Autocorrelation Robust Inference Using Series Variance Estimator and Testing-optimal Smoothing Parameter

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

First version: June 2009; This version: February 2011

*Email: yisun@ucsd.edu. The author gratefully acknowledges partial research support from NSF under Grant No. SES-0752443. He thanks Xiaohong Chen, Chung-ming Kuan, Peter Phillips, participants at the 2010 California Econometrics Conference at Stanford, a coeditor, and two anonymous referees for helpful comments and suggestions. Correspondence to: Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508.

ABSTRACT

The paper develops a new procedure for hypothesis testing in time series models. The test is based on a series long run variance matrix estimator that involves projecting the time series onto a set of orthonormal bases and using the sample variance of the projection vectors as the variance estimator. The series long run variance estimator is asymptotically invariant to model parameters and thus does not suffer from the usual estimation bias that hurts the performance of conventional kernel estimators. The number of basis functions K, the underlying smoothing parameter, plays a key role in determining the asymptotic properties of the series long run variance estimator and the associated semiparametric test. When K is fixed, the (modified) Wald statistic converges to an F-distribution while when K grows with the sample size, the Wald statistic converges to a chi-square distribution. We show that critical values from the fixed-K asymptotics are second order correct under the large-K asymptotics. We propose a new approach to select K which minimizes the type II error hence maximizes the power of the test while controlling for the type I error. This testing-oriented selection rule is fundamentally different from the conventional rule based on the mean square error criterion. A plug-in procedure for implementing the new rule is suggested and simulations show that the new plug-in procedure works remarkably well in finite samples.

JEL Classification: C13; C14; C32; C51

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

1 Introduction

One objective of time series analysis is to estimate some unknown characteristic or parameter of the system being studied. We often want not only an estimate of this parameter value, but also some measure of the estimator's precision in order to conduct inference. Hypothesis testing is widely used for this purpose. In this paper, we first consider a multivariate time series whose mean value is the parameter of interest. We use this simple model to illustrate our ideas and then discuss how the basic ideas can be extended to more general settings. Nevertheless, the mean inference problem includes a large number of situations. For example, the data might consist of a multivariate time series of forecasting loss that are produced by different forecasting methods. We can test equal predictive accuracy of these forecasting methods by examining whether the loss differential series has mean zero. Diebold and Mariano (1995) consider the case with two forecasts while Christensen, Diebold, Rudebusch and Strasser (2008) extend it to general multivariate scenarios. There is also a large and active literature on inference for the mean of simulated time series. Various variance estimation methods have been proposed in the fields of operation research and industrial engineering. Some methods are familiar to statisticians and econometricians, for example, the spectrum method of Heidelberger and Welch (1981) and the autoregression approximation method of Fishman (1978). Some methods are less familiar, for example, the batch mean method, discussed as early as Conway (1963). These procedures all rely on central limit theorems that describe the asymptotic behavior of the parameter estimator. The methods differ in the manner of estimating the standard deviation of this estimator. A major difficulty in performing hypothesis testing for dependent stationary time series comes in estimating the value of this unknown scaling constant. In the econometrics literature, this scaling constant is also referred to as the long run variance (LRV).

In this paper, we follow Phillips (2005) and consider estimating the LRV using a nonparametric series method. The basic idea is to project the time series onto a set of basis functions designed to represent the long-run behavior directly. The outer-product of each projection coefficient is a direct and asymptotically unbiased estimator of the LRV. The series LRV estimator is simply an average of these direct estimators. By construction, the series LRV estimator is automatically positive semidefinite, a desirable property for practical use of the LRV estimator.

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 Wald statistic converges to a nonstandard distribution rather than the standard chi-square distribution. This type of asymptotics captures the randomness of the LRV estimator and delivers a test that is more accurate in size than the conventional chi-square test. The fixed-K asymptotics has been widely used in the literature on computer simulations, see, for example, Foley and Goldman (1999). It is also in the spirit of the fixed-b asymptotics of Kiefer and Vogelsang (2005).

The novelty here is that we design a set of basis functions so that the nonstandard limiting distribution becomes a standard F distribution. We require the basis functions to be orthonormal and have zero mean in that they integrate to zero on [0,1]. These two conditions ensure that the direct LRV estimators are asymptotically independent and the series LRV estimator converges to a scaled Wishart distribution. As a result, a modified Wald statistic converges to an F-distribution. So a great advantage of using the series LRV estimator is that the critical values from the fixed-K asymptotics are readily available from statistical tables and software programs. The computational burden of simulating critical values from nonstandard distributions is completely removed. The zero mean condition ensures that the series LRV estimator is asymptotically invariant to the model parameters of interest. As a result, it does not suffer from the bias due to the estimation uncertainty of these model parameters. This is in contrast with the conventional kernel LRV estimators where this type of bias is often present. See for example, Hannan (1957) and Ng and Perron (1994). Although carefully crafted polynomials meet the orthonormality and zero mean conditions, it is more convenient to use cosine and sine functions. In this paper, we use the sine bases as they do not suffer from the bias due to the edge effect.

The challenge is how to select the number of basis terms, an important tuning parameter that determines the asymptotic properties of the series LRV estimator. It turns out that the fixed-K asymptotics does not provide an internally consistent framework for selecting the optimal number of bases. For a fixed value of K, the series LRV estimator has a bias of order $O(1/T)$ while its variance is $O(1)$ and decreases with K. To minimize the variance, we should select the value of K that is as large as possible. However, the (absolute) asymptotics bias also increases with K. The fundamental problem is that the bias order of $O(1/T)$ obtained under the fixed-K asymptotics is not uniform across K. To overcome this problem, we consider the asymptotic behavior of the series LRV estimator when K grows with the sample size at a certain rate. Following the conventional approach (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. However, the MSE-optimal choice is not necessarily best suited for semiparametric testing.

To develop an optimal choice of K for semiparametric testing, we first have to decide on which test to use. There are two choices. One is the traditional Wald test which is based the Wald statistic and uses critical values from a chi-square distribution. The other is the new F^* test given in this paper, which is based on a modified Wald statistic and uses critical values from an F distribution. The modification involves multiplying the Wald statistic by a finite sample correction factor $(K-p+1)/K$ where p is the number of restrictions being tested. The correction factor can be large when K is small or when p is large. One of main contributions of the paper is to show that critical values from the fixed-K asymptotics are higher order correct under the conventional large-K asymptotics. This implies that the F^* test generally is more accurate in size than the traditional Wald test. On the basis of this theoretical result and the emphasis on the size control in the statistics and econometrics literature, we employ the F^* test to conduct inference on the mean of the time series.

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

To select the testing-optimal K, we use high order asymptotic expansions to obtain approximate measures of type I and type II errors of the F^* test. We show that the type I error depends on the nonparametric bias of the LRV estimator. The type II error of the F^* test depends on the local alternative hypothesis through the noncentrality parameter $\|\tilde{c}\|^2$, where $\tilde{c}$ is a vector that characterizes the local departure of the alternative hypothesis from the null. To the first order, the type II error depends on $\tilde{c}$ only through its squared length $\|\tilde{c}\|^2$. So it is reasonable to assume that $\tilde{c}$ is uniformly distributed on a sphere. This assumption greatly facilitates the higher order expansion under the local alternative hypothesis. In a transformed space, the null hypothesis is a fixed point while the alternative hypothesis is a random point uniformly distributed on a sphere centered at the fixed null. We choose the radius of the sphere so that the power of the test is 50% under the first order asymptotics. This strategy is similar to that used in the optimal testing literature, see for example, Elliott, Rothenberg and Stock (1996).

The paper contributes to a large literature on semiparametric testing for time series models. In particular, we provide a rigorous framework for optimal smoothing parameter choice for mean inference. In the fields of operation research, industrial engineering, simulation, statistics and econometrics, various methods have been proposed for robust mean inference. See, for examples, Conway (1963), Albers (1978), Fishman (1978), Heidelberger and Welch (1981), Kabaila and Nelson (1985), Foley and Goldman (1999), Alexopoulos (2007) and references therein. The fundamental problem is how to select the smoothing parameter in the nonparametric variance estimator so that the associated test has good size and power properties. Existing proposals are either ad hoc or based on the MSE criterion. The present paper proposes a new practical procedure for selecting the smoothing parameter that addresses the central concern of hypothesis testing.

Some discussions of this paper's contributions relative to the author's other work are in order. Sun (2011a) employs the series LRV estimator and investigates the testing-optimal smoothing parameter choice for autocorrelation-robust trend inference, while the focus here is on stationary time series. The idea of testing-optimal smoothing parameter choice first appears in Sun, Phillips and Jin (2008) where a simple univariate Gaussian location model is considered. Subsequently, Sun and Phillips (2008) extends Sun, Phillips, and Jin (2008) to linear IV regressions and relaxes the Gaussian assumption but considers optimal bandwidth that minimizes the coverage error of confidence intervals for a single parameter or a single linear parameter combination. Sun (2011b) employs the idea of the F approximation in this paper to tackle autocorrelation robust inference based on kernel LRV estimators.

The rest of the paper is organized as follows. Section 2 describes the problem at hand. Section 3 discusses the series LRV estimator and its relationship with other popular estimators. The section also establishes the asymptotic properties of the series LRV estimator under the fixed-K and large-K asymptotics. Section 4 investigates the Wald test under both the fixed-K and large-K asymptotics. Section 5 gives a higher order expansion of the finite sample distribution of the modified Wald statistic. On the basis of this expansion, the next section proposes a selection rule for K that is most suitable for implementation in semiparametric testing. Section 7 discusses the applicability of our procedure for location tests to more general testing problems. The subsequent section reports simulation evidence on the performance of the new procedure. The last section provides some concluding discussion. Proofs of the main results and their extensions are given in the Appendix.

2 The Model and Preliminaries

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

$$y_t = \theta + u_t, t = 1, 2, ..., T \tag{1}$$

where $y_t = (y_{1t}, ..., y_{nt})'$, $\theta = (\theta_1, ..., \theta_n)'$, $u_t = (u_{1t}, ..., u_{nt})'$ is a weakly dependent process with zero mean. Our focus of interest is on inference about the mean θ.

Assumption 1 *We assume that*

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

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

$$\sum_{j=0}^{\infty} j^a \|C_j\| < \infty \text{ for } a > 3,\ C(1)\Sigma C(1)' > 0 \tag{2}$$

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

The summability assumption in (2) ensures that

$$\sum_{h=-\infty}^{\infty} |h|^3 \|\Gamma_u(h)\| < \infty$$

where $\Gamma_u(h) = Eu_t u_{t+h}'$. The convergence result, which is helpful in some technical derivations below, means that the spectral density matrix of u_t has continuous second order derivatives.

Under Assumption 1, we can prove that

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

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

$$\Omega = \Lambda\Lambda' = \sum_{j=-\infty}^{\infty} Eu_t u_{t-j}' = \sum_{j=-\infty}^{\infty} \Gamma_u(j).$$

The OLS estimator of θ is the average of $\{y_t\}$, viz. $\hat{\theta}_{OLS} = T^{-1} \sum_{t=1}^{T} y_t$. It follows from (3) that

$$\sqrt{T}\left(\hat{\theta}_{OLS} - \theta\right) \to^d \Lambda W_n(1) :=^d N(0, \Omega).$$

Let $\hat{\theta}_{GLS}$ be the GLS estimator of θ given by

$$\hat{\theta}_{GLS} = \left[(\ell_T \otimes \mathbb{I}_n)' \, \Omega_u^{-1} \, (\ell_T \otimes \mathbb{I}_n)\right]^{-1} (\ell_T \otimes \mathbb{I}_n)' \, \Omega_u^{-1} y,$$

where $\Omega_u = var([u_1', u_2', ..., u_T']')$, $y = [y_1', y_2', ..., y_T']'$ and ℓ_T is a vector of ones. Under Assumption 1, the OLS estimator $\hat{\theta}_{OLS}$ is asymptotically equivalent to the GLS estimator $\hat{\theta}_{GLS}$. To see this, let c be any vector in $\mathbb{R}^n$, then $x_t = c'y_t$ is a univariate time series with mean $\mu = c'\theta$. According to Grenander and Rosenblatt (1957), the OLS estimator $\hat{\mu}_{OLS}$ of μ is asymptotically equivalent to the GLS $\hat{\mu}_{GLS}$ of μ. In addition, $var(\hat{\mu}_{OLS}) = var(\hat{\mu}_{GLS}) + O\left(1/T\right)$. Note that $\hat{\mu}_{OLS} = c'\hat{\theta}_{OLS}$, $\hat{\mu}_{GLS} = c'\hat{\theta}_{GLS}$, so

$$var(c'\hat{\theta}_{OLS}) = var(c'\hat{\theta}_{GLS}) + O\left(1/T\right) \text{ for any } c \in \mathbb{R}^n.$$

This is to say, $\hat{\theta}_{OLS}$ and $\hat{\theta}_{GLS}$ are asymptotically equivalent. Thus, the simple OLS estimator has a nice optimality property. We note in passing that the GLS estimator $\hat{\theta}_{GLS}$ will be employed as a technical device in later developments.

The hypotheses of interest in this paper are

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

where R is a $p \times n$ matrix and r is a $p \times 1$ vector. For example, we may want to test whether the means are jointly zero. In this case, R is an identity matrix and r is a vector of zeros. The problem is the same as that considered by Hotelling (1931) except that we allow for time series dependence. As a second example, in the equal predictability test of Christensen, Diebold, Rudebusch and Strasser (2008), the null hypothesis is that the means of forecasting loss are the same across different forecasts.

3 Series LRV Estimator and its Asymptotic Properties

3.1 Series LRV Estimator

To conduct inference regarding θ, we need to first estimate the long run variance matrix Ω. Many nonparametric estimation methods are available in both statistics and econometrics literature. Most LRV estimators use kernel-smoothing methods that involve taking a weighted sum of sample autocovariances. In this paper, we consider a nonparametric series method which involves projecting the time series onto orthogonal functions.

Let $\{\phi_k\}_{k=1}^{\infty}$ be a sequence of basis functions in $L_2[0,1]$ and $\hat{u}_t = y_t - \hat{\theta}_{OLS}$. Define the inner product

$$\hat{\Lambda}_k = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \phi_k(\frac{t}{T}) \hat{u}_t$$

and construct the direct LRV estimator:

$$\hat{\Omega}_k = \hat{\Lambda}_k \hat{\Lambda}_k'$$

for each $k = 1, 2, ..., K$. Taking a simple average of these direct estimators yields our series estimator:

$$\hat{\Omega} = \frac{1}{K} \sum_{k=1}^{K} \hat{\Omega}_k$$

where K is the number of basis functions used.

The series estimator has different interpretations. First, it can be regarded as a multiple-window estimator with window function $\phi_k(t/T)$, see Thomson (1982). Depending on the choice of ϕ_k, the estimator $\hat{\Omega}$ includes many existing estimators as special cases. Percival and Walden (1993, page 353) point out that all quadratic estimators with a real valued, symmetric, positive semidefinite matrix of weights can be written in the form of $\hat{\Omega}$. Examples of such quadratic estimators are the kernel LRV estimators, which are used widely in the econometric literature (e.g. Andrews (1991) and Newey and West (1987)). Quadratic estimators that cannot be written in the kernel form have been considered by Phillips (2005), Müller (2007) and Sun (2006) in the econometric literature. Second, when $\phi_k(1-x) = \phi_k(x)$, we can write $\hat{\Lambda}_k = (1/\sqrt{T})\sum_{\tau=0}^{T-1}\phi_k(\tau/T)\hat{u}_{T-\tau}$, which is a linear filter of the residual process $\hat{u}_t$. The transfer function of the linear filter for each k is

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

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. For more discussions on filter-bank estimators, see Stoica and Moses (2005). Finally, when $\phi_k(\cdot)$ are orthonormal in finite samples so that $T^{-1}\sum_{t=1}^{T}\phi_k(t/T)\phi_{k'}(t/T) = \delta_{k,k'}$, $\hat{\Lambda}_k$ is the vector of projection coefficients of projecting $\hat{u}_t$ onto the basis function ϕ_k. There exist basis functions such that $T^{-1}\sum_{t=1}^{T}\phi_k(t/T)\phi_{k'}(t/t) = \int_0^1\phi_k(r)\phi_{k'}(r)dr = \delta_{k,k'}$. Following and extending Phillips (2005), we can cast our estimator in the seemingly unrelated regression system below:

$$\begin{aligned}\hat{u}_{t1} &= \sum_{k=1}^{K}\frac{1}{\sqrt{T}}\phi_k(\frac{t}{T})\Lambda_{k1} + e_{t1},\\ &\dots\\ \hat{u}_{tn} &= \sum_{k=1}^{K}\frac{1}{\sqrt{T}}\phi_k(\frac{t}{T})\Lambda_{kn} + e_{tn}.\end{aligned}$$

In the vector form, the system becomes

$$\hat{u}_t = \sum_{k=1}^{K}\frac{1}{\sqrt{T}}\phi_k(\frac{t}{T})\Lambda_k + e_t.$$

$\hat{\Lambda}_k$ is simply the OLS estimator of Λ_k and $\hat{\Omega}$ is the sample variance of $\hat{\Lambda}_k$. $\hat{\Omega}_k$ 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 $\phi_k(\cdot)$. 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 variation in the series. A closely related paper along this line is Phillips (2006), which discusses relations of basis function series projection estimators to kernel procedures, and how the series projections can be used to optimally estimate cointegration systems.

The series estimator $\hat{\Omega}$ can also be written as the so-called area estimator in industrial and system engineering. Let $S_0 = 0$ and

$$S_t = \frac{1}{\sqrt{T}} \sum_{\tau=1}^{t} \hat{u}_\tau = \frac{1}{\sqrt{T}} \sum_{\tau=1}^{t} (u_\tau - \bar{u}) \text{ for } t = 1, 2, ..., T$$

then using summation by parts, we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \phi_k \left(\frac{t}{T}\right) \hat{u}_t := \frac{1}{T} \sum_{t=1}^{T} w_k \left(\frac{t}{T}\right) S_t \tag{4}$$

where

$$w_k \left(\frac{t}{T}\right) = T \left[\phi_k \left(\frac{t}{T}\right) - \phi_k \left(\frac{t+1}{T}\right)\right] \sim \phi_k' \left(\frac{t}{T}\right)$$

and $\phi_k ([T+1]/T)$ can be arbitrarily defined as $S_T = 0$. So

$$\hat{\Omega} = \frac{1}{K} \sum_{k=1}^{K} \left[\frac{1}{T} \sum_{t=1}^{T} w_k \left(\frac{t}{T}\right) S_t\right] \left[\frac{1}{T} \sum_{t=1}^{T} w_k \left(\frac{t}{T}\right) S_t\right]'.$$

For a given weighting function, the estimator formulated as above is called the weighted area estimator (Foley and Goldsman (1999)). This terminology originates from the observation that

$$\frac{1}{T} \sum_{t=1}^{T} w_k \left(\frac{t}{T}\right) S_t \to^d \int_0^1 w_k (r) V_n(r) dr$$

which is the weighted area of the Brownian bridge $V_n(r) = W_n(r) - rW_n(1)$.

Finally, the series estimator can be written as

$$\hat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \hat{u}_t \mathcal{K}_G(\frac{t}{T}, \frac{s}{T}) \hat{u}_s'$$

where

$$\mathcal{K}_G(r, s) = \frac{1}{K} \sum_{k=1}^{K} \phi_k (r) \phi_k (s).$$

So the series estimator $\hat{\Omega}$ can be regarded as a kernel LRV estimator with the generalized kernel function $\mathcal{K}_G(r, s)$. For regular kernel estimators, the kernel function satisfies $\mathcal{K}(r, s) = \mathcal{K}(r - s)$. Here for any finite K, $\mathcal{K}_G(r, s) \neq \mathcal{K}_G(r - s)$ in general.

By selecting different basis functions, we can obtain many series LRV estimators. However, in nonparametric series estimation, it is the 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

To obtain the asymptotic distribution $\hat{\Omega}$ for a fixed K, we assume that $\phi_k(r)$ is continuously differentiable. Under this assumption and Assumption 1, we can prove Theorem 1.

Theorem 1 *Let Assumption 1 hold. If $\phi_k(r)$ is continuously differentiable, then for a fixed K,*

$$\hat{\Omega} \to^d \Lambda \left(\frac{1}{K} \sum_{k=1}^{K} \zeta_k \zeta_k' \right) \Lambda', \tag{5}$$

where

$$\zeta_k = \int_0^1 \left(\phi_k(r) - \int_0^1 \phi_k(s) ds \right) dW_n(r).$$

Assumption 1 is sufficient but not necessary for Theorem 1 to hold. All we need is the FCLT in (3).

In general, the limiting distribution is nonstandard, which is not very convenient for practical use. To simplify the asymptotic distribution, we make the following assumption:

Assumption 2 *(i)* $\int_0^1 \phi_k(r)\, dr = 0$; *(ii)* $\int_0^1 \phi_{k_1}(r) \phi_{k_2}(r) dr = \delta_{k_1,k_2} := 1\{k_1 = k_2\}$.

The "zero mean" assumption $\int_0^1 \phi_k(r)\, dr = 0$ removes the second part in the definition of ζ_k so that $\zeta_k = \int_0^1 \phi_k(r)\, dW_n(r)$. Note that the second part in ζ_k reflects the estimation error in $\hat{\theta}$. So the zero mean assumption ensures that our LRV estimator is not affected by the estimation error in the mean. This is an important point. For conventional LRV estimators, Hannan (1957) points out that mean correction can lead to large bias in LRV estimation. Based on the MSE criterion, this bias is of smaller order than the nonparametric smoothing bias. However for robust testing, this type of bias is of the same order of magnitude as the asymptotic variance; see SPJ (2008). Ignoring this type of bias can lead to severe size distortion in finite samples.

The orthonormality assumption ensures that ζ_k is iid $N(0, \mathbb{I}_n)$, where $\mathbb{I}_n$ is the $n \times n$ identity matrix. That is, each of $\hat{\Omega}_k$ is asymptotically unbiased and $\hat{\Omega}_{k_1}$ and $\hat{\Omega}_{k_2}$ for $k_1 \neq k_2$ are asymptotically independent. This property is analogous to the well-known property for periodogram of a stationary time series. Under this assumption and for a fixed K, $\hat{\Omega}$ converges to a Wishart type distribution as shown below.

Corollary 2 *Let Assumptions 1 and 2 hold. If $\phi_k(r)$ is continuously differentiable, then for a fixed K,*

$$K\Lambda^{-1}\hat{\Omega}\left(\Lambda'\right)^{-1} \to^d \mathbb{W}_n(\mathbb{I}_n, K) \text{ a Wishart distribution.}$$

In the scalar case with $n = 1$, the limiting distribution reduces to the chi-square distribution χ^2_K. In addition, for any conforming constant vector z, we have

$$\frac{z'\hat{\Omega}z}{z'\Omega z} = \frac{z'\Lambda\left[\Lambda^{-1}\hat{\Omega}\left(\Lambda'\right)^{-1}\right]\Lambda' z}{z'\Lambda\Lambda^{-1}\Omega\left(\Lambda'\right)^{-1}\Lambda' z} \to^d \frac{\tilde{z}'}{\|\tilde{z}\|}\left(\frac{1}{K}\sum_{k=1}^{K}\zeta_k\zeta_k'\right)\frac{\tilde{z}}{\|\tilde{z}\|} \sim^d \frac{\chi^2_K}{K},$$

where $\tilde{z} = \Lambda' z$ and $\zeta_k \sim iidN(0, \mathbb{I}_n)$. That is, $z'\hat{\Omega}z/z'\Omega z$ converges to a scaled chi-square distribution. This result can be used to test hypotheses regarding Ω. The resulting test may have better size properties than the asymptotic chi-square test. See Phillips, Sun and Jin (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 θ.

For convenience, we call the above asymptotics the fixed-K asymptotics. This type of asymptotics has been used widely in the operation research and simulation literature. See for example, Foley and Goldman (1999), Alexopoulos (2007). In fact, it is the only type of asymptotics considered in that literature. The fixed-K asymptotics is similar in spirit to the fixed-b asymptotics of Kiefer and Vogelsang (2005) who consider conventional kernel LRV estimators and assume that the bandwidth is equal to a fixed proportion of the sample size. The fixed-b asymptotics can be traced back to Neave (1970).

3.3 Large-K Asymptotics

In this section, we consider the asymptotic properties of $\hat{\Omega}$ when both K and T go to infinity such that $K/T \to 0$. We focus the special case $\phi_k(r) = \sqrt{2}\sin 2\pi kr$ as the results for other cases can be proved analogously. More importantly, this series of basis functions satisfies $\phi_k(0) = \phi_k(1) = 0$, a condition that ensures that the LRV estimator puts relatively less weight on the boundary points and is highly effective in removing the edge effect. As a result, the asymptotic bias of order $O(1/T)$ vanishes. $\phi_k(r) = \sqrt{2}\sin 2\pi kr$ is also consistent with the weight function $w_k(r) = \sqrt{8}\pi\cos(2\pi kr)$ used in the weighted area estimator of Foley and Goldsman (1999). According to the relationship between $w_k(x)$ in that paper and $\phi_k(x)$ in the present setting, we have

$$\phi_k(r) = \int_0^r w_k(s)\,ds = \sqrt{8}\pi\int_0^r \cos(2\pi s)\,ds = \sqrt{2}\sin 2\pi kr,$$

which is exactly the sine function. We use the sine bases throughout the rest of the paper.

Theorem 3 *Let Assumption 1 hold. If $\phi_k(r) = \sqrt{2}\sin 2\pi kr$, then*
(a) The bias of $\hat{\Omega}$ is

$$E\hat{\Omega} - \Omega = \frac{K^2}{T^2}B + o\left(\frac{K^2}{T^2}\right) + o\left(\frac{1}{T}\right) \text{ for } B = -\frac{2\pi^2}{3}\sum_{h=-\infty}^{\infty} h^2\Gamma_u(h). \tag{6}$$

(b) The variance of $\hat{\Omega}$ is

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

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

The theorem can be proved using the same arguments in Phillips (2005, Theorem 1). We omit the details to conserve space.

Let

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

be the mean squared error of $vec(\hat{\Omega})$ with weighting matrix $\mathcal{W}$. It follows from Theorem 3 that, up to smaller order terms:

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

So the MSE optimal K is given by

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

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

4 Autocorrelation Robust Inference

The usual Wald statistic $F_{T,OLS}$ for testing H_0 against H_1 is given by

$$F_{T,OLS} = \left[\sqrt{T}(R\hat{\theta}_{OLS}-r)\right]'\left(R\hat{\Omega}R'\right)^{-1}\left[\sqrt{T}(R\hat{\theta}_{OLS}-r)\right]. \tag{7}$$

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

$$t_{T,OLS} = \frac{\sqrt{T}(R\hat{\theta}_{OLS}-r)}{\left(R\hat{\Omega}R'\right)^{1/2}}.$$

4.1 Fixed-K Asymptotics

Under the fixed-K asymptotics and the null hypothesis

$$\begin{aligned} F_{T,OLS} &= \left[R\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t\right]'\left(R\hat{\Omega}R'\right)^{-1}\left[R\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t\right] \\ &\to^d \left(R\Lambda W_n(1)\right)'\left\{R\Lambda\frac{1}{K}\sum_{k=1}^{K}\left[\int_0^1\phi_k(r)dW_n(r)\right]\left[\int_0^1\phi_k(s)dW_n'(s)\right]\Lambda'R'\right\}^{-1} \\ &\times\left(R\Lambda W_n(1)\right). \end{aligned}$$

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

$$\begin{aligned} F_{T,OLS} &\to^d W_p^*(1)'\left\{\frac{1}{K}\sum_{k=1}^{K}\left[\int_0^1\phi_k(r)\,dW_p^*(r)\right]\left[\int_0^1\phi_k(s)\,dW_p^*(s)'\right]\right\}^{-1}W_p^*(1) \\ &:= \eta'\left(\frac{1}{K}\sum_{k=1}^{K}\xi_k\xi_k'\right)^{-1}\eta, \end{aligned}$$

where

$$\eta = W_p^*(1) \text{ and } \xi_k = \int_0^1 \phi_k(r)\, dW_p^*(r).$$

Since

$$cov\left[W_p^*(1), \int_0^1 \phi_k(r)\, dW_p^*(r)\right] = \int_0^1 \phi_k(r)\, dr = 0 \text{ for all } k,$$

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

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

Using the well-known relationship between the T-squared distribution and the F-distribution, we have

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

In other words,

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

where χ_p^2 and χ_{K-p+1}^2 denote independent χ^2 random variables. Of course, for the above distribution to be well defined, we need to assume that $K \geq p$, a necessary condition to ensure that $R\hat{\Omega}R'$ is invertible. In general, we need to assume $K \geq n$, a necessary condition for the positive semi-definiteness of $\hat{\Omega}$.

When $p = 1$, the above result reduces to $t_T \to^d t_K$. That is, the t-statistic converges to the t-distribution with K degrees of freedom. The asymptotic t-distribution theory is not new in the literature. For the batch mean method, when the number of batches is fixed, the t-statistic also converges to a t-distribution, see Alexopoulos (2007). Ibragimov and Müller (2010) employ a closely related method and establish the robustness of t-approximation to variance heterogeneity.

We have therefore shown that when K is held fixed, the t-statistic converges to the t distribution with degrees of freedom K and the scaled Wald statistic converges to the F distribution with degrees of freedom p and $K-p+1$. These results are very convenient in practical situations as critical values from the t distribution or the F distribution can be easily obtained from statistical tables and packages.

Under the local alternative hypothesis,

$$H_1\left(\delta^2\right): R\theta = r + c/\left(\sqrt{T}\right) \text{ where } c = \left(R\Omega R'\right)^{1/2}\tilde{c} \tag{8}$$

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

$$\begin{aligned}\frac{(K-p+1)}{pK} F_{T,OLS} &\to^d \frac{(K-p+1)}{p}(\eta+\tilde{c})'\left(\sum_{k=1}^K \xi_k\xi_k'\right)^{-1}(\eta+\tilde{c}) \\ &=^d F_{p,K-p+1}\left(\delta^2\right),\end{aligned}$$

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

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

Similarly, the t-statistic converges to the noncentral t distribution with degrees of freedom K and noncentrality parameter $\delta = c/\left(R\Omega R'\right)^{1/2} = \tilde{c}$.

The local alternative power depends on c only through the noncentrality parameter $\delta^2 = \|\tilde{c}\|^2$, the squared length of $\tilde{c}$. The power is invariant to the direction of $\tilde{c}$. Hence, for the first order asymptotics given here, it is innocuous to assume that $\tilde{c}$ is uniformly distributed on the sphere $\mathfrak{S}_p(\delta) = \{x \in \mathbb{R}^p : \|x\|^2 = \delta^2\}$. As we show later, this assumption greatly simplifies the development of the higher order expansion in Section 5.

4.2 Large-K Asymptotics

When $K \to \infty$ such that $K/T \to 0$, the LRV estimator $\hat{\Omega}$ is consistent. In this case, we obtain the standard results:

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

When $p = 1$, the above results reduce to

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

Under the null hypothesis, the fixed-K asymptotics can be rewritten as

$$F_{T,OLS} \to^d \frac{\chi_p^2}{\chi_{K-p+1}^2/(K-p+1)} \frac{K}{K-p+1} =^d \left(pF_{p,K-p+1}\right) \frac{K}{K-p+1}.$$

Comparing it with the fixed-K asymptotics above, we find that the large-K asymptotics uses the fact that both $\chi_{K-p+1}^2/(K-p+1)$ and $K/(K-p+1)$ converge to one as $K \to \infty$. To the first order, the large-K asymptotics can be regarded as a sequential asymptotics where $T \to \infty$ for a fixed K and then $K \to \infty$.

Since both the random denominator $\chi_{K-p+1}^2/(K-p+1)$ and the proportional factor $K/(K-p+1)$ shift the probability mass to the right, critical values based on the fixed-K asymptotics are larger than those based on the large-K asymptotics. More rigorously, let $G_p(\cdot)$ be the CDF of a χ^2 random variable with degrees of freedom p and χ_p^α be the α-level critical value such that $1 - G_p\left(\chi_p^\alpha\right) = \alpha$. Then for typical α used in empirical applications:

$$\begin{aligned}
& P\left(\frac{\chi_p^2}{\chi_{K-p+1}^2/(K-p+1)} \frac{K}{K-p+1} > \chi_p^\alpha\right) \\
= \;& 1 - EG_p\left(\chi_p^\alpha \frac{K-p+1}{K} \frac{\chi_{K-p+1}^2}{(K-p+1)}\right) \\
> \;& 1 - G_p\left(\chi_p^\alpha \frac{K-p+1}{K}\right) > 1 - G_p\left(\chi_p^\alpha\right) = \alpha,
\end{aligned}$$

where we have used the concavity of $G_p(\cdot)$ at the right tail and Jensen's inequality. So critical values from the fixed-K asymptotics are indeed larger than the corresponding standard chi-square critical values.

5 High Order Expansion of the Finite Sample Distribution

In this section, we consider a high order expansion of the Wald statistic in order to design a procedure to select K. We follow SPJ (2008) and make the simplification assumption that u_t is normal. The assumption could be relaxed but at the cost of much greater complexity. See, for example, Velasco and Robinson (2001) and Sun and Phillips (2008).

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

$$F_{T,GLS} = \left[R\sqrt{T}(\hat{\theta}_{GLS} - \theta)\right]' \left(R\hat{\Omega}R'\right)^{-1} \left[R\sqrt{T}(\hat{\theta}_{GLS} - \theta)\right].$$

Define

$$\Delta = \hat{\theta}_{OLS} - \theta - (\hat{\theta}_{GLS} - \theta),$$

Then it can be shown that $\hat{\theta}_{GLS}$ is independent of Δ and $\hat{u}$. Using this independence result and the asymptotic equivalence of the OLS and GLS estimators, we can prove the following Lemma.

Lemma 1 *Let Assumption 1 hold and assume that $\varepsilon_t \sim iidN(0, \Sigma)$. Then*
(a) $P\left(F_{T,GLS} < z\right) = EG_p\left(z\Xi^{-1}\right) + O\left(\frac{1}{T}\right)$,
(b) $P\left(F_{T,OLS} < z\right) = P\left(F_{T,GLS} < z\right) + O\left(\frac{1}{T}\right)$,
where

$$\Xi = \left[e_\theta'\left(R\Omega R'\right)^{1/2}\left(R\hat{\Omega}R'\right)^{-1}\left(R\Omega R'\right)^{1/2} e_\theta\right],$$

$$e_\theta = \frac{\left(R\Omega_{T,GLS}R'\right)^{-1/2} R\sqrt{T}(\hat{\theta}_{GLS} - \theta)}{\left\|\left(R\Omega_{T,GLS}R'\right)^{-1/2} R\sqrt{T}(\hat{\theta}_{GLS} - \theta)\right\|}.$$

When $p = 1$, Lemma 1 reduces to a result in SPJ (2008). The lemma shows that the estimation uncertainty of $\hat{\Omega}$ affects the distribution of the Wald statistic only through Ξ. Taking a Taylor expansion, we have

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

where L is linear in $\hat{\Omega} - \Omega$ and Q is quadratic in $\hat{\Omega} - \Omega$. The exact expressions for L and Q are not important here but are given in the proof of Theorem 4. Using this stochastic expansion and Lemma 1, we can establish a higher order expansion of the finite sample distribution of $F_{T,OLS}$ under the conventional asymptotics where $K \to \infty$ such that $K/T \to 0$.

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

$$\begin{aligned} &P\left(\frac{(K-p+1)}{K} F_{T,OLS} < z\right) \\ &= G_p(z) + \frac{K^2}{T^2} G_p'(z)\, z\bar{B} + \frac{1}{K} G_p''(z)\, z^2 + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right) \end{aligned} \tag{9}$$

where

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

The first term in (9) 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.

Consider $p = 1$ and the kernel LRV estimator $\hat{\Omega}_D$:

$$\hat{\Omega}_D = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} k_D(\frac{t-s}{M})\hat{u}_t\hat{u}_s' \tag{10}$$

where

$$k_D(x) = \frac{\sin \pi x}{\pi x} \tag{11}$$

and $M = T/K$. If $\hat{\Omega}_D$ is used in constructing $F_{T,OLS}$, SPJ establish that, up to smaller order terms:

$$\begin{aligned} &P\left(\frac{(K-p+1)}{K}F_{T,OLS}(\hat{\Omega}_D) < z\right) \\ &= G_p(z) + \frac{K^2}{T^2}G_p'(z)\, z\bar{B}^* + \frac{1}{K}G_p''(z)\, z^2 - \frac{1}{K}G_p'(z)\, z\int_{-\infty}^{\infty} k_D(x)dx. \end{aligned}$$

This expansion is of the same form as (9) except that the SPJ expansion has an additional term:

$$-\frac{1}{K}G_p'(z)\, z\int_{-\infty}^{\infty} k_D(x)dx = -\frac{1}{2K}G_p'(z)\, z.$$

This term reflects the bias due to the estimation error of the model parameters. Such a term does not appear in (9) because the basis functions we employ are asymptotically orthogonal to the regressor, which is the column of ones in our model.

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

$$\begin{aligned} P\left(pF_{p,K-p+1} < z\right) &= P\left(\chi_p^2 < z\frac{\chi^2_{K-p+1}}{K-p+1}\right) = EG_p\left(z\frac{\chi^2_{K-p+1}}{K-p+1}\right) \\ &= G_p(z) + \frac{1}{K}G_p''(z)\, z^2 + o\left(\frac{1}{K}\right). \end{aligned}$$

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. A direct implication is that

$$\begin{aligned} &P\left(\frac{(K-p+1)}{K}F_{T,OLS} < pF^{\alpha}_{p,K-p+1}\right) \\ &= 1-\alpha + \frac{K^2}{T^2}G_p'\left(pF^{\alpha}_{p,K-p+1}\right) pF^{\alpha}_{p,K-p+1}\bar{B} + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right). \end{aligned} \tag{12}$$

Therefore, use of critical value $pF^{\alpha}_{p,K-p+1}$ removes the variance term $K^{-1}G_p''(z)\, z^2$ in the higher order expansion. The size distortion is then of order $O\left(K^2/T^2\right)$. In contrast, if the critical value from the conventional χ_p^2 distribution is used, the size distortion is of order $O\left(K^2/T^2\right) + O\left(1/K\right)$. So when $K^3/T^2 \to 0$, using critical value $pF^{\alpha}_{p,K-p+1}$ should lead to size improvement. We have thus shown that critical values from the fixed-K asymptotics are second order correct under the large-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 that can be interpreted as a Bartlett type correction. See Sun (2011b) for more details. Without this correction, we can show that, up to smaller order terms

$$\begin{aligned}&P\left(F_{T,OLS}<\chi_p^\alpha\right)\\&=G_p\left(\chi_p^\alpha\right)+\frac{K^2}{T^2}G_p'\left(\chi_p^\alpha\right)\chi_p^\alpha\bar{B}-\frac{1}{K}G_p'\left(\chi_p^\alpha\right)\chi_p^\alpha(p-1)+\frac{1}{K}G_p''\left(\chi_p^\alpha\right)\left[\chi_p^\alpha\right]^2.\end{aligned}$$

Comparing this with (9), we find that the above expansion has an additional term $-K^{-1}G_p'\left(\chi_p^\alpha\right)\chi_p^\alpha(p-1)$. 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 confidence regions tend to have large under-coverage when the number of joint hypotheses is large.

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

$$F_{T,OLS}^*=\frac{(K-p+1)}{K}F_{T,OLS}\tag{13}$$

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

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

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

$$\begin{aligned}&P\left(F_{T,OLS}^*>pF_{p,K-p+1}^\alpha\right)-\alpha\\&=-\frac{K^2\bar{B}}{T^2}G_p'\left(\chi_p^\alpha\right)\chi_p^\alpha+o\left(\frac{1}{K}\right)+o\left(\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right).\end{aligned}\tag{14}$$

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

$$\begin{aligned}P\left(F_{T,OLS}^*>pF_{p,K-p+1}^\alpha\right)&=1-G_{p,\delta^2}\left(\chi_p^\alpha\right)-\frac{K^2}{T^2}G_{p,\delta^2}'\left(\chi_p^\alpha\right)\chi_p^\alpha\bar{B}\\&-\frac{1}{K}\frac{\chi_p^\alpha\delta^2}{2}G_{(p+2),\delta^2}'\left(\chi_p^\alpha\right)+o\left(\frac{1}{K}\right)+o\left(\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right),\end{aligned}\tag{15}$$

where $G_{p,\delta^2}'(z)$ is the pdf of noncentral χ^2 distribution with degrees of freedom p and noncentrality parameter δ^2.

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_θ is uniformly distributed on the unit sphere $\mathfrak{S}_p(1)$, which follows from the rotation invariance of the multivariate standard normal distribution. The uniformity of $\tilde{c}$ ensures that the same property holds for the corresponding statistic

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

under the local alternative hypothesis.

6 Optimal Smoothing Parameter Selection

We have shown that the optimal K that minimizes the asymptotic mean squared error in LRV estimation has the form $K = O\left(T^{4/5}\right)$. However, there is no reason to expect that such a choice is the most appropriate in statistical testing using nonparametrically scaled statistics. This section attempts to provide a new approach for optimal K selection that addresses the central concern of classical hypothesis testing, which can be expressed as minimizing type II error subject to controlling the type I error. The approach is also used in Sun, Phillips and Jin (2010) who consider univariate Gaussian location models and employ the method of exponentiated kernels.

6.1 Test-Optimal K

In view of the asymptotic expansion in (14) and ignoring the higher order terms, the type I error for the F^* test can be measured by

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

Similarly, from (15), the type II error of the F^* test can be measured by

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

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

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

where κ is a constant greater than 1. The optimal K value for the case $\kappa = 1$ can be obtained by letting $\kappa \to 1+$. Ideally, the type I error is less than or equal to the nominal type I error α. In finite samples, approximation errors are unavoidable and we allow for some discrepancy by introducing the tolerance factor κ. For example, when $\alpha = 10\%$ and $\kappa = 1.2$, we aim to control the type I error such that it is not greater than 12%. Note that κ may depend on the sample size T. For a larger sample size, we may require κ to take smaller values.

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

$$K_{\text{opt}} = \left(\frac{\delta^2 G'_{(p+2),\delta^2}\left(\chi_p^\alpha\right)}{4\bar{B}G'_{p,\delta^2}\left(\chi_p^\alpha\right)}\right)^{1/3} T^{2/3}. \tag{16}$$

When $\bar{B} < 0$, it can be shown that the constraint $e_I \leq \kappa\alpha$ is binding and we have to form the expanded objective function as follows:

$$L(K,\lambda) = G_{p,\delta^2}\left(\chi_p^\alpha\right) + \frac{K^2}{T^2} G'_{p,\delta^2}\left(\chi_p^\alpha\right)\chi_p^\alpha \bar{B} + \frac{1}{K}\frac{\chi_p^\alpha}{2}\delta^2 G'_{(p+2),\delta^2}\left(\chi_p^\alpha\right) + \lambda\left(\left(\alpha - \frac{K^2\bar{B}}{T^2}G'_p\left(\chi_p^\alpha\right)\chi_p^\alpha\right) - \kappa\alpha\right). \tag{17}$$

where λ is the Lagrange multiplier. Using the Kuhn-Tucker theorem, we find the optimal K is

$$K_{\text{opt}} = \left(\frac{(\kappa-1)\,\alpha}{\left|\bar{B}\right| G'_p\left(\chi_p^\alpha\right)\chi_p^\alpha}\right)^{1/2} T, \tag{18}$$

and the corresponding Lagrange multiplier is

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

Formulae (16) and (18) can be written collectively as

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

where

$$\lambda_{\text{opt}} = \begin{cases} 0, & \text{if } \bar{B} > 0 \\ \frac{G'_{p,\delta^2}\left(\chi_p^\alpha\right)}{G'_p\left(\chi_p^\alpha\right)} + \delta^2\frac{\left|\bar{B}\right|^{1/2} G'_{(p+2),\delta^2}\left(\chi_p^\alpha\right)\left[\chi_p^\alpha\right]^{3/2}\left[G'_p\left(\chi_p^\alpha\right)\right]^{1/2}}{4\left[(\kappa-1)\alpha\right]^{3/2} T}, & \text{if } \bar{B} < 0 \end{cases} \tag{20}$$

Since K is an integer greater than or equal to n, in practice, we take $\max(int(K_{\text{opt}}), n)$ as the selected K value, where $int(\cdot)$ is the nearest integer function.

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

The tolerance parameter κ affects the choice of K_{opt} only when $\bar{B} < 0$. As the tolerance parameter κ decreases toward 1, the weight to the type I error increases and the optimal K_{opt} decreases. The case of $\kappa = 1$ can be obtained by letting $\kappa \to 1$ from the right. When $\bar{B} < 0$ and $\kappa \to 1+$, the multiplicative constant in (19) approaches zero. As a result, the theoretically optimal K_{opt} will be smaller than n, the lower bound that ensures the positive semi-definiteness of the LRV estimator $\hat{\Omega}$. In this case, we use the lower bound as the data-driven choice of K. This is a reasonable choice. When $\bar{B} < 0$, the LRV estimator has a downward bias and the associated test tends to be oversized. To alleviate the problem, we choose K as small as possible in order to reduce the bias. On the other hand, the variance inflation from using a small K can be accommodated by using the F approximation rather than the standard χ^2 approximation.

The formula for K_{opt} depends on the noncentrality parameter δ^2. For practical implementations, we recommend choosing δ^2 such that the first order power of the test, as measured by $1 - G_{p,\delta^2}(\chi_p^\alpha)$, is 50%. That is, we solve $1 - G_{p,\delta^2}(\chi_p^\alpha) = 50\%$ for a given p and a given significance level α. The marginal effect of δ^2 on K_{opt} when the corresponding power is in the middle range, say [0.35, 0.75], is small. More specifically, compared to the K_{opt} value obtained under the recommended power of 50%, the K_{opt} values for other power requirements in the range [0.35, 0.75] differ by less than 10%. In the simulation study, we have experimented with 50% and 75% as the power levels to choose δ^2. The simulation results are very close to each other.

To sum up, the testing-optimal K that minimizes the type II error while controlling the type I error 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.

6.2 Data Driven Implementation

The optimal bandwidth in (20) can be written as $K_{opt} = K_{opt}(\bar{B})$. It involves unknown parameter $\bar{B}$, which can be estimated nonparametrically (e.g. Newey and West (1994)) or by a standard plug-in procedure based on a simple parametric model like a VAR (e.g. Andrews (1991)). Both methods achieve a valid order of magnitude and the procedure is analogous to conventional data-driven methods for kernel LRV estimation.

We focus the discussion on the plug-in procedure, which involves the following steps. First, we estimate the model using the OLS estimator and compute the residuals $\{\hat{u}_t\}$. Second, we specify a multivariate approximating parametric model and fit the model to $\{\hat{u}_t\}$ by the standard OLS method. Third, we treat the fitted model as if it were the true model for the process $\{u_t\}$ and compute $\bar{B}$ as a function of the parameters of the parametric model. Plugging the estimate $\bar{B}$ into (20) gives the automatic bandwidth $\hat{K}$.

As in the case of MSE-optimal bandwidth choice, the automatic bandwidth considered here deviates from the finite sample optimal one due to the error introduced by estimation, the use of approximating parametric models, and the approximation inherent in the asymptotic formula employed. It is hoped that in practical work the deviation is not large so that the test based on the automatic bandwidth still has good performance. Some simulation evidence reported in Section 8 supports this argument.

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

$$\hat{\Omega} = (\mathbb{I}_n - \hat{A})^{-1}\hat{\Sigma}(\mathbb{I}_n - \hat{A}')^{-1}, \tag{21}$$

$$\hat{B} = -\frac{2\pi^2}{3}(\mathbb{I}_n - \hat{A})^{-3}\left(\hat{A}\hat{\Sigma} + \hat{A}^2\hat{\Sigma}\hat{A}' + \hat{A}^2\hat{\Sigma} - 6\hat{A}\hat{\Sigma}\hat{A}'\right.$$
$$\left.+\hat{\Sigma}(\hat{A}')^2 + \hat{A}\hat{\Sigma}(\hat{A}')^2 + \hat{\Sigma}\hat{A}'\right)(\mathbb{I}_n - \hat{A}')^{-3}. \tag{22}$$

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 K can be computed as

$$\hat{K} = \max\left\{int\left[\hat{K}_{opt}(\bar{B}(R, \hat{B}, \hat{\Omega})\right], n\right\}. \tag{23}$$

It should be pointed out that the computational cost involved in this automatic smoothing parameter selection is the same as that of the conventional plug-in bandwidth based on the MSE criterion.

7 Extension to General Settings

In the previous sections, we use the simple multivariate location model to highlight the importance of smoothing parameter choice in semiparametric testing. Hypothesis testing in location models, as simple as it seems, includes more general testing problems as special cases.

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

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

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

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

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

Define $s_t(\theta) = \partial\rho(\theta, Z_t)/\partial\theta$. Then

$$Es_t(\theta) = 0,$$

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

The above extension applies to hypothesis testing that involves the whole parameter vector θ. Suppose we are only interested in some linear combinations of θ such that the null hypothesis is $H_0 : R\theta = r$ and the alternative hypothesis is $H_1 : R\theta \neq r$, where R is a $p \times n$ matrix. We can construct F^*-statistic as before:

$$F_T^* = \frac{(K - p + 1)}{K} F_T$$

where F_T is usual Wald statistic:

$$F_T = \left[\sqrt{T}\left(R\hat{\theta}_T - r\right)\right]' \left(R\hat{H}_T^{-1}\hat{\Omega}\hat{H}_T^{-1}R'\right)^{-1} \sqrt{T}\left(R\hat{\theta}_T - r\right),$$

$$\hat{\Omega} = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k(\frac{t}{T})\hat{s}_t\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k(\frac{t}{T})\hat{s}_t\right)',$$

and

$$\hat{s}_t = s_t\left(\hat{\theta}_T\right),\ \hat{H}_T = H_T(\hat{\theta}_T) \text{ and } H_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial s_t\left(\theta\right)}{\partial\theta}.$$

Here $\hat{\Omega}$ is the series estimator of the LRV Ω of $s_t\left(\theta_0\right)$.

If $s_t\left(\theta_0\right)$ satisfies a FCLT:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[rT]} s_t\left(\theta_0\right) \to^d \Lambda W_n(r) \tag{24}$$

and a uniform law of large numbers:

$$\sup_{\theta\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^{[rT]}\frac{\partial s_t\left(\theta\right)}{\partial\theta} - rH_\infty(\theta)\right\| \to^p 0$$

uniformly in r, for a nonsingular Hessian matrix $H_\infty(\theta)$. Then both the fixed-K asymptotics and the large-K asymptotics in Section 4 hold for F_T^*. As a result, we can employ the critical value $pF_{p,K-p+1}^{\alpha}$ from the $F_{p,K-p+1}$ distribution and perform the F^* test as before.

The question is how to select the testing-optimal smoothing parameter K. Let

$$\mathfrak{s}_t(\theta_1,\theta_2) = RH_T^{-1}\left(\theta_1\right)s_t\left(\theta_2\right)$$

be the transformed score type of process, then under standard assumptions we can write F_T as

$$F_T = \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathfrak{s}(\tilde{\theta}_T,\theta_0)\right)'\tilde{\Omega}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathfrak{s}(\tilde{\theta}_T,\theta_0)\right)'$$

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

$$\tilde{\Omega} = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k(\frac{t}{T})\mathfrak{s}_t(\hat{\theta}_T,\hat{\theta}_T)\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k(\frac{t}{T})\mathfrak{s}_t(\hat{\theta}_T,\hat{\theta}_T)\right)'.$$

It follows that the F_T statistic can be viewed as a Wald statistic for testing whether the mean of the multivariate process $\mathfrak{s}_t(\tilde{\theta}_T,\theta_0)$ is zero. In principle, we can use formulae in (19) and (20) to compute the testing-optimal K. However, $\mathfrak{s}_t(\tilde{\theta}_T,\theta_0)$ is not observable and a plug-in estimator of $\bar{B}$ is not feasible. As an empirical and practical procedure, we may replace $\mathfrak{s}(\tilde{\theta}_T,\theta_0)$ by $\mathfrak{s}_t(\hat{\theta}_T,\hat{\theta}_T)$ and fit an approximating parametric model to the observed process $\mathfrak{s}_t(\hat{\theta}_T,\hat{\theta}_T)$. In other words, we implement (19) and (20) using the standard plug-in procedure based on the process $\mathfrak{s}_t(\hat{\theta}_T,\hat{\theta}_T)$. We can then perform the F^* test using the plug-in implementation of K_{opt}. A rigorous study of this testing procedure is reported in Appendix B.

8 Simulation Evidence

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

We consider the case with 6 time series, i.e. $n = 6$. The error follows either a VAR(1) or VMA(1) process:

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

where $A = \rho\mathbb{I}_n$, $\varepsilon_t = (v_{1t} + \mu f_t, v_{2t} + \mu f_t, ..., v_{nt} + \mu f_t)' / \sqrt{1+\mu^2}$ and $(v_t, f_t)'$ is a multivariate Gaussian white noise process with unit variance. 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 dependence among the time series considered. When $\mu = 0$, the six series are uncorrelated with each other. When $\mu = 1$, the six series have the same pairwise correlation coefficient 0.5. The variance-covariance matrix of u_t is normalized so that the variance of each series u_{it} is equal to one for all values of $|\rho| < 1$. For the VAR(1) process, $\Omega = \left(1-\rho^2\right)\left(\mathbb{I}_n - A\right)^{-1}\Sigma\left(\mathbb{I} - A'\right)^{-1}$. For the VMA(1) process $\Omega = \left(1-\rho^2\right)\left(\mathbb{I}_n + A/\sqrt{1-\rho^2}\right)\Sigma(\mathbb{I}_n + A/\sqrt{1-\rho^2})'$.

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

We consider the following null hypotheses:

$$\begin{aligned}
&H_{01} : \theta_1 = 0,\\
&H_{02} : \theta_1 = \theta_2 = 0,\\
&H_{03} : \theta_1 = \theta_2 = \theta_3 = 0,\\
&H_{04} : \theta_1 = \theta_2 = ... = \theta_6 = 0,
\end{aligned}$$

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 $H_1\left(\delta^2\right)$.

We examine the finite sample performance of four different testing methods. The first one is the new F^* test, which is based on the modified Wald statistic and uses the testing-optimal K and critical values from the F-distribution. The second one is the conventional Wald test, which is based on the unmodified Wald statistic and uses the MSE-optimal K and critical values from the χ^2 distribution. The third one is the same as the new F^* test but with $K = n$. The last one is the test proposed by Kiefer and Vogelsang (2002) and is based on the Bartlett kernel LRV estimator with bandwidth equal to the sample size and uses the nonstandard asymptotic theory. The four methods are referred as 'K-OPT', 'K-MSE', 'K-n' and 'KV' respectively in the tables and figures below.

Table 1: Empirical Type I error of different 5% tests for VAR(1) error with $T = 100, \kappa = 1.1$

	K-OPT	K-MSE	K-n	K-MSE*	KV	K-OPT	K-MSE	K-n	K-MSE*	KV
			$p = 1$					$p = 2$		
$\rho = 0$	.0540	.0617	.0471	.0499	.0485	.0506	.0685	.0460	.0453	.0452
$\rho = .25$	.0839	.0820	0.0489	.0609	.0520	.0949	.1128	.0474	.0632	.0522
$\rho = .50$	.0923	.1038	.0544	.0643	.0586	.0993	.1687	.0542	.0657	.0656
$\rho = .75$	.1079	.1525	.0908	.0908	.0798	.1175	.2778	.0899	.0900	.1059
			$p = 3$					$p = 6$		
$\rho = 0$	.0567	.0900	.0449	.0489	.0526	.0539	.1698	.0462	.0498	.0512
$\rho = .25$	.1122	.1631	.0467	.0675	.0663	.1539	.3673	.0483	.0755	.0857
$\rho = .50$	.1160	.2660	.0597	.0709	.0888	.1586	.5788	.0706	.0747	.1485
$\rho = .75$	.1521	.4272	.1227	.1227	.1575	.2677	.8017	.2109	.2109	.3342

The K-n test and K-MSE test can be regarded as the two ends of the power and size tradeoff. While the K-n test aims at controlling the size, the K-MSE test puts more weights on power maximization. Many other testing methods are available in the literature. We do not include all of them here as their performances are likely to be between the K-n test and the K-MSE test.

Table 1 gives the type I error of the four testing methods for the VAR(1) error with sample size $T = 100$, tolerance parameter $\kappa = 1.1$ and $\mu = 0$. The table also includes a hybrid procedure denoted 'K-MSE*' which employs the MSE-optimal K and critical values from the fixed-K asymptotics. The only difference between the conventional method and the hybrid method lies in the critical value used. The significance level is 5%, which is also the nominal type I error. Several patterns emerge. First, as it is clear from the table, the conventional method has large size distortion. The size distortion increases with both the error dependence and the number of restrictions being tested. This result is consistent with our theoretical analysis. The size distortion can be very severe. For example, when $\rho = 0.75$ and $p = 6$, the empirical type I error of the test is 0.8017, which is far from 0.05, the nominal type I error. Second, comparing the K-MSE test with the K-MSE* test, we find that using critical values from the fixed-K asymptotics eliminates the size distortion of the conventional Wald test 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 fixed-K asymptotics and large-K asymptotics is larger. Note that the hybrid procedure is rigorously justified as the critical value from the fixed-K asymptotics is second order correct under the conventional large-K asymptotics. Third, the size distortion of the K-OPT, K-n, and KV tests is substantially smaller than the conventional method. This is because these three tests employ asymptotic approximations that capture the estimation uncertainty of the LRV estimator. Finally, compared with the K-n and KV methods, the K-OPT test has slightly larger size distortion. Since the bandwidth is set equal to the sample size, the KV method is designed to achieve the smallest possible size distortion. Similarly, the K-n test uses the smallest possible K value in order to achieve the best size accuracy. Given these observations, we can conclude that the K-OPT test succeeds in controlling the type I error.

Table 2 presents the simulated type I errors for the VMA(1) error process. The qualitative observations for the VAR(1) error remain valid. In fact, these qualitative observations

Table 2: Empirical Type I error of different 5% tests based on VMA(1) error with $T = 100, \kappa = 1.1$

	K-OPT	K-MSE	K-n	K-MSE*	KV	K-OPT	K-MSE	K-n	K-MSE*	KV
			$p = 1$					$p = 2$		
$\rho = 0$	.0540	.0617	.0471	.0499	.0485	.0506	.0685	.0460	.0453	.0452
$\rho = .25$	.0729	.0733	.0484	.0554	.0505	.0775	.0982	.0455	.0532	.0493
$\rho = .50$	.0660	.0784	.0495	.0522	.0527	.0681	.1155	.0459	.0512	.0517
$\rho = .75$	.0616	.0807	.0492	.0500	.0523	.0634	.1271	.0465	.0480	.0503
			$p = 3$					$p = 6$		
$\rho = 0$	.0567	.0900	.0449	.0489	.0526	.0539	.1698	.0462	.0498	.0512
$\rho = .25$	.0893	.1378	.0447	.0580	.0610	.1208	.3156	.0452	.0632	.0716
$\rho = .50$	.0774	.1749	.0454	.0540	.0646	.1032	.4417	.0468	.0597	.0818
$\rho = .75$	.0741	.1994	.0459	.0529	.0646	.0999	.4783	.0457	.0509	.0834

hold for other parameter configurations such as different sample sizes and different values of μ. All else being equal, the size distortion of the K-OPT test for $\kappa = 1.2$ is slightly larger than that for $\kappa = 1.1$. This is expected as we allow for higher tolerance for the type I error when the value of κ is larger.

Figures 1-4 present the finite sample power under the VAR(1) error for different values of p. We compute the power using the 5% empirical finite sample critical values obtained from the null distribution. So the finite sample power is size-adjusted and power comparisons are meaningful. It should be pointed out that the size-adjustment is not feasible in practice. The parameter configuration is the same as those for Table 1 except the DGP is generated under the local alternatives. Three observations can be drawn from these figures. First, the power of the K-OPT test is never lower than the conventional Wald test. This result demonstrates the advantage of using the criterion that focuses on the testing problem at hand. Consistent with the asymptotic result, the focused criterion has superior performance in finite samples than the MSE criterion which is not suitable for hypothesis testing. Second, the K-OPT test has higher power than the KV test in most cases except when the error dependence is very high and the number of restrictions is large. When the error dependence is low, the selected K value is relatively large and the variance of the associated LRV estimator is small. In contrast, the LRV estimator used in the KV test has a large variance. As a result, the K-OPT test is more powerful than the KV test. On the other hand, when the error dependence is very large, the selected K value is very small. In this case, both the KV test and the K-OPT test employ LRV estimators with large variance. The KV test can be more powerful in this scenario. Finally, the K-OPT test is consistently more powerful than the K-n test. The power improvement is substantial in most of the cases. This is not surprising as the K-n test, like the KV test and many other nonstandard tests proposed by Vogelsang and his coauthors, is designed to have good size properties but often at the cost of power loss.

To save space, we do not report the figures for the power curves under the VMA(1) error but make a brief comment. We find that the figures reinforce and strengthen the observations for the VAR(1) error. It is clear under the VMA(1) error that the K-OPT test is more powerful than the conventional K-MSE test, the nonstandard KV test and the K-n test. This is true for all parameter combinations considered.

In simulations not reported here, we have considered VAR(1) and VMA(1) errors with negative values of ρ and hypotheses of the form $\theta_1 = \theta_2 = ... = \theta_{j_0}$ for some $j_0 \geq 2$. For some of these configurations, $\bar{B} > 0$. Regardless of the sign of $\bar{B}$, in terms of the type I error and size adjusted power, the performance of the new K-OPT test is at least as good as the conventional Wald test and often much better. It also dominates the KV test and the K-n test in most scenarios considered. Details are available upon request.

9 Conclusion

Using the series LRV estimator, the paper proposes a new approach to multivariate mean inference in the presence of nonparametric autocorrelation. The series LRV estimator is asymptotically invariant to the intercept parameters. This property releases us from worrying about the estimation uncertainly of those parameters. Another advantage of the series LRV estimator is that the associated (modified) Wald statistic converges to a standard distribution regardless of the asymptotic specifications of the smoothing parameter. This property releases practitioners from the computational burden of simulating nonstandard critical values. 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. The idea is in the spirit of the Neyman-Pearson Lemma. Monte Carlo simulations show that our inference procedure enjoys superior performance in finite samples. The procedure can be applied to more general hypothesis testing problems.

There are many extensions to the current paper. One possibility is to optimally select the smoothing parameter for confidence region construction. We can minimize the volume of the region subject to the constraint that the coverage probability is at least at some level. The volume can be the physical volume or an indirect measure such as the probability of including the false values (see Neyman (1937)). The idea of optimal smoothing choice and the inference procedure given here may be also used in general linear and non-linear models with nonstationary time series. These extensions will be reported in future work.

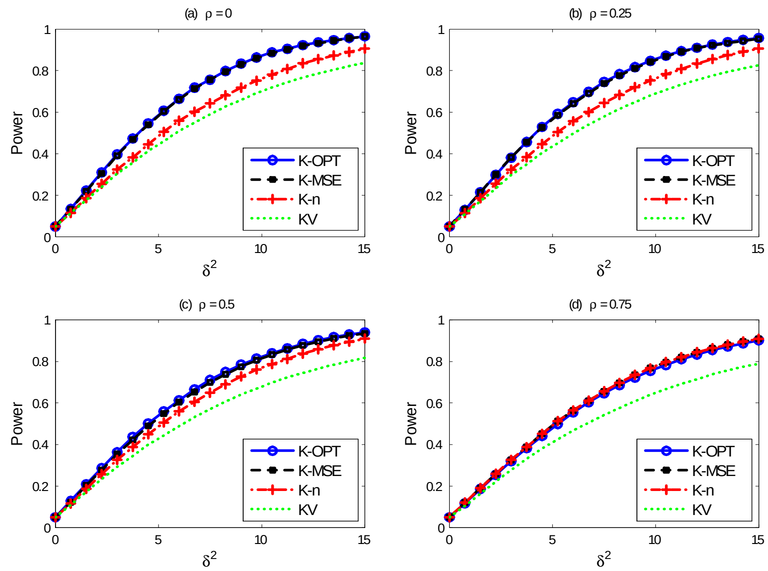

Figure 1: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with $T = 100, \kappa = 1.1$ and $p = 1$.

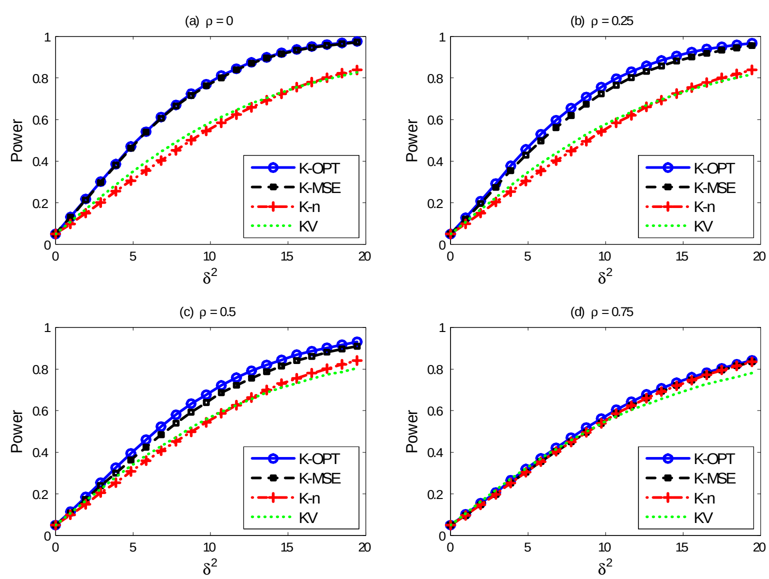

Figure 2: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with $T = 100, \kappa = 1.1$ and $p = 2$.

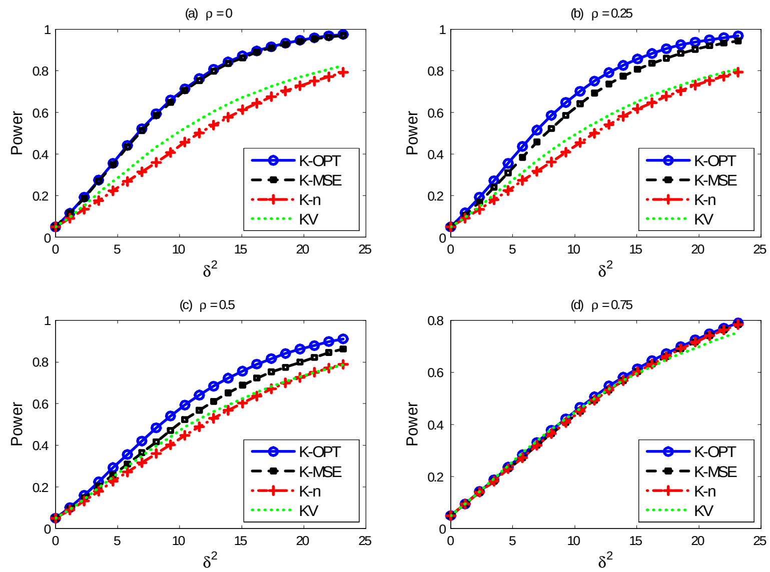

Figure 3: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with $T = 100, \kappa = 1.1$ and $p = 3$.

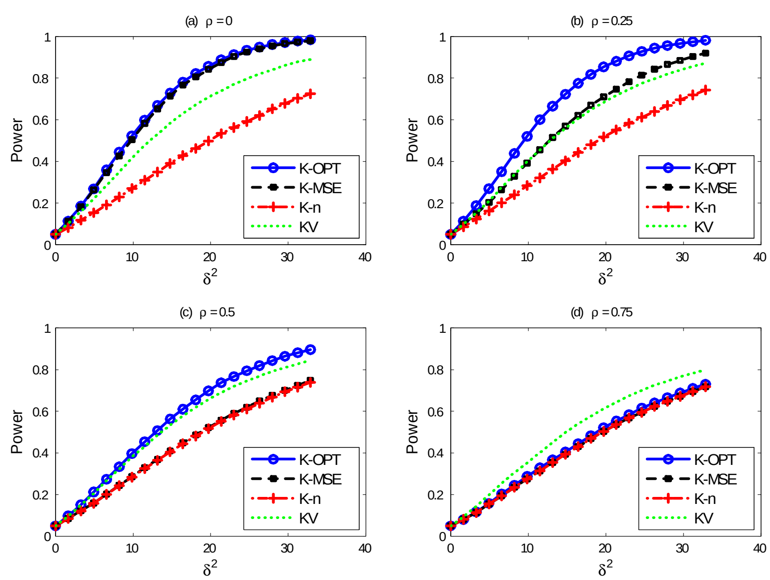

Figure 4: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with $T = 100, \kappa = 1.1$ and $p = 6$.

10 Appendix of Proofs

10.1 Appendix A: Proof of Main Results

Proof of Theorem 1. As in (4), we use summation by parts to obtain:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k\left(\frac{t}{T}\right)\hat{u}_t=\frac{1}{T}\sum_{t=1}^{T}\frac{\left[\phi_k\left(t/T\right)-\phi_k\left(\left(t+1\right)/T\right)\right]}{1/T}S_t.$$

Since

$$\frac{\left[\phi_k\left(\left[Tr\right]/T\right)-\phi_k\left(\left(\left[Tr\right]+1\right)/T\right)\right]}{1/T}\to-\phi_k'(r)$$

and

$$S_{[Tr]}\to^d \Lambda V_n(r),$$

invoking the continuous mapping theorem, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k\left(\frac{t}{T}\right)\hat{u}_t\to^d\Lambda\int_0^1-\phi_k'(r)V_n(r)dr.$$

Using integration by parts, we can show that

$$\int_0^1-\phi_k'(r)V_n(r)dr=\int_0^1\left(\phi_k(r)-\int_0^1\phi_k(r)dr\right)dW_n(r).$$

Hence

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k\left(\frac{t}{T}\right)\hat{u}_t\to^d\Lambda\int_0^1\left(\phi_k(r)-\int_0^1\phi_k(r)dr\right)dW_n(r):=\Lambda\zeta_k,$$

and

$$\hat{\Omega}\to^d\Lambda\frac{1}{K}\sum_{k=1}^{K}\zeta_k\zeta_k'\Lambda'. \tag{25}$$

■

Proof of Lemma 1. Part (a). Define

$$\Omega_{T,GLS}=var\left[T^{1/2}\left(\hat{\theta}_{GLS}-\theta\right)\right].$$

We write the statistic $F_{T,GLS}$ as

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

where

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

and

$$\Xi = e_\theta' \left(R\Omega R'\right)^{1/2} \left(R\hat{\Omega}R'\right)^{-1} \left(R\Omega R'\right)^{1/2} e_\theta.$$

Here we have used

$$\Omega_{T,GLS} = \Omega\left(1 + O\left(\frac{1}{T}\right)\right).$$

Note that Θ is independent of Ξ because (i) Θ is the squared length of a standard normal vector and e_θ is the direction of this vector. The length is independent of the direction. (ii) $(\hat{\theta}_{GLS} - \theta)$ is independent of $\hat{\Omega}$. To see the second independence result, recall that the GLS estimator $\hat{\theta}_{GLS}$ of θ satisfies

$$\hat{\theta}_{GLS} - \theta = \left[\left(\ell_T \otimes \mathbb{I}_n\right)' \Omega_u^{-1} \left(\ell_T \otimes \mathbb{I}_n\right)\right]^{-1} \left(\ell_T \otimes \mathbb{I}_n\right)' \Omega_u^{-1} u$$

and the OLS estimator satisfies

$$\hat{\theta}_{OLS} - \theta = \left[\left(\ell_T \otimes \mathbb{I}_n\right)' \left(\ell_T \otimes \mathbb{I}_n\right)\right]^{-1} \left(\ell_T \otimes \mathbb{I}_n\right)' u = \left(\frac{1}{T}\ell_T' \otimes \mathbb{I}_n\right) u.$$

As a result,

$$\begin{aligned} \hat{u} &= \left[\hat{u}_1', \hat{u}_2', ..., \hat{u}_T'\right]' = u - \ell_T \otimes \left(\hat{\theta}_{OLS} - \theta\right) \\ &= \left[\mathbb{I}_T \otimes \mathbb{I}_n - \ell_T \otimes \left(\frac{1}{T}\ell_T' \otimes \mathbb{I}_n\right)\right] u. \end{aligned}$$

So

$$\begin{aligned} & E\left[\hat{\theta}_{GLS} - \theta\right] \hat{u}' \\ = \; & E\left[\left(\ell_T \otimes \mathbb{I}_n\right)' \Omega_u^{-1} \left(\ell_T \otimes \mathbb{I}_n\right)\right]^{-1} \left(\ell_T \otimes \mathbb{I}_n\right)' \Omega_u^{-1} u \\ & u \times \left[\mathbb{I}_T \otimes \mathbb{I}_n - \ell_T \otimes \left(\frac{1}{T}\ell_T' \otimes \mathbb{I}_n\right)\right] \\ = \; & \left[\left(\ell_T \otimes \mathbb{I}_n\right)' \Omega_u^{-1} \left(\ell_T \otimes \mathbb{I}_n\right)\right]^{-1} \left(\ell_T \otimes \mathbb{I}_n\right)' \left[\mathbb{I}_T \otimes \mathbb{I}_n - \ell_T \otimes \left(\frac{1}{T}\ell_T' \otimes \mathbb{I}_n\right)\right] \\ = \; & \left[\left(\ell_T \otimes \mathbb{I}_n\right)' \Omega_u^{-1} \left(\ell_T \otimes \mathbb{I}_n\right)\right]^{-1} \left\{\ell_T' \otimes \mathbb{I}_n - \ell_T' \otimes \mathbb{I}_n\right\} = 0. \end{aligned}$$

That is, $\hat{\theta}_{GLS}$ is independent of $\hat{u}$. But $\hat{\Omega}$ is a function of $\hat{u}$. Hence $\hat{\theta}_{GLS}$ is independent of $\hat{\Omega}$.

Hence

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

as stated.

Part (b). Let

$$\zeta_{1T} = 2T^{1/2}(R\Delta)'\left(R\hat{\Omega}R'\right)^{-1}\left(R\Omega_{T,GLS}R'\right)^{1/2}e_\theta$$
$$\zeta_{2T} = T^{1/2}(R\Delta)'\left(R\hat{\Omega}R'\right)^{-1}RT^{1/2}\Delta$$

and $\zeta_T = \sqrt{\Theta}\zeta_{1T} + \zeta_{2T}$. Then $F_{T,OLS} = F_{T,GLS} + \zeta_T$. We first show that Θ is independent of ζ_{1T} and ζ_{2T}. Note that Θ is independent of $\hat{\Omega}$ and e_θ, it is sufficient to show that $(\hat{\theta}_{GLS} - \theta)$ is independent of $\Delta := \hat{\theta}_{OLS} - \theta - (\hat{\theta}_{GLS} - \theta)$. This follows because

$$\begin{aligned}
& E(\hat{\theta}_{GLS} - \theta)\Delta' \\
= \; & E\left[(\ell_T \otimes \mathbb{I}_n)'\,\Omega_u^{-1}(\ell_T \otimes \mathbb{I}_n)\right]^{-1}(\ell_T \otimes \mathbb{I}_n)'\,\Omega_u^{-1}uu'\left(\frac{1}{T}\ell_T \otimes \mathbb{I}_n\right) \\
& -\left[(\ell_T \otimes \mathbb{I}_n)'\,\Omega_u^{-1}(\ell_T \otimes \mathbb{I}_n)\right]^{-1} \\
= \; & \left[(\ell_T \otimes \mathbb{I}_n)'\,\Omega_u^{-1}(\ell_T \otimes \mathbb{I}_n)\right]^{-1}(1 \otimes \mathbb{I}_n) - \left[(\ell_T \otimes \mathbb{I}_n)'\,\Omega_u^{-1}(\ell_T \otimes \mathbb{I}_n)\right]^{-1} \\
= \; & 0.
\end{aligned}$$

Given that Θ is independent of ζ_{1T}, ζ_{2T} and Ξ, we have

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

where

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

But

$$\begin{aligned}
& EF\left(\zeta_{1T}, \zeta_{2T}, \Xi\right) \\
= \; & EF(0,0,\Xi) + EF_1'(0,0,\Xi)\,\zeta_{1T} + O\left(E\zeta_{1T}^2\right) + O\left(E\left|\zeta_{1T}\zeta_{2T}\right|\right) + O\left(E\zeta_{2T}\right) \\
= \; & EF(0,0,\Xi) + EF_1'(0,0,\Xi)\,\zeta_{1T} + O\left(\frac{1}{T}\right).
\end{aligned}$$

where $F_1' = \partial F/\partial a$. Here we have used: $O\left(E\zeta_{1T}^2\right) = O(1/T)$ and $O\left(E\zeta_{2T}\right) = O(1/T)$, which follows from $var(c'\Delta\Delta'c) = O(1/T)$ for any constant c. Next, let $f_{e_\theta}(x)$ be the pdf of e_θ. Since e_θ is independent of $\hat{\Omega}$ and Δ, we have

$$\begin{aligned}
& EF_1'(0,0,\Xi)\,\zeta_{1T} \\
= \; & \int E\left[F_1'(0,0,\Xi)\,\zeta_{1T}\,|e_\theta = x\right] f_{e_\theta}(x)dx \\
= \; & \int E\left[F_1'\left(0,0,x'\left(R\Omega R'\right)^{1/2}\left(R\hat{\Omega}R'\right)^{-1}\left(R\Omega R'\right)^{1/2}x\right)\right. \\
& \left.\times 2T^{1/2}(R\Delta)'\left(R\hat{\Omega}R'\right)^{-1}\left(R\Omega_{T,GLS}R'\right)^{1/2}x\right] f_{e_\theta}(x)dx.
\end{aligned}$$

Note that $\hat{\Omega}(u) = \hat{\Omega}(-u)$ and $\Delta = -\Delta(-u)$, we have

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

As a result,

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

So

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

We have therefore shown that

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

as desired. ∎

Proof of Theorem 4. Writing $\Xi = \Xi(\hat{\Omega})$ and taking a Taylor expansion of $\Xi(\hat{\Omega})$ around $\Xi(\Omega) = 1$, we have

$$\left[\Xi\left(\hat{\Omega}\right)\right]^{-1} = 1 + L + Q + remainder \tag{26}$$

where *remainder* is the remainder term of the expansion,

$$\begin{aligned}
L &= Dvec\left(\hat{\Omega}-\Omega\right) \\
Q &= \frac{1}{2}vec\left(\hat{\Omega}-\Omega\right)'\left(J_1+J_2\right)vec\left(\hat{\Omega}-\Omega\right)
\end{aligned}$$

and

$$\begin{aligned}
D &= \left(\left[e_\theta'\left(R\Omega R'\right)^{-1/2}R\right]\otimes\left[e_\theta'\left(R\Omega R'\right)^{-1/2}R\right]\right) \\
J_1 &= 2R'\left(R\Omega R'\right)^{-1/2}\left(e_\theta e_\theta'\right)\left(R\Omega R'\right)^{-1/2}R\otimes R'\left(R\Omega R'\right)^{-1/2}\left(e_\theta e_\theta'\right)\left(R\Omega R'\right)^{-1/2}R \\
J_2 &= -\left[R'\left(R\Omega R'\right)^{-1/2}e_\theta e_\theta'\left(R\Omega R'\right)^{-1/2}R\otimes R'\left(R\Omega R'\right)^{-1}R\right]\mathbb{K}_{nn}\left(\mathbb{I}_{n^2}+\mathbb{K}_{nn}\right).
\end{aligned}$$

It is not difficult to show that

$$EL = \frac{K^2}{T^2}tr\left[\left(R\Omega R'\right)^{-1/2}\left(RBR'\right)\left(R\Omega R'\right)^{-1/2}\right]\frac{1}{p}\left(1+o\left(1\right)\right)+O\left(\frac{1}{T}\right),$$

$$EL^2 = \frac{2}{K} + o\left(\frac{1}{K}+\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right),$$

and

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

Hence

$$\left[\Xi\left(\hat{\Omega}\right)\right]^{-1}=1+L+Q+o_p\left(\frac{1}{K}+\frac{K^2}{T^2}\right)+o_p\left(\frac{1}{T}\right). \tag{27}$$

Using the above asymptotic expansion, we have

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

as desired. ■

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

$$\begin{aligned}
&P\left(F_{T,OLS}^{*}>pF_{p,K-p+1}^{\alpha}\right)-\alpha\\
&=-\frac{K^2}{T^2}G_p'\left(pF_{p,K-p+1}^{\alpha}\right)pF_{p,K-p+1}^{\alpha}\bar{B}+o\left(\frac{1}{K}\right)+o\left(\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right).
\end{aligned} \tag{28}$$

But

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

hence

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

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

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

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

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

then

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

where

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

and Θ_δ is independent of Ξ_δ. In addition, $\Theta_\delta \sim \chi_p^2\left(\delta^2\right)$ and $e_{\theta\delta}$ is uniformly distributed on the unit sphere $\mathfrak{S}_p(1)$. Using the same calculation as in the proof of Theorem 4, we have,

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

Plugging

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

we have

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

where we have used

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

and the last equality follows from straightforward calculations. ■

10.2 Appendix B: Proof of Extended Results

In this appendix, we make some simplifying assumptions and show that the optimal K formula remains valid for hypothesis testing based on a general M-estimator.

We consider the conventional large-K asymptotics under which $K \to \infty$ and $K/T \to 0$. We first assume that

$$H_T(\tilde{\theta}_T) = H_\infty(\theta_0) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

for any $\tilde{\theta}_T = \theta_0 + O_p\left(1/\sqrt{T}\right)$. This assumption holds if $\partial H_T(\theta)/\partial\theta'$ satisfies a uniform law of large numbers. Under this assumption, we have:

$$\sqrt{T}R\left(\hat{\theta}_T - \theta_0\right) = -RH_\infty^{-1}(\theta_0)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_t(\theta_0) + O_p\left(\frac{1}{\sqrt{T}}\right). \tag{29}$$

We further assume that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\left[\frac{\partial s_t(\theta_0)}{\partial\theta'} - E\frac{\partial s_t(\theta_0)}{\partial\theta'}\right] = O_p(1)$$

uniformly over $r \in [0,1]$. This assumption holds if the left hand side, as an empirical process, converges to a continuous time Gaussian process. The assumption can be rewritten as

$$\frac{1}{T}\sum_{t=1}^{[Tr]}\frac{\partial s_t(\theta_0)}{\partial\theta'} - rH_\infty(\theta_0) = O_p\left(\frac{1}{\sqrt{T}}\right)$$

uniformly over $r \in [0,1]$. Let

$$S_t(\theta) = \sum_{\tau=1}^{t} s_\tau(\theta) \text{ and } S_0(\theta) = 0,$$

then

$$\begin{aligned}
\frac{1}{\sqrt{T}}S_t\left(\hat{\theta}_T\right) &= \frac{1}{\sqrt{T}}S_t(\theta_0) + \left(\frac{1}{T}\sum_{\tau=1}^{t}\frac{\partial s_\tau\left(\bar{\theta}_T\right)}{\partial\theta'}\right)\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \\
&= \frac{1}{\sqrt{T}}S_t(\theta_0) - \left[\frac{1}{T}\sum_{\tau=1}^{t}\frac{\partial s_\tau(\theta_0)}{\partial\theta'}\right]\left[H_\infty^{-1}(\theta_0)\frac{1}{\sqrt{T}}S_T(\theta_0) + O_p\left(\frac{1}{\sqrt{T}}\right)\right] \\
&- \left[\sum_{i=1}^{d}\frac{1}{T}\sum_{\tau=1}^{t}\frac{\partial}{\partial\theta_i}\left[\frac{\partial s_\tau\left(\check{\theta}_T\right)}{\partial\theta'}\right]\left(\bar{\theta}_{Ti} - \theta_{0i}\right)\right]\left[H_\infty^{-1}(\theta_0)\frac{1}{\sqrt{T}}S_T(\theta_0) + O_p\left(\frac{1}{\sqrt{T}}\right)\right] \\
&= \frac{1}{\sqrt{T}}S_t(\theta_0) - \frac{t}{T}\frac{1}{\sqrt{T}}S_T(\theta_0) + O_p\left(\frac{1}{\sqrt{T}}\right)
\end{aligned}$$

where we have used

$$\frac{1}{T}\sum_{\tau=1}^{T}\left\|\frac{\partial}{\partial\theta_i}\frac{\partial s_\tau\left(\check{\theta}_T\right)}{\partial\theta'}\right\| = O_p(1) \text{ for any } i.$$

The preceding equation holds if $T^{-1}\sum_{\tau=1}^{T}\left\|\frac{\partial}{\partial\theta_i}\frac{\partial s_\tau(\theta)}{\partial\theta'}\right\|$ satisfies a ULLN. As a result,

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k(\frac{t}{T})s_t\left(\hat{\theta}_T\right)=\sum_{t=1}^{T}\left[\phi_k(\frac{t}{T})-\phi_k(\frac{t+1}{T})\right]\frac{S_t\left(\hat{\theta}_T\right)}{\sqrt{T}}\\
&=\sum_{t=1}^{T}\left[\phi_k(\frac{t}{T})-\phi_k(\frac{t+1}{T})\right]\frac{1}{\sqrt{T}}\left[S_t\left(\theta_0\right)-\frac{t}{T}S_T\left(\theta_0\right)\right]+O_p\left(\frac{1}{\sqrt{T}}\right)\\
&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k(\frac{t}{T})\left[s_t\left(\theta_0\right)-\frac{1}{T}S_T\left(\theta_0\right)\right]+O_p\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k(\frac{t}{T})\mathfrak{s}_t(\hat{\theta}_T,\hat{\theta}_T) &= RH_T^{-1}\left(\hat{\theta}_T\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k(\frac{t}{T})s_t\left(\hat{\theta}_T\right)\\
&= RH_\infty^{-1}\left(\theta_0\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k(\frac{t}{T})\left[s_t\left(\theta_0\right)-\frac{1}{T}S_T\left(\theta_0\right)\right]+O_p\left(\frac{1}{\sqrt{T}}\right)\\
&:= R\Lambda_k+O_p\left(\frac{1}{\sqrt{T}}\right)
\end{aligned}$$

where

$$\Lambda_k=H_\infty^{-1}\left(\theta_0\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k(\frac{t}{T})\left[s_t\left(\theta_0\right)-\frac{1}{T}S_T\left(\theta_0\right)\right].$$

Now

$$\begin{aligned}
F_T&=\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}RH_T^{-1}(\tilde{\theta}_T)^{-1}s_t\left(\theta_0\right)\right]'\tilde{\Omega}^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}RH_T^{-1}(\tilde{\theta}_T)s_t\left(\theta_0\right)\right]\\
&=\left[RH_\infty^{-1}\left(\theta_0\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_t\left(\theta_0\right)+O_p\left(\frac{1}{\sqrt{T}}\right)\right]'\\
&\times\left\{\frac{1}{K}\sum_{k=1}^{K}\left[R\Lambda_k+O_p\left(\frac{1}{\sqrt{T}}\right)\right]\left[R\Lambda_k+O_p\left(\frac{1}{\sqrt{T}}\right)\right]'\right\}^{-1}\\
&\times\left[RH_\infty^{-1}\left(\theta_0\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_t\left(\theta_0\right)+O_p\left(\frac{1}{\sqrt{T}}\right)\right]\\
&=F_{LT}+\frac{1}{\sqrt{T}}\psi_T+\left[\left(\frac{1}{\sqrt{K}}+\frac{K^2}{T^2}\right)\left(\frac{1}{\sqrt{T}}\right)\right]\psi_T^*
\end{aligned}$$

where ψ_T does not depend on K, both ψ_T and ψ_T^* are $O_p(1)$, and

$$F_{LT}=\left[RH_\infty^{-1}\left(\theta_0\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_t\left(\theta_0\right)\right]'\left\{R\hat{\Omega}_LR'\right\}^{-1}\left[RH_\infty^{-1}\left(\theta_0\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_t\left(\theta_0\right)\right],$$

for

$$\hat{\Omega}_L=\frac{1}{K}\sum_{k=1}^{K}\Lambda_k\Lambda_k'.$$

Let

$$\tilde{Y}_t = H_\infty^{-1}(\theta_0)\, s_t(\theta_0) := A_0 + u_t,$$

then

$$F_{LT} = \left[R \frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_t \right]' \left\{ R \tilde{\Omega}_L R' \right\}^{-1} \left[R \frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_t \right].$$

So F_{LT} is the Wald statistic for testing whether the mean of the process $\tilde{Y}_t = H_\infty^{-1}(\theta_0)\, s_t(\theta_0)$ satisfies $R\left(E\tilde{Y}_t\right) = 0$. In other words, F_{LT} is in exactly the same form as the statistic F_T defined in (7). Its finite sample corrected version $(K-p+1)/(pK) \times F_{LT}$ is exactly the same as the statistic F_T^* defined in (13). Therefore, if u_t satisfies Assumption 1, we have, under the null hypothesis,

$$P\left(\frac{K-p+1}{(pK)} F_{LT} > F_{p,K-p+1}^{\alpha} \right) = \alpha - \frac{K^2 \bar{B}_L}{T^2} G_p'\left(\chi_p^\alpha\right) \chi_p^\alpha + o\left(\frac{K^2}{T^2} + \frac{1}{K} \right).$$

Consequently,

$$P\left(F_T^* > F_{p,K-p+1}^{\alpha} \right) = \alpha - \frac{K^2 \bar{B}_L}{T^2} G_p'\left(\chi_p^\alpha\right) \chi_p^\alpha + o\left(\frac{K^2}{T^2} + \frac{1}{K} \right) + O\left(\frac{1}{\sqrt{T}} \right).$$

where the $O\left(1/\sqrt{T}\right)$ term does not depend on K.

Next, we consider the alternative hypothesis

$$H_1\left(\delta_o^2\right) : R\theta = r + \frac{\left(R H_\infty^{-1}(\theta_0)\, \Omega^{-1} H_\infty^{-1}(\theta_0)\, R' \right)^{1/2} \tilde{c}}{\sqrt{T}}$$

where $\tilde{c}$ is uniformly distributed on the sphere $\mathfrak{S}_p(\delta)$. Using similar arguments, we have

$$\begin{aligned} & P\left(F_T^* > p F_{p,K-p+1}^{\alpha} | H_1\left(\delta^2\right) \right) \\ & = 1 - G_{p,\delta^2}\left(\chi_p^\alpha\right) - \frac{K^2}{T^2} G_{p,\delta^2}'\left(\chi_p^\alpha\right) \chi_p^\alpha \bar{B} - \frac{1}{K} \frac{\chi_p^\alpha \delta^2}{2} G_{(p+2),\delta^2}'\left(\chi_p^\alpha\right) \\ & + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{\sqrt{T}}\right) \end{aligned} \tag{30}$$

where again the $O(1/\sqrt{T})$ term does not depend on K.

We collect our results in the theorem below.

Theorem 6 *Assume (i) $\rho(\theta, Z_T)$ is three times continuously differentiable in θ (ii) $plim_{T\to\infty} \hat{\theta}_T = \theta_0$, (iii) for sufficiently large C, $P(|\psi_T| < C$ and $|\psi_T^*| < C) = 1 + O(1/T)$, (iv) $Y_t = H_\infty^{-1}(\theta_0)\, s_t(\theta_0)$ satisfies Assumption 1, (v) $T^{-1} S_{[Tr]}\left(\tilde{\theta}_T\right) - r H_\infty(\theta_0) = O_p\left(1/\sqrt{T}\right)$ uniformly over $r \in [0,1]$ for any $\tilde{\theta}_T$ between $\hat{\theta}_T$ and θ_0, (vi) $H_T(\tilde{\theta}_T) = H_\infty + O_p\left(1/\sqrt{T}\right)$ for any $\tilde{\theta}_T = \theta_0 + O_p\left(1/\sqrt{T}\right)$ (vii) $sup_{\theta\in\Theta} T^{-1} \sum_{\tau=1}^{T} \left\| \frac{\partial}{\partial\theta_i} \frac{\partial s_\tau(\theta)}{\partial\theta'} \right\| = O_p(1)$ for any i. Let*

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

$$B = -\frac{2\pi^2}{3} \sum_{j=-\infty}^{\infty} |j|^2 s_t(\theta_0)\, s_{t-j}'(\theta_0).$$

If $K \to \infty$ such that $K/T \to 0$, then

(i) the type I error of the F^* test is

$$e_I(K) = \alpha - \frac{K^2 \bar{B}_L}{T^2} G_p' \left(\chi_p^\alpha\right) \chi_p^\alpha + o\left(\frac{K^2}{T^2} + \frac{1}{K}\right) + O\left(\frac{1}{\sqrt{T}}\right),$$

where the $O(1/\sqrt{T})$ term does not depend on K.

(ii) The average type II error of the F^* test under $H_1\left(\delta^2\right)$ is

$$\begin{aligned} e_{II}(K) = {} & G_{p,\delta^2}\left(\chi_p^\alpha\right) + \frac{K^2}{T^2} G'_{p,\delta^2}\left(\chi_p^\alpha\right) \chi_p^\alpha \bar{B} + \frac{1}{K} \frac{\chi_p^\alpha \delta^2}{2} G'_{(p+2),\delta^2}\left(\chi_p^\alpha\right) \\ & + o\left(\frac{K^2}{T^2} + \frac{1}{K}\right) + O\left(\frac{1}{\sqrt{T}}\right) \end{aligned}$$

where the average is over the sphere $\tilde{c} \in \mathfrak{S}_p(\delta)$ and the $O(1/\sqrt{T})$ term does not depend on K.

Theorem 6 shows that, if we ignore the terms that do not depend on K, the approximate type I and type II errors are exactly the same as those for Gaussian location models. Hence the optimal K formula in (19) remains valid in the general setting.

References

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

[2] Albers, W. (1978): "Testing the Mean of a Normal Population Under Dependence." *Annals of Statistics* 6, 1337–1344.

[3] Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59, 817–854.

[4] Christensen, J.H., F. Diebold, G. D. Rudebusch and G. Strasser (2008): "Multivariate Comparison of Predictive Accuracy." Working paper, Department of Economics, University of Pennsylvania.

[5] Conway, R. W. (1963): "Some Tactical Problems in Digital Simulation." *Management Science* 10, 47–61.

[6] Diebold, F. X. and R.S. Mariano (1995): "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 13, 253–263.

[7] Elliott, G., T.J. Rothenberg and J.H. Stock (1996): "Efficient Tests for an Autoregressive Unit Root." *Econometrica* 64, 813–836.

[8] Fishman, G.S. (1978): *Principles of Discrete Event Simulation*, Wiley, New York.

[9] Foley, R.D. and D. Goldman (1999): "Confidence Intervals Using Orthonormally Weighted Standardized Time Series." *ACM Transactions on Modeling and Computer Simulation,* 19(4), 297–325.

[10] Grenander, U. and M. Rosenblatt (1957): *Statistical Analysis of Stationary Time Series.* Wiley, New York.

[11] Hashimzade, N. and T. J. Vogelsang (2007): "Fixed-b Asymptotic Approximation of the Sampling Behavior of Nonparametric Spectral Density Estimators." *Journal of Time Series Analysis* 29, 142–162.

[12] Hannan, E. J. (1957). "The Variance of the Mean of a Stationary Process." *Journal of the Royal Statistical Society* Series B, 19, 282–285

[13] Heidelberger, P. and P. D. Welch (1981): "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations." *Communications of the ACM* 24, 233–245

[14] Hotelling H. (1931): "The Generalization of Student's Ratio." *Annals of Mathematical Statistics* 2, 360–378.

[15] Ibragimov, R. and U.K. Müller (2010): "t-Statistic Based Correlation and Heterogeneity Robust Inference." *Journal of Business and Economic Statistics*, 28, 453–468.

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

[17] Kabaila, P. and G. Nelson (1985): "On the Confidence Regions for the Mean of a Multivariate Time Series." *Communication in Statistics – Theory and Methods* 14, 735–753.

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

[19] ——— (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests." *Econometric Theory* 21, 1130–1164.

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

[21] Neyman, J. (1937): "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability." *Philosophical Transactions of the Royal Society of London*, 236, 333–380

[22] Ng, S. and P. Perron (1994): "The Exact Error in Estimating of the Spectral Density at the Origin." *Journal of Time Series Analysis* 17, 379–408.

[23] Neave, H. R. (1970): "An Improved Formula for the Asymptotic Variance of Spectrum Estimates." *Annals of Mathematical Statistics* 41, 70–77.

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

[25] ——— (1994): "Automatic Lag Selection in Covariance Estimation." *Review of Economic Studies* 61, 631–654.

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

[27] Phillips, P. C. B. (2006): "Optimal Estimation of Cointegrated Systems with Irrelevant Instruments", Cowles Foundation Discussion Paper # 1547.

[28] 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.

[29] ——— (2007): "Long-Run Variance Estimation and Robust Regression Using Sharp Origin Kernels with No Truncation." *Journal of Statistical Planning and Inference* 137, 985–1023.

[30] Percival D. and A. T. Walden (1993): *Spectral Analysis For Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge University Press.

[31] Stoica, P. and R. Moses (2005): *Spectral Analysis of Signals*, Pearson Prentice Hall.

[32] Sun, Y. and P. C. B. Phillips (2008): "Optimal Bandwidth Choice for Interval Estimation in GMM Regression." Working paper, Department of Economics, UC San Diego.

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

[34] Sun, Y., P. C. B. Phillips, and S. Jin (2010): "Power Maximization and Size Control in Heteroskedasticity and Autocorrelation Robust Tests with Exponentiated Kernels." Forthcoming, *Econometric Theory.*

[35] 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.

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

[37] Sun, Y. (2011b): "Asymptotic F Test in a GMM Framework." Working paper, Department of Economics, UC San Diego.

[38] Thomson, D. J. (1982): "Spectrum Estimation and Harmonic Analysis." *IEEE Proceedings*, 70, 1055–1096.

[39] Velasco, C. and P. M. Robinson (2001): "Edgeworth Expansions for Spectral Density Estimates and Studentized Sample Mean." *Econometric Theory* 17, 497–539.