p^\alpha)} + \delta^2 \frac{|\bar{B}|^{1/2} G'_{(p+2), \delta^2}(\chi_p^\alpha) [\chi_p^\alpha]^{3/2} [G'_p(\chi_p^\alpha)]^{1/2}}{4[(\kappa-1)\alpha]^{3/2} T}, & \text{if } \bar{B} < 0 \end{cases} \quad (22)$$

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 the size distortion is expected to be negative, the optimal Lagrange multiplier is zero and we assign all weight to the type II error. In this case, the expansion rate of the optimal K is $O(T^{2/3})$. When the size distortion is expected to be positive, the Lagrange multiplier is positive. In this case, the loss function is a genuine weighted sum of type I and type II errors. The optimal K has an expansion rate that increases with the tolerance on the type I error. When the permitted tolerance is very low so that $\kappa - 1 \sim 1/T^2$, the optimal K is bounded. The fixed- K rule can be interpreted as assigning increasingly more weight to the type I error as the sample size increases. On the other hand, when the permitted tolerance is high so that $\kappa - 1 = O(1)$, the optimal K has an expansion rate of $O(T)$, which is faster than the MSE-optimal expansion rate.

All else being equal, the optimal K decreases with $|\bar{B}|$. This is expected, as the asymptotic bias of $\hat{\Omega}$ increases with both K and $|\bar{B}|$. When $|\bar{B}|$ is large, we should choose a small K to offset the bias effect.

The formula for K_{opt} depends on the noncentrality parameter δ^2 . For practical implementation, we suggest choosing δ^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 α . As usual, we consider $\alpha = 5\%$ and 10% . The value of δ^2 can be easily computed using standard statistical programs. Since K is an integer greater than or equal to p , in practice, we take $\max(\lceil K_{\text{opt}} \rceil, p)$ as the K value, where $\lceil \cdot \rceil$ is the ceiling function.

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 permitted tolerance is very low to $O(T)$ when the permitted tolerance is very high.

6.2 Data Driven Implementation

The optimal K in (22) depends on the data generating process only through the parameter \bar{B} . We can therefore write $K_{\text{opt}} = K_{\text{opt}}(\bar{B})$. The unknown parameter \bar{B} can be estimated by a standard plug-in procedure based on a simple parametric model like VAR (e.g. Andrews (1991)). More specifically, the plug-in procedure 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 (22) gives the automatic bandwidth \hat{K} .

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 Ω and B are

$$\hat{\Omega} = (\mathbb{I}_n - \hat{A})^{-1} \hat{\Sigma} (\mathbb{I}_n - \hat{A}')^{-1}, \quad (23)$$

$$\begin{aligned} \hat{B} = & -\frac{2\pi^2}{3} (\mathbb{I}_n - \hat{A})^{-3} \left(\hat{A} \hat{\Sigma} + \hat{A}^2 \hat{\Sigma} \hat{A}' + \hat{A}^2 \hat{\Sigma} - 6 \hat{A} \hat{\Sigma} \hat{A}' \right. \\ & \left. + \hat{\Sigma} (\hat{A}')^2 + \hat{A} \hat{\Sigma} (\hat{A}')^2 + \hat{\Sigma} \hat{A}' \right) (\mathbb{I}_n - \hat{A}')^{-3}. \end{aligned} \quad (24)$$

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 Ω and B , the data-driven automatic bandwidth can be computed as

$$\hat{K}_{opt}^* = \max \left(\left[\hat{K}_{opt}(\bar{B}(R, \hat{B}, \hat{\Omega})) \right], p \right). \quad (25)$$

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

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

where \mathbb{J}_n is a matrix of ones. The parameter μ determines the degree of cross-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 pair wise 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, $\Omega = (1 - \rho^2) (\mathbb{I}_n - A)^{-1} \Sigma (\mathbb{I}_n - A')^{-1}$. For the VMA(1) process $\Omega = (1 - \rho^2) (\mathbb{I}_n + A/\sqrt{1 - \rho^2}) \Sigma (\mathbb{I}_n + A/\sqrt{1 - \rho^2})'$.

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 two different sample sizes $T = 300, 500$.

Table 1: Type I error of different tests for VAR(1) error with $T = 300, \kappa = 1.1$ and $\mu = 1$

	New	MSE	Hybrid	VF	New	MSE	Hybrid	VF
	$p = 1$				$p = 2$			
$\rho = 0$	0.0517	0.0549	0.0508	0.0491	0.0505	0.0609	0.0499	0.0477
$\rho = 0.25$	0.0571	0.0681	0.0590	0.0519	0.0614	0.0814	0.0601	0.0521
$\rho = 0.50$	0.0579	0.0784	0.0619	0.0557	0.0622	0.1023	0.0659	0.0571
$\rho = 0.75$	0.0637	0.0976	0.0665	0.0627	0.0685	0.1492	0.0682	0.0717
	$p = 3$				$p = 6$			
$\rho = 0$	0.0505	0.0666	0.0495	0.0497	0.0504	0.0949	0.0508	0.0484
$\rho = 0.25$	0.0647	0.0965	0.0611	0.0552	0.0815	0.1645	0.0664	0.0602
$\rho = 0.50$	0.0704	0.1353	0.0685	0.0636	0.0936	0.2730	0.0753	0.0787
$\rho = 0.75$	0.0781	0.2236	0.0731	0.0881	0.1124	0.5328	0.0776	0.1401

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 testing-optimal K and uses the F-distribution as the reference distribution. The second one is the conventional Wald test, which is based on the unmodified Wald statistic and MSE-optimal K and uses the χ_p^2 distribution as the reference distribution. 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 empirical type I error of the three testing methods for the VAR(1) error with sample size $T = 300$, tolerance parameter $\kappa = 1.1$, and $\mu = 1$. The table also includes a hybrid procedure that 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 values used. More specifically, let \hat{K}_{mse} be the plug-in estimate of the MSE-optimal K given in (11) and $\hat{F}_{T,OLS}(\hat{K}_{mse})$ be the associated Wald statistic. The hybrid method rejects the null if $\hat{F}_{T,OLS}(\hat{K}_{mse})$ is larger than the critical value $(p\hat{K}_{mse})(\hat{K}_{mse} - p + 1)^{-1}F_{p,\hat{K}_{mse}-p+1}^\alpha$ where $F_{p,\hat{K}_{mse}-p+1}^\alpha$ is the α -level critical value from the F distribution $F_{p,\hat{K}_{mse}-p+1}$. In contrast, the conventional method uses critical values from the χ_p^2 distribution. 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 jointly 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 0.5328, 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 χ^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.

Due to the approximation error, the bound we impose on the approximate type I error does not fully control the empirical type I error. This is demonstrated in Table 1. The quality of approximation depends on the persistence of the time series. When the time series is highly persistent, the first order asymptotic bias of the LRV estimator may not approximate the finite sample bias very well. As a result, the approximate type I error, which is based on the first order asymptotic bias, may not fully capture the empirical type I error. So it is important to keep in mind that the empirical type I error may still be larger than the nominal type I error even if we exert some control over the approximate type I error.

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 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 being jointly tested 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 is therefore expected to have 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 high, the selected K values are 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. Second, the new test is as powerful as the conventional Wald test. This result is encouraging, as the size accuracy of a test is often achieved at the cost of sizable power loss. An example is the VF test. While it has more accurate size than the corresponding Wald test based on the Bartlett kernel LRV estimator, it is also less powerful. See Vogelsang and Franses (2005) for details.

To shed further light on the size and power properties of the new test and the corresponding Wald test under the conventional asymptotics, we present the mean and median of the selected K values for sample size $T = 300$ in Table 2. The (sample) mean and median are computed over 10000 simulation replications. It is clear that for both the testing-oriented criterion and the MSE criterion, the mean and median of the selected K

Table 2: The selected K values based on the testing-oriented criterion and the MSE criterion for VAR(1) error with $T = 300$, $\kappa = 1.1$, and $\mu = 1$

	$\hat{K}_{test}, p = 1$		$\hat{K}_{test}, p = 2$		$\hat{K}_{test}, p = 3$		$\hat{K}_{test}, p = 6$		\hat{K}_{mse}	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
$\rho = 0$	53	44	62	52	62	53	53	50	63	62
$\rho = 0.25$	29	24	33	27	37	29	45	37	36	36
$\rho = 0.50$	16	13	17	14	19	14	25	20	21	21
$\rho = 0.75$	9	6	9	6	11	7	15	12	11	10

Table 3: Type I error for various tests based on VMA(1) error with $T = 100$, $\kappa = 1.1$

	New	MSE	Hybrid	VF	New	MSE	Hybrid	VF
	$p = 1$				$p = 2$			
$\rho = 0$	0.0517	0.0549	0.0508	0.0491	0.0505	0.0609	0.0499	0.0477
$\rho = 0.25$	0.0546	0.0627	0.0540	0.0510	0.0537	0.0727	0.0522	0.0511
$\rho = 0.50$	0.0516	0.0655	0.0539	0.0522	0.0522	0.0804	0.0514	0.0514
$\rho = 0.75$	0.0502	0.0657	0.0531	0.0531	0.0517	0.0827	0.0511	0.0513
	$p = 3$				$p = 6$			
$\rho = 0$	0.0505	0.0666	0.0495	0.0497	0.0504	0.0949	0.0508	0.0484
$\rho = 0.25$	0.0561	0.0850	0.0534	0.0531	0.0661	0.1430	0.0579	0.0560
$\rho = 0.50$	0.0561	0.0989	0.0543	0.0544	0.0644	0.1940	0.0553	0.0599
$\rho = 0.75$	0.0567	0.1066	0.0541	0.0544	0.0645	0.2154	0.0560	0.0615

value increase with the error dependence. While the sample mean and median of the MSE-optimal K are about the same, the sample mean of the testing-optimal K is less than its sample median, implying that the testing-optimal K has relatively few high values. When the number of constraints is small, e.g. $p = 1, 2, 3$, the testing-optimal K is smaller than the MSE-optimal K . This explains why the type I error of the new test is smaller than or about the same as that of the hybrid test. When the number of constraints is 6, the testing-optimal K is larger, which explains the higher size-adjusted power of the new test as compared to the hybrid test. When the sample size increases to 500, the testing-optimal K becomes smaller than the MSE-optimal K for all values of p considered. In this case, the new test has smaller size distortion than the hybrid test for all parameter configurations considered but is also slightly less powerful.

Table 3 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 μ . 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 κ is larger.

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 the VF test and is as powerful as the conventional Wald test based on the MSE-optimal K and χ^2 approximation. 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 ρ 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} , the new F^* test is often as powerful as, albeit sometimes slightly less powerful than, the conventional Wald test we consider here. On the other hand, the new F^* test is much more accurate in size than the Wald test. In terms of the type I error and size-adjusted power, the new F^* test dominates the VF test in an overall sense. Compared to the hybrid test, the new F^* test achieves a smaller type I error for a medium sample size at the cost of very small power loss.

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 series type LRV estimator. Compared to the conventional kernel type LRV estimators, the series LRV estimator enjoys two advantages. First, it is asymptotically invariant to the intercept and trend parameters. This property releases us from worrying about the estimation uncertainty of those parameters. Second, 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.

As a primary contribution of this paper, we propose a new method to select the smoothing parameter in the series 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. This new selection criterion is fundamentally different from the MSE criterion for the point estimation of the LRV. Depending on the permitted tolerance on the type I error, the expansion rate of the testing-optimal smoothing parameter may be larger or smaller than the MSE-optimal smoothing parameter. The fixed smoothing parameter rule can be interpreted as exerting increasingly tight control on the type I error. Monte Carlo experiments show that the size of the new testing procedure is as accurate as the nonstandard test of Vogelsang and Franses (2005) with bandwidth equal to the sample size. It is also as powerful as the conventional Wald test that is based on the series LRV estimator and uses the MSE-optimal smoothing parameter.

The idea of testing-optimal smoothing parameter choice can be extended to usual kernel HAC estimator. Sun (2010) considers kernel HAC estimation in a general GMM framework and develops a testing-optimal procedure for smoothing parameter choice. The method in Sun (2010) can be adopted for trend estimation and inference, leading to a testing-oriented bandwidth choice for the VF type test.

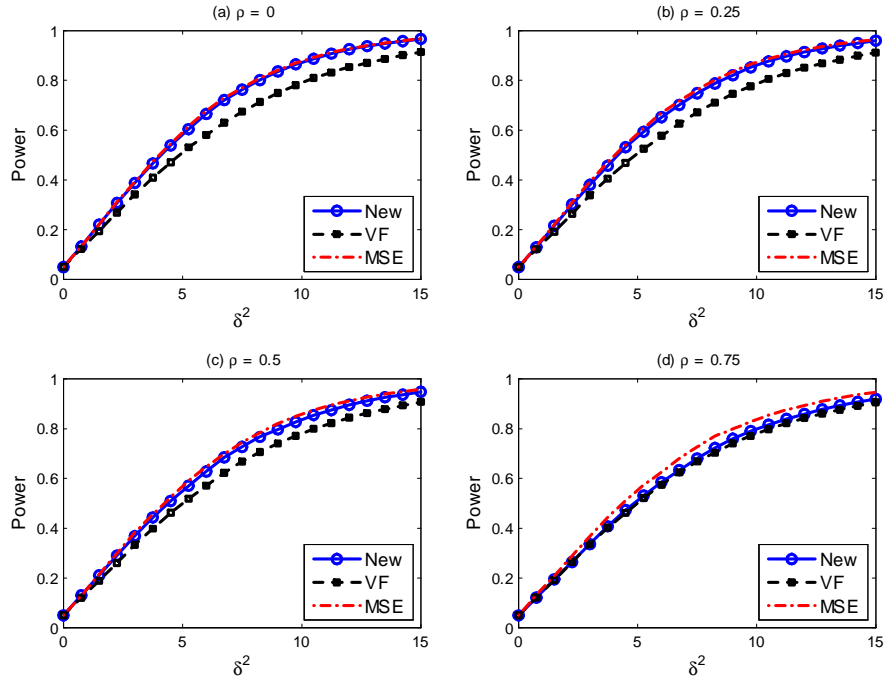


Figure 1: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with $T = 300$, $\kappa = 1.1$ and $p = 1$.

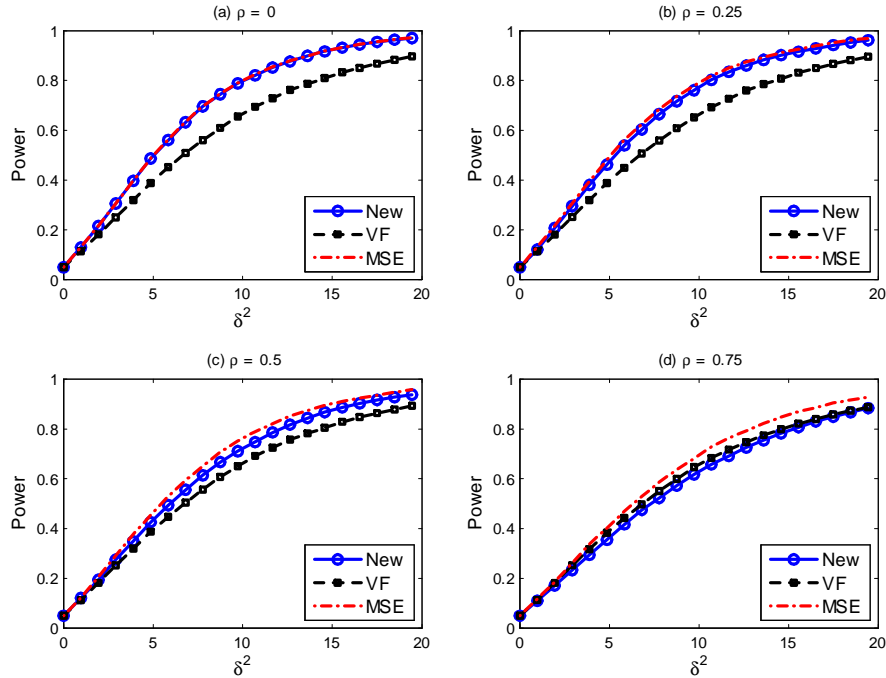


Figure 2: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with $T = 300$, $\kappa = 1.1$ and $p = 2$.

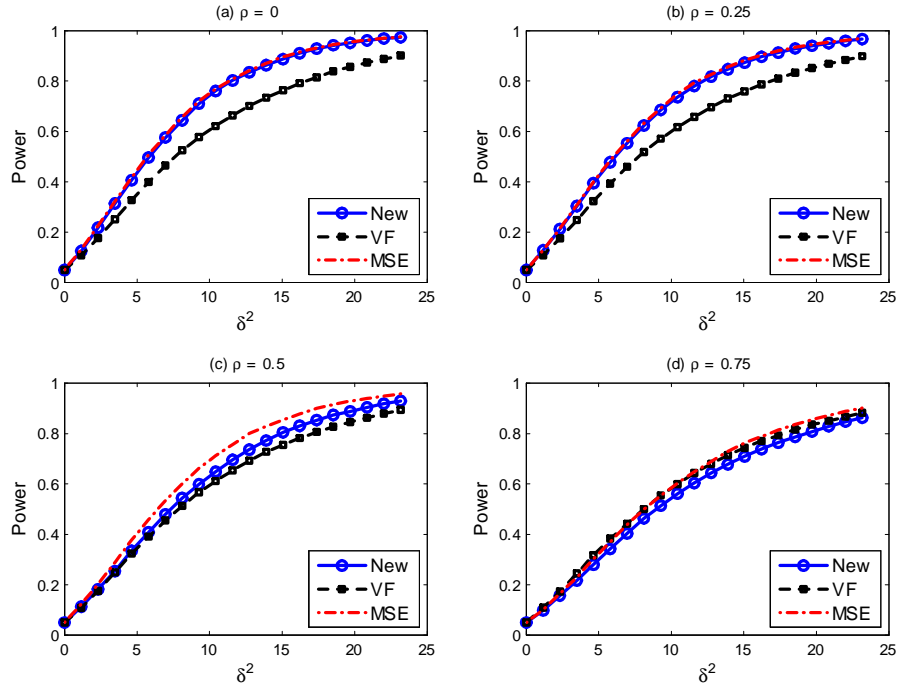


Figure 3: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with $T = 300$, $\kappa = 1.1$ and $p = 3$.

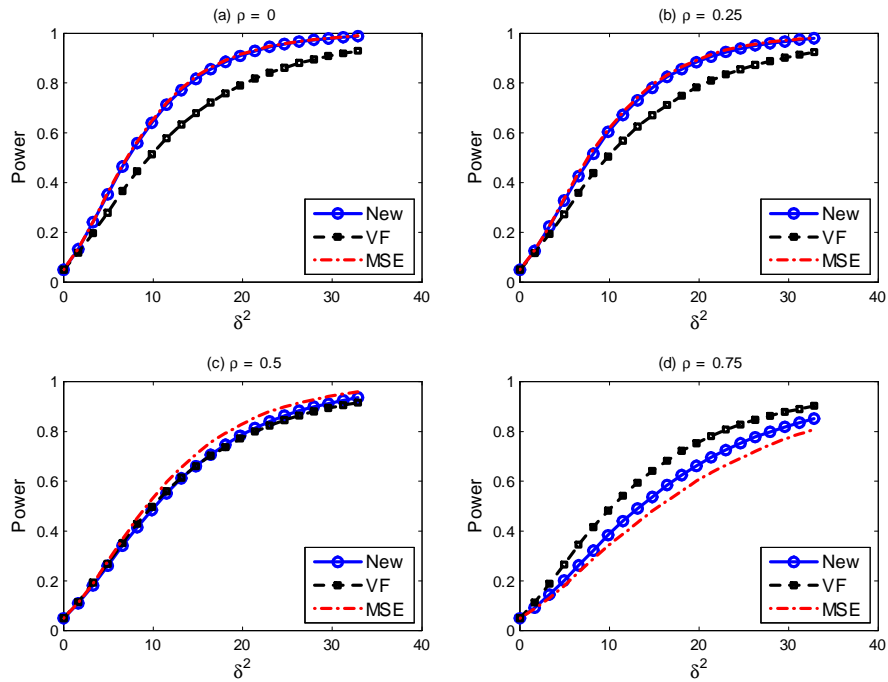


Figure 4: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with $T = 300$, $\kappa = 1.1$ and $p = 6$.

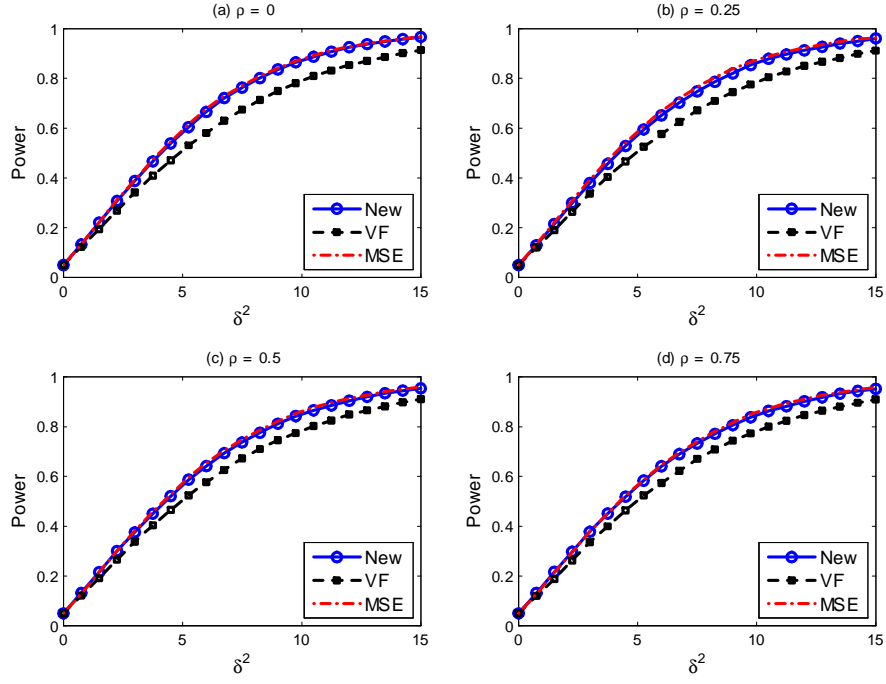


Figure 5: Size-adjusted Power of Different Testing Procedures for VMA(1) Error with $T = 300, \kappa = 1.1$ and $p = 1$.

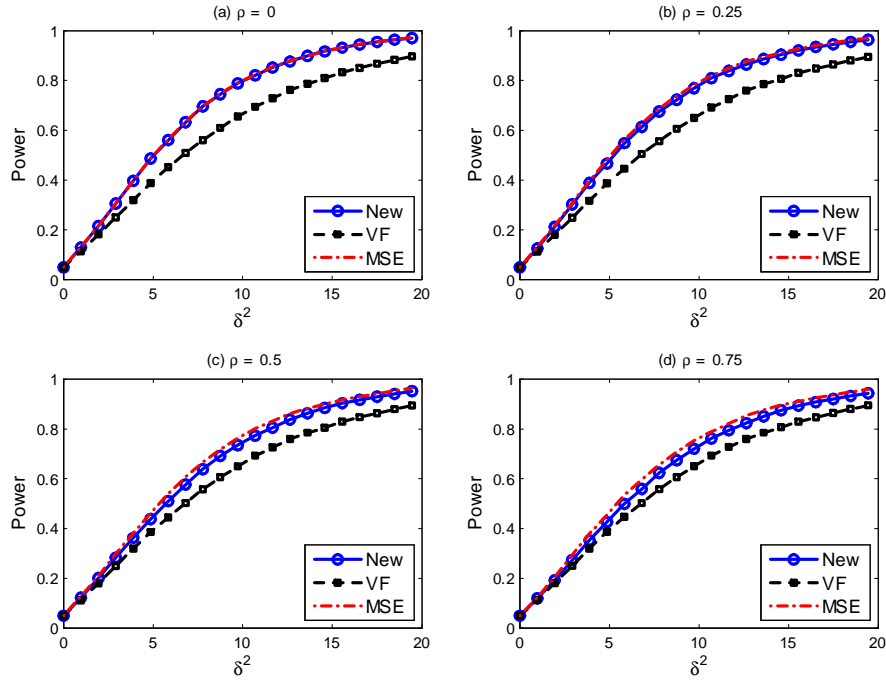


Figure 6: Size-adjusted Power of Different Testing Procedures for VMA(1) Error with $T = 300, \kappa = 1.1$ and $p = 2$.

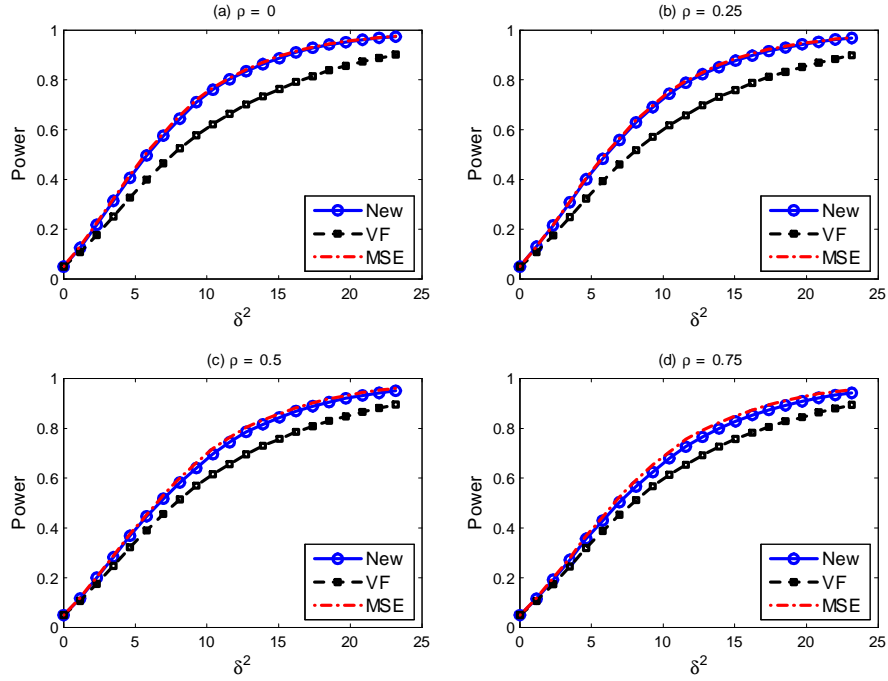


Figure 7: Size-adjusted Power of Different Testing Procedures for VMA(1) Error with $T = 300, \kappa = 1.1$ and $p = 3$.

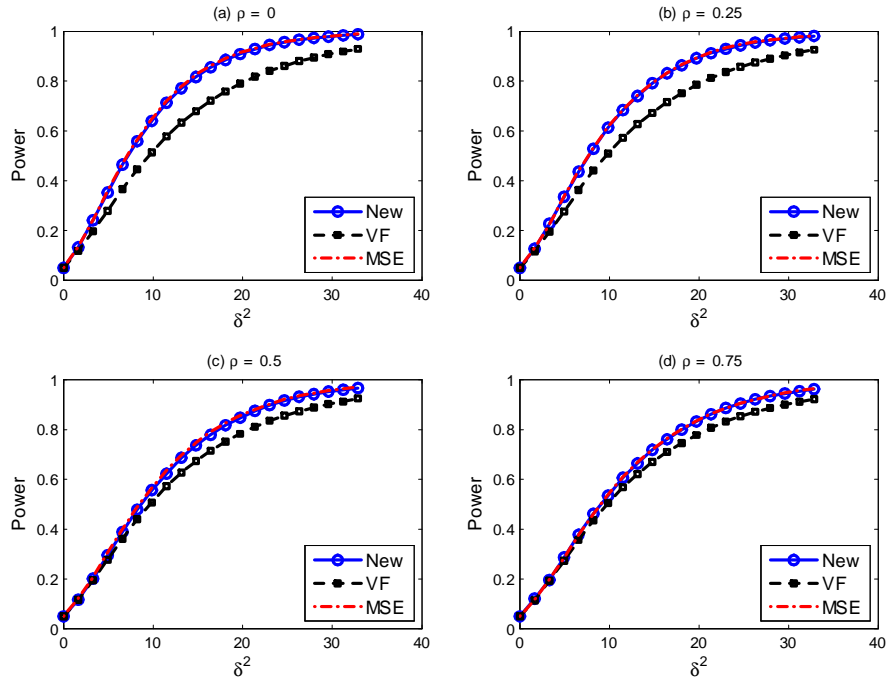


Figure 8: Size-adjusted Power of Different Testing Procedures for VMA(1) Error with $T = 300, \kappa = 1.1$ and $p = 6$.

9 Appendix of Proofs

Proof of Theorem 2. Part (a). Let \bar{x} be the sample 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

$$\tilde{\Omega} = \frac{1}{K} \sum_{k=1}^K \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \check{\phi}_k \left(\frac{t}{T} \right) u_t \right] \left[\frac{1}{\sqrt{T}} \sum_{s=1}^T \phi_k \left(\frac{s}{T} \right) u_s \right]'. \quad (26)$$

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

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

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

$$\frac{T}{L_T^{3/2} K} + \frac{L_T K}{T} = o(1). \quad (27)$$

Next consider

$$\begin{aligned}
& \frac{1}{T} \sum_{1 \leq t, t+h \leq T} \phi_k \left(\frac{t}{T} \right) \left[\phi_k \left(\frac{t+h}{T} \right) - \phi_k \left(\frac{t}{T} \right) \right] \quad (28) \\
&= \frac{2}{T} \sum_{1 \leq t, t+h \leq T} \cos \frac{2\pi kt}{T} \left[\cos \frac{2\pi k(t+h)}{T} - \cos \frac{2\pi kt}{T} \right] \\
&= \frac{1}{T} \sum_{1 \leq t, t+h \leq T} \left\{ \cos \left(\frac{2\pi k(2t+h)}{T} \right) + \cos \left(\frac{2\pi kh}{T} \right) - \left(\cos \left(\frac{4\pi kt}{T} \right) + 1 \right) \right\} \\
&= \left(1 - \frac{|h|}{T} \right) \left(\cos \left(\frac{2\pi kh}{T} \right) - 1 \right) \\
&\quad + \frac{1}{T} \sum_{1 \leq t, t+h \leq T} \left(\cos \left(\frac{4\pi k(t+\frac{1}{2}h)}{T} \right) - \cos \left(\frac{4\pi kt}{T} \right) \right). \quad (29)
\end{aligned}$$

Taking the first term of (29), averaging over k and using the fact that $\|h\| \leq L_T$ and L_T satisfies (27), we get

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

Approximating the sums by integrals, we can show that

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

uniformly over $h \in [-L_T, L_T]$. Hence

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

So

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

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

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

It is not hard but tedious to show that

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

To save space, we omit the details but they are available upon request. Now

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

and

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

Therefore

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

as stated. ■

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

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

where

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

and

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

Here we have used

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

Note that Θ is independent of Ξ because (i) $(\hat{\beta}_{GLS} - \beta)$ is independent of $\hat{\Omega}$. (ii) Θ is the squared length of a standard normal vector and e_β is the direction of this vector. By the

rotation-invariance of standard normals, the length is independent of the direction. Hence

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

as stated.

Part (b). Let

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

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

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

where

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

But

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

where $F'_1 = \partial F / \partial a$. Here we have used: $O(E\zeta_{1T}^2) = O(1/T)$ and $O(E\zeta_{2T}) = O(1/T)$, which follows from $E(c'\Delta_\beta\Delta'_\beta c) = O(1/T)$ for any constant c . Next, let $f_{e_\beta}(x)$ be the pdf of e_β . Since e_β is independent of $\hat{\Omega}$ and Δ_β , we have

$$\begin{aligned} &EF'_1(0, 0, \Xi) \zeta_{1T} \\ &= \int E \left[F'_1(0, 0, \Xi) \zeta_{1T} | e_\beta = x \right] f_{e_\beta}(x) dx \\ &= \int E \left[F'_1 \left(0, 0, x' (R\Omega R')^{1/2} \left(R\hat{\Omega}R' \right)^{-1} (R\Omega R')^{1/2} x' \right) \right. \\ &\quad \left. \times 2 \left(RT^{3/2} \Delta_\beta \right)' \left(R12\hat{\Omega}R' \right)^{-1} \left(R12\Omega_{T,GLS}R' \right)^{1/2} x \right] f_{e_\beta}(x) dx. \end{aligned}$$

Writing $\hat{\Omega} = \hat{\Omega}(u)$ and $\Delta_\beta = \Delta_\beta(u)$ as functions of the error process u , we have $\hat{\Omega}(u) = \hat{\Omega}(-u)$ and $\Delta_\beta(u) = -\Delta_\beta(-u)$. Hence

$$\begin{aligned} & E \left[F_1' \left(0, 0, x' (R\Omega R')^{1/2} (R\hat{\Omega}R')^{-1} (R\Omega R')^{1/2} x' \right) \right. \\ & \quad \left. \times 2 \left(RT^{3/2} \Delta_\beta \right)' (R12\hat{\Omega}R')^{-1} (R12\Omega_{T,GLS}R')^{1/2} x \right] \\ & = 0 \text{ for all } x. \end{aligned}$$

As a result,

$$EF_1'(0, 0, \Xi) \zeta_{1T} = 0.$$

So

$$EF(\zeta_{1T}, \zeta_{2T}, \Xi) = EF(0, 0, \Xi) + O\left(\frac{1}{T}\right).$$

We have therefore shown that

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

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 $\hat{\Omega}$:

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

Hence

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

Evaluating the above derivative at Ω yields:

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

Next, we compute the second order derivative:

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

where

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

Furthermore

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

Therefore

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

Now

$$\left[\Xi \left(\hat{\Omega} \right) \right]^{-1} = 1 + L + Q + \text{remainder} \quad (30)$$

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

$$EX_1^2 = \int_{-1}^1 x^2 f_{X_1}(x) dx = \frac{1}{2} \frac{\Gamma(\frac{1}{2}p)}{\Gamma(\frac{1}{2}p+1)} = \frac{1}{p}, \quad (31)$$

$$EX_1^4 = \int_{-1}^1 x^4 f_{X_1}(x) dx = \frac{3}{4} \frac{\Gamma(\frac{1}{2}p)}{\Gamma(\frac{1}{2}p+2)} = \frac{3}{p(p+2)}. \quad (32)$$

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

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

which implies that

$$EX_1^2X_2^2 = \frac{1}{p(p+2)}. \quad (33)$$

Using (31) and $EX_1X_2 = 0$, we have

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

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

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

where $A = (A_{ij}) = (R\Omega R')^{-1/2} R (\hat{\Omega} - \Omega) R' (R\Omega R')^{-1/2}$. Plugging in (32) and (33), we have

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

Now

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

The trace in the last line is the sum of

$$\begin{aligned}
& tr \left\{ \left(\left[R' (R\Omega R')^{-1} R \right] \otimes \left[R' (R\Omega R')^{-1} R \right] \right) (\Omega \otimes \Omega) \mathbb{I}_{n^2} \right\} \\
&= tr \left\{ \left[R' (R\Omega R')^{-1} R\Omega \right] \otimes \left[R' (R\Omega R')^{-1} R\Omega \right] \right\} \\
&= tr \left(R' (R\Omega R')^{-1} R\Omega \right) tr \left(R' (R\Omega R')^{-1} R\Omega \right) \\
&= p^2
\end{aligned}$$

and

$$\begin{aligned}
& tr \left\{ \left(\left[R' (R\Omega R')^{-1} R \right] \otimes \left[R' (R\Omega R')^{-1} R \right] \right) (\Omega \otimes \Omega) \mathbb{K}_{nn} \right\} \\
&= tr \left\{ \left(\left[R' (R\Omega R')^{-1} R \right] \otimes \left[R' (R\Omega R')^{-1} R \right] \right) (\Omega \otimes \Omega) \sum_{i=1}^n \sum_{j=1}^n (e_i e_j') \otimes (e_i e_j') \right\} \\
&= \sum_{i=1}^n \sum_{j=1}^n tr \left((R\Omega R')^{-1} R\Omega e_i e_j' R' \right) tr \left(R' (R\Omega R')^{-1} \Omega e_i e_j' R \right) \\
&= \sum_{i=1}^n \sum_{j=1}^n e_j' R' (R\Omega R')^{-1} R\Omega e_i e_i' R' (R\Omega R')^{-1} R\Omega e_j \\
&= \sum_{i=1}^n \sum_{j=1}^n \left(\left[R' (R\Omega R')^{-1} R\Omega \right]_{ij} \right)^2 \\
&= tr \left(R' (R\Omega R')^{-1} R\Omega R' (R\Omega R')^{-1} R\Omega \right) = p
\end{aligned}$$

where we use the representation that $\mathbb{K}_{nn} = \sum_{i=1}^n \sum_{j=1}^n (e_i e_j') \otimes (e_i e_j')$ with e_i being the i -th column unit vector of order n . See Magnus and Neudecker (1979, definition 3.1). Therefore

$$E tr(AA) = E vec(A)' vec(A) = \frac{1}{K} (p^2 + p) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right).$$

Next, since $\mathbb{K}_{nn} vec(A) = vec(A')$, we have

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

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

Therefore

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

It is easy to show that

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

Hence

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

and

$$\left[\Xi(\hat{\Omega}) \right]^{-1} = 1 + L + Q + o_p \left(\frac{K^2}{T^2} \right) + O_p \left(\frac{1}{T} \right). \quad (34)$$

Using the above expansion, we have

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

But

$$\begin{aligned}
EL^2 &= \frac{1}{K} E \left(\left[e'_\beta (R\Omega R')^{-1/2} R \right] \otimes \left[e'_\beta (R\Omega R')^{-1/2} R \right] (\Omega \otimes \Omega) (\mathbb{I}_{n^2} + \mathbb{K}_{nn}) \right) \\
&\quad \times \left[e'_\beta (R\Omega R')^{-1/2} R \right]' \otimes \left[e'_\beta (R\Omega R')^{-1/2} R \right]' + o \left(\frac{1}{K} + \frac{K^2}{T^2} \right) + O \left(\frac{1}{T} \right). \\
&= \frac{2}{K} E [e'_\beta e_\beta]^2 + o \left(\frac{1}{K} + \frac{K^2}{T^2} \right) + O \left(\frac{1}{T} \right) \\
&= \frac{2}{K} + o \left(\frac{1}{K} + \frac{K^2}{T^2} \right) + O \left(\frac{1}{T} \right)
\end{aligned}$$

so

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

as desired. ■

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

$$\begin{aligned}
&P (F_{T,OLS}^* > pF_{p,K-p+1}^\alpha) - \alpha \\
&= -\frac{K^2}{T^2} G'_p(pF_{p,K-p+1}^\alpha) pF_{p,K-p+1}^\alpha \bar{B} + o \left(\frac{1}{K} \right) + o \left(\frac{K^2}{T^2} \right) + O \left(\frac{1}{T} \right). \quad (35)
\end{aligned}$$

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

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

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

$$e_{\beta\delta} = \frac{(R12\Omega_{T,GLS}R')^{-1/2} RT^{3/2}(\hat{\beta}_{GLS} - \beta) + \tilde{c}}{\left\| (R12\Omega_{T,GLS}R')^{-1/2} RT^{3/2}(\hat{\beta}_{GLS} - \beta) + \tilde{c} \right\|},$$

then

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

where

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

and Θ_δ is independent of Ξ_δ . 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,

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

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

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

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

$$G'_{p\delta^2}(z) = \sum_{j=0}^{\infty} \frac{(\delta^2/2)^j}{j!} e^{-\delta^2/2} \frac{1}{2^{j+p/2}\Gamma(j+p/2)} z^{j+p/2-1} e^{-z/2}, \quad (36)$$

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

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

Some algebra shows that

$$\mathcal{Q}_{p,\delta^2}(z) = \frac{\delta^2}{2z} G'_{(p+2),\delta^2}(z)$$

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] Beveridge, S., and C. R. Nelson (1981): “A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle,” *Journal of Monetary Economics*, 7, pp. 151–174.
- [3] Bilodeau, M. and D. Brenner (1999): *Theory of Multivariate Statistics*, Springer.
- [4] Foley, R.D. and Goldman, D. (1999): “Confidence Intervals Using Orthonormally Weighted Standardized Time Series.” *ACM Transactions on Modeling and Computer Simulation*, 19(4), 297-325.
- [5] Grenander, U. and M. Rosenblatt (1957): *Statistical Analysis of Stationary Time Series*. New York: Wiley.

- [6] 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.
- [7] Hannan, E. J. (1957): “The Variance of the Mean of a Stationary Process,” *Journal of the Royal Statistical Society, Series B*, Vol. 19, No. 2, pp. 282–285.
- [8] Hotelling H. (1931): “The Generalization of Student’s Ratio,” *Annals of Mathematical Statistics*, 2, 360–378.
- [9] Jansson, M. (2004): “On the Error of Rejection Probability in Simple Autocorrelation Robust Tests,” *Econometrica*, 72, 937–946.
- [10] 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.
- [11] Kiefer, N. M. and T. J. Vogelsang (2002a): “Heteroskedasticity-autocorrelation Robust Testing Using Bandwidth Equal to Sample Size,” *Econometric Theory*, 18, 1350–1366.
- [12] ——— (2002b): “Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation,” *Econometrica*, 70, 2093–2095.
- [13] ——— (2005): “A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests,” *Econometric Theory*, 21, 1130–1164.
- [14] Magnus, J.R. and H. Neudecker (1979): “The Commutation Matrix: Some Properties and Applications,” *The Annals of Statistics*, 7(2), 381–394.
- [15] Mercer, J. (1909): “Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations,” *Philosophical Transactions of the Royal Society of London, Series A*, 209, 415–446.
- [16] Müller, U.K. (2007): “A Theory of Robust Long-Run Variance Estimation,” *Journal of Econometrics*, 141(2), 1331–1352
- [17] Newey, W. K. and K. D. West (1987): “A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703–708.
- [18] ——— (1994): “Automatic Lag Selection in Covariance Estimation,” *Review of Economic Studies*, 61, 631–654.
- [19] Ng, S. and Perron, P. (1994): “The Exact Error in Estimating of the Spectral Density at the Origin,” *Journal of Time Series Analysis*, 17(4), 379–408.
- [20] Phillips, P. C. B. (2005): “HAC Estimation by Automated Regression,” *Econometric Theory*, 21, 116–142.
- [21] Phillips, P. C. B. and V. Solo (1992): “Asymptotics for Linear Processes,” *Annals of Statistics*, 20, 971–1001
- [22] 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.

- [23] ——— (2007): “Long-Run Variance Estimation and Robust Regression Using Sharp Origin Kernels with No Truncation,” *Journal of Statistical Planning and Inference*, 137, 985–1023.
- [24] Percival D. and A. T. Walden (1993): *Spectral Analysis For Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge University Press.
- [25] Priestley, M. B. (1981): *Spectral Analysis and Time Series*, New York: Academic Press.
- [26] Stoica, P. and Moses, R. (2005): *Spectral Analysis of Signals*, Pearson Prentice Hall.
- [27] 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.
- [28] Sun, Y. (2010): “Let’s Fix It: Fixed- b Asymptotics versus Small- b Asymptotics in Heteroscedasticity and Autocorrelation Robust Inference.” Working paper, Department of Economics, UC San Diego.
- [29] Sun, Y., P. C. B. Phillips, and S. Jin (2008): “Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing,” *Econometrica*, 76, 175–194.
- [30] Sun, Y., P. C. B. Phillips, and S. Jin (2010): “Power Maximization and Size Control in Heteroskedasticity and Autocorrelation Robust Tests with Exponentiated Kernels,” *Econometric Theory*, forthcoming.
- [31] Sun, Y. and P. C. B. Phillips (2009): “Bandwidth Choice for Interval Estimation in GMM Regression,” Working paper, Department of Economics, UC San Diego.
- [32] Thomson, D. J. (1982): “Spectrum Estimation and Harmonic Analysis,” *IEEE Proceedings*, 70, 1055–1096.
- [33] Vogelsang, T. J. and P. H. Franses (2005): “Testing for Common Deterministic Trend Slopes,” *Journal of Econometrics*, 126, 1–24.