Heteroscedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator

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

March 11, 2012

*Email: yisun@ucsd.edu. The paper replaces a working paper entitled "Autocorrelation Robust Inference Using Series Variance Estimator and Testing-optimal Smoothing Parameter." The author gratefully acknowledges partial research support from NSF under Grant No. SES-0752443. He thanks Xiaohong Chen, Chung-ming Kuan, Peter Phillips 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 heteroscedasticity and autocorrelation robust test in a time series setting. The test is based on a series long run variance matrix estimator that involves projecting the time series of interest onto a set of orthonormal bases and using the sample variance of the projection coefficients as the long run variance estimator. When the number of orthonormal bases K is fixed, a finite sample corrected Wald statistic converges to a standard F distribution. When K grows with the sample size, the usual Wald statistic converges to a chi-square distribution. We show that critical values from the F distribution are second-order correct under the conventional increasing-smoothing asymptotics. Simulations show that the F approximation is more accurate than the chi-square approximation in finite samples.

JEL Classification: C13; C14; C32; C51

Keywords: Asymptotic expansion, F-distribution, Fixed-smoothing asymptotics, Increasing-smoothing asymptotics, long run variance, robust standard error, nonparametric series method.

1 Introduction

The paper considers heteroscedasticity and autocorrelation robust inference in a time series setting. In the presence of nonparametric autocorrelation, it is standard practice to use a kernel-based method to estimate the long run variance (LRV) of the moment process, e.g., Newey and West (1987) and Andrews (1991). Recent research on kernel LRV estimation has developed new and more accurate asymptotic approximations to the associated test statistics. See, for example, Kiefer and Vogelsang (2005), Jansson (2004), Sun and Phillips and Jin (2008, SPJ hereafter). The new asymptotic approximation is obtained under the limiting thought experiment where the amount of nonparametric smoothing is held fixed. However, the new approximation has not been widely adopted in empirical studies. A possible reason is that the new approximation is not standard and critical values have to be simulated. In this paper, we employ a nonparametric series method for LRV estimation. A great advantage of this method is that the asymptotic distribution of the associated Wald statistic is standard regardless of the limiting thought experiments we use. So more accurate inference can be conducted without any extra computational cost.

The basic idea behind the series LRV estimation is to project the moment process 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. It can also be regarded as the sample variance of the projection coefficients. By construction, the series LRV estimator is automatically positive semidefinite, a desirable property for practical use of this estimator.

The smoothing parameter in the series LRV estimator is the number of basis functions K employed. Depending on the asymptotic specifications on K, there are two types of asymptotics: the fixed-smoothing asymptotics under which K is assumed to be fixed and the increasing-smoothing asymptotics under which K is assumed to grow with the sample size but at a slower rate. These two types of asymptotics correspond to the fixed-b asymptotics and the small-b asymptotics for kernel LRV estimation, as introduced in Kiefer and Vogelsang (2005). By capturing the estimation uncertainty of the LRV estimator, the fixed-smoothing asymptotics is often more accurate than the increasing-smoothing asymptotics in finite samples.

To obtain a standard limiting distribution under the fixed-smoothing asymptotics, 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 normalized Wishart distribution. As a result, under the fixed-smoothing asymptotics, a modified Wald statistic converges to a standard F distribution. This is in contrast with the conventional increasing-smoothing asymptotics under which the Wald statistic converges to a chi-square distribution.

In finite samples and in practical situations, we often choose the smoothing parameter K to reflect the temporary dependence and the sample size. In general, K increases with the sample size. This practice is more compatible with the increasing-smoothing asymptotics. To justify the use of the fixed-smoothing asymptotics in finite samples, we establish a high order expansion of the modified Wald statistic under the increasing-smoothing asymptotics. Using the high order expansion, we show that critical values from the fixed-smoothing limiting distribution, i.e. the F distribution, are second-order correct under the increasing-smoothing asymptotics. A direct implication is that we can use fixed-smoothing critical values regardless whether the finite sample situation is more compatible with the increasing-smoothing asymptotics or not.

On the basis of the high order expansion, we obtain an approximate measure of the coverage probability error (CPE) of a finite sample corrected Wald confidence region. We select the smoothing parameter K to minimize the CPE. Our proposed CPE-optimal K is different from the MSE-optimal K as given in Phillips (2005) in terms both the orders of magnitude and the constant coefficients.

Some discussions of this paper's contributions relative to the author's other work are in order. Sun (2011a) employs the series LRV estimator for OLS trend inference in a simple linear trend model, while the focus here is on general stationary time series and M estimation. So the present paper complements Sun (2011a). The idea of CPE-optimal smoothing parameter choice was used in the unpublished working papers Sun and Phillips (2008) and Sun (2011b) but these two papers focus on kernel LRV estimation and inference. In addition, the present paper relaxes the functional CLT assumption, which is typically maintained in the literature on the fixed-smoothing asymptotics, to a multivariate CLT. This can be viewed as a technical advantage of using the series LRV estimator.

The rest of the paper is organized as follows. Section 2 describes the problem at hand and provides an overview of the series LRV estimator. Section 3 investigates the limiting behavior of the Wald statistic under the fixed-smoothing asymptotics. Section 4 gives a higher order distributional expansion of the modified Wald statistic under the conventional increasing-smoothing asymptotics. On the basis of this expansion, we consider selecting K to minimize the coverage probability error in Section 5. The subsequent section reports simulation evidence. The last section provides some concluding discussion. Proofs of the main results and their extensions are given in the Appendix.

2 Basic Setting and Series LRV Estimator

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

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

where Θ is a compact parameter space, $\left\{Z_t \in \mathbb{R}^{d_z}\right\}_{t=1}^{T}$ are the time series observations, and $\rho(\theta, Z_t)$ is a twice continuously differentiable criterion function. To achieve identification, we assume that $\theta = \theta_0$ is the unique minimizer of $Q(\theta) = E\rho(\theta, Z_t)$ over Θ. Define

$$s_t(\theta) = \frac{\partial \rho(\theta, Z_t)}{\partial \theta} \text{ and } H_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial s_t(\theta)}{\partial \theta'}.$$

Then under some additional regularity conditions, we have

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) = -H_T^{-1}\left(\tilde{\theta}_T\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_t(\theta_0),$$

where $H_T(\tilde{\theta}_T)$ is defined to be

$$H_T\left(\tilde{\theta}_T\right) := T^{-1}\sum_{t=1}^{T}\frac{\partial s_t\left(\tilde{\theta}_T\right)}{\partial \theta'},\ \frac{\partial s_t\left(\tilde{\theta}_T\right)}{\partial \theta'} := \left[\frac{\partial s_t\left(\tilde{\theta}_{T1}\right)}{\partial \theta_1}, ..., \frac{\partial s_t\left(\tilde{\theta}_{Td}\right)}{\partial \theta_d}\right]$$

with $\tilde{\theta}_{Ti} = \theta_0 + \lambda_{Ti}\left(\hat{\theta}_T - \theta_0\right)$ for some $\lambda_{Ti} \in [0,1]$ and $i = 1, 2, ..., d$.

Suppose we want to test the null hypothesis that $\mathcal{H}_0 : R\theta_0 = r$ against $\mathcal{H}_1 : R\theta_0 \neq r$ for some $p \times d$ matrix R with full rank. Under the null hypothesis, we have

$$\sqrt{T}\left(R\hat{\theta}_T - r\right) = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t\left(\tilde{\theta}_T, \theta_0\right)$$

where

$$u_t\left(\theta_1, \theta_2\right) = RH_T^{-1}\left(\theta_1\right) s_t\left(\theta_2\right).$$

Invoking a CLT, we can obtain: $\sqrt{T}\left(R\hat{\theta}_T - r\right) \xrightarrow{d} N(0, \Omega)$ where $\Omega = \Lambda\Lambda' = \sum_{j=-\infty}^{\infty}\Gamma(j)$,

$$\Gamma(j) = RH^{-1}\left[Es_t\left(\theta_0\right)s_{t-j}'\left(\theta_0\right)\right]H^{-1}R' \text{ and } H = EH_T\left(\theta_0\right).$$

The above asymptotic result provides the usual basis for the Wald test. To implement the test, we need to first estimate the long run variance matrix Ω. Many nonparametric estimation methods are available in the econometrics literature. Most LRV estimators use kernel-smoothing methods that involve taking a weighted sum of sample autocovariances. In this paper, following Phillips (2005, 2006), Sun (2006, 2011a), and Müller (2007), we consider a nonparametric series method that involves projecting the time series onto orthonormal basis functions.

Let $\{\phi_k\}_{k=1}^{\infty}$ be a sequence of orthonormal basis functions in $L_2[0,1]$ and $\hat{u}_t = u_t(\hat{\theta}_T, \hat{\theta}_T)$. Define the sample 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 LRV estimator:

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

where K is the number of basis functions used.

The series LRV estimator has different interpretations. It can be regarded as a multiple-window estimator with window function $\phi_k(t/T)$, see Thomson (1982). It also belongs to the class of filter-bank estimators and $\hat{\Omega}$ is a simple average of the individual filter-bank estimators $\hat{\Omega}_k$. For more discussions along this line, see Stoica and Moses (2005, ch. 5). Furthermore, the series LRV 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

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

So the series LRV 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. Sun (2011a) provides additional motivation for the series LRV estimator.

Under some conditions, it is not hard to show that each of the direct LRV estimator $\hat{\Omega}_k$ is an asymptotically unbiased but inconsistent estimator of Ω. As a result, $\hat{\Omega}$ is inconsistent when K is fixed. However, the asymptotic variance of $\hat{\Omega}$ decreases as K increases. So as $K \to \infty$, $\hat{\Omega}$ becomes consistent. These properties of the series LRV estimator are analogous to those of kernel LRV estimators.

3 Fixed-smoothing Asymptotics

The usual Wald statistic F_T for testing linear hypothesis $\mathcal{H}_0$ against $\mathcal{H}_1$ is given by

$$F_T = \sqrt{T}\left(R\hat{\theta}_T - r\right)' \hat{\Omega}^{-1} \sqrt{T}\left(R\hat{\theta}_T - r\right). \tag{1}$$

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

$$t_T = \frac{\sqrt{T}\left(R\hat{\theta}_T - r\right)}{\sqrt{\hat{\Omega}}}.$$

Our results extend straightforwardly to nonlinear hypotheses.

Let $\phi_0(x) = 1$ be the constant function. To establish the large sample asymptotic distributions of F_T and t_T, we maintain the following assumption.

Assumption 1 *(i) $\rho(\theta, Z_t)$ is twice continuously differentiable, $\hat{\theta}_T = \theta_0 + o_p(1)$, and θ_0 is an interior point of Θ.*

(ii) Uniformly in r

$$\left\| \frac{1}{T}\sum_{t=1}^{[rT]} \frac{\partial s_t\left(\tilde{\theta}_T\right)}{\partial \theta'} - rH \right\| \to^p 0$$

for nonsingular matrix $H = EH_T(\theta_0)$ where $\|\cdot\|$ is the Frobenius norm.

(iii) The following CLT holds:

$$RH^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \phi_k\left(\frac{t}{T}\right) s_t(\theta_0) \xrightarrow{d} \Lambda \xi_k$$

jointly for $k = 0, ..., K$ where $\xi_k \sim iidN(0, \mathbb{I}_p)$.

(iv) $\{\phi_k(x)\}_{k=1}^K$ is a sequence of continuously differentiable and orthonormal basis functions on $L_2[0,1]$ satisfying $\int_0^1 \phi_k(x)\,dx = 0$.

The consistency result in Assumption 1(i) can be verified under more primitive assumptions and using standard textbook arguments. Assumption 1(ii) is standard in the literature on the fixed-smoothing asymptotics, e.g. see Kiefer and Vogelsang (2005, Assumption 3). Assumption 1(iii) is a multivariate CLT. It is much weaker than the functional CLT (FCLT) assumption that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} s_t(\theta_0) \xrightarrow{d} W_d(r)$$

where $W_d(r)$ is a Brownian motion process. The above FCLT is often assumed for deriving the fixed-smoothing asymptotics. See for example Kiefer and Vogelsang (2005, Assumption 2) and Sun and Kim (2012, Assumption 3). It is easy to see that the FCLT and Assumption 1(iv) imply Assumption 1(iii).

There are many basis functions that satisfy Assumption 1(iv). For example, $\phi_k(x) = \sqrt{2}\sin 2\pi kx$ and $\phi_k(x) = \sqrt{2}\cos 2\pi kx$. However, the function $\phi_k(x) = \sqrt{2}\sin\pi(k-0.5)x$ used in Phillips (2005) and Müller (2007) does not satisfy Assumption 1(iv) as it does not have zero 'mean' in that $\int_0^1 \phi_k(x)\,dx \neq 0$. The zero mean condition ensures that $T^{-1/2}\sum_{t=1}^T \phi_k(t/T)\,s_t(\theta_0)$ and $T^{-1/2}\sum_{t=1}^T \phi_k(t/T)\,s_t(\hat{\theta}_T)$ have the same limiting distribution. That is, the estimation uncertainty in $\hat{\theta}_T$ can be ignored in large samples. Without this assumption, the limiting distributions of $\hat{\Omega}$ and F_T may be nonstandard.

Theorem 1 *Let Assumption 1 hold. Then for $K \geq p$,*

(i) $K\Lambda^{-1}\hat{\Omega}(\Lambda')^{-1} \xrightarrow{d} \mathbb{W}_p(\mathbb{I}_p, K) \overset{d}{=} \sum_{k=1}^K \xi_k\xi_k'$, $\xi_k \sim iidN(0, \mathbb{I}_p)$.

(ii)

$$\frac{(K-p+1)}{pK}F_T \xrightarrow{d} F_{p,K-p+1} = \frac{\chi_p^2/p}{\chi_{K-p+1}^2/(K-p+1)}$$

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

By definition, $\mathbb{W}_p(\mathbb{I}_p, K)$ is a Wishart distribution. In the scalar case with $d = 1$, the limiting distribution of $K\Lambda^{-1}\hat{\Omega}(\Lambda')^{-1}$ reduces to the chi-square distribution χ_K^2. In addition, for any conforming constant vector z, we have

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

where $\tilde{z} = \Lambda'z$. That is, $z'\hat{\Omega}z/z'\Omega z$ converges weakly to a normalized 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 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.

The fixed-smoothing asymptotics of F_T can be rewritten as

$$F_T \xrightarrow{d} \frac{\chi_p^2}{\chi_{K-p+1}^2/(K-p+1)}\frac{K}{K-p+1}$$

where χ_p^2 and χ_{K-p+1}^2 are independent χ^2 variates. As $K \to \infty$, the fixed-smoothing asymptotic distribution approaches the usual chi-square distribution χ_p^2. However, when K is fixed, critical values based on the fixed-smoothing asymptotics are larger than those based on the usual chi-square approximation. This is because 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. More rigorously, let $G_p(\cdot)$ be the CDF of a χ^2 random variable with degrees of freedom p and $\mathcal{X}_p^\alpha$ be the α-level critical value such that $1 - G_p\left(\mathcal{X}_p^\alpha\right) = \alpha$. Then for typical α used in empirical applications:

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

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

When $p=1$, the above result reduces to $t_T \xrightarrow{d} t_K$. That is, the t-statistic converges to the t distribution with K degrees of freedom. The asymptotic t-distribution theory has appeared in the literature. Ibragimov and Müller (2010) employ a closely related method and establish the robustness of t-approximation to variance heterogeneity.

In the special case when $\rho(\theta, Z_t)=\left\|Z_t-\theta\right\|^2$, we have $\theta_0=EZ_t$. By selecting R and r appropriately, we can obtain the null hypothesis $\mathcal{H}_0:\theta_{01}=\theta_{02}=...=\theta_{0d}$. So the problem reduces to testing whether the means are the same across different time series. Our result can be viewed as an extension of Hotelling's T^2 test for iid data to the time series setting. We recover the asymptotic T^2 distribution but with different degrees of freedom that reflect the time series dependence. An application of multivariate mean comparison is the equal predictability test. 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 mean of the loss series are the same. Diebold and Mariano (1995) consider the case with two forecasts while Christensen, Diebold, Rudebusch and Strasser (2008) extend it to general multivariate scenarios.

Under the local alternative hypothesis,

$$
\mathcal{H}_1\left(\delta^2\right):R\theta=r+c/\left(\sqrt{T}\right) \text{ where } c=\Omega^{1/2}\tilde{c} \tag{2}
$$

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

$$
\begin{aligned}
& \frac{\left(K-p+1\right)}{pK}F_T \xrightarrow{d} \frac{\left(K-p+1\right)}{p}\left(\xi_0+\tilde{c}\right)'\left(\sum_{k=1}^{K}\xi_k\xi_k'\right)^{-1}\left(\xi_0+\tilde{c}\right) \\
& =^d F_{p,K-p+1}\left(\delta^2\right),
\end{aligned}
$$

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

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

Similarly, the t-statistic converges to the noncentral t distribution with degrees of freedom K and noncentrality parameter $\delta=c/\Omega^{1/2}=\tilde{c}$.

4 High Order Expansion under Increasing-smoothing Asymptotics

In this section, we establish a high order expansion of the Wald statistic under the increasing-smoothing asymptotics under which $K \to \infty$ and $K/T \to 0$. Using the high order expansion, we show that the fixed-smoothing approximation is second-order correct under the increasing-smoothing asymptotics.

To simplify the presentation and for the sake of clarity, we consider the special case when R is the identity matrix $\mathbb{I}_p$ with $p = d$. That is, we are interested in the hypothesis about the whole parameter vector. Since $Es_t(\theta) = 0$ if and only if $\theta = \theta_0$, the null hypothesis $\mathcal{H}_0 : \theta = \theta_0$ is equivalent to the hypothesis that the multivariate process $s_t(\theta_0)$ has mean zero. In view of this equivalence, we can construct the $\tilde{F}_T$ statistic as follows

$$\tilde{F}_T = \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_t(\theta_0) \right]' \tilde{\Omega}^{-1} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_t(\theta_0) \right] \tag{3}$$

where $\tilde{\Omega} = K^{-1} \sum_{k=1}^{K} \tilde{\Lambda}_k \tilde{\Lambda}_k'$ for

$$\tilde{\Lambda}_k = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \phi_k (\frac{t}{T}) s_t(\theta_0).$$

The above test statistic $\tilde{F}_T$ is the same as F_T introduced before except that we do not need to estimate the Hessian matrix under the null. Define the finite sample corrected statistic

$$\tilde{F}_T^* = \frac{K - p + 1}{K} \tilde{F}_T.$$

Theorem 2 *Assume that (i) $s_t(\theta_0)$ is a stationary Gaussian process; (ii) For any $\kappa \in \mathbb{R}^d$, the spectral density of $\kappa' s_t(\theta_0)$ is twice continuously differentiable and is bounded above and away from zero in a neighborhood around the origin. If $K \to \infty$ such that $K/T \to 0$, then*

$$\begin{aligned} &P\left(\tilde{F}_T^* < z\right) \\ &= G_p(z) + \frac{K^2}{T^2} G_p'(z) z \bar{B} + \frac{1}{K} G_p''(z) z^2 + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{1}{T}\right) \end{aligned} \tag{4}$$

where $B = E\tilde{\Omega} - \Omega$, $\bar{B} = tr\left(B\Omega^{-1}\right)/p$, and Ω is the LRV of $s_t(\theta_0)$.

In the above theorem, we have followed SPJ (2008) and made the simplification assumption that $s_t(\theta_0)$ is normal. The normality assumption allows us to decompose $(1/\sqrt{T}) \sum_{t=1}^{T} s_t(\theta_0)$ into two parts: one part that is independent of $\tilde{\Omega}$ and the other part that is stochastically small. The main terms in the high order expansion are driven by the first part. The normality assumption could be relaxed but at the cost of much greater complexity and tedious calculations. The expansion under non-normality is also much hard to understand. It contains high order terms that cannot be reliably estimated in finite samples. See, for example, Velasco and Robinson (2001) and Sun and Phillips (2008).

The high order expansion in Theorem 2 has the same form as the expansion in Sun (2011a). The first term in (4) comes from the standard chi-square approximation of the

Wald statistic. The second term captures the nonparametric bias of the LRV estimator. While B is the bias of $\tilde{\Omega}$, $\bar{B}$ can be regarded as the relative bias as it summarizes the bias matrix relative to the true matrix Ω. When $p=1$, $\bar{B}=B/\Omega$, which is the percentage of the bias. The third term in (4) reflects the variance of the LRV estimator. It decreases as K increases.

With some abuse of notation, we use $F_{p,K-p+1}$ to denote a random variable with the distribution $F_{p,K-p+1}$. It is straightforward to show that

$$P\left(pF_{p,K-p+1}<z\right)=EG_p\left(z\frac{\chi^2_{K-p+1}}{K-p+1}\right)=G_p\left(z\right)+\frac{1}{K}G_p''\left(z\right)z^2+o\left(\frac{1}{K}\right). \tag{5}$$

So the fixed-smoothing asymptotics agrees with the higher-order increasing-smoothing asymptotics up to the order of $O\left(1/K\right)$. This implies that $pF^{\alpha}_{p,K-p+1}$ is high order correct under the conventional increasing-smoothing asymptotics.

Let $\mathcal{F}^{\alpha}_{p,K-p+1}$ be the α-level critical value of the $F_{p,K-p+1}$ distribution, i.e.

$$P\left(pF_{p,K-p+1}<p\mathcal{F}^{\alpha}_{p,K-p+1}\right)=1-\alpha.$$

In view of (5), we have

$$G_p\left(p\mathcal{F}^{\alpha}_{p,K-p+1}\right)+\frac{1}{K}G_p''\left(p\mathcal{F}^{\alpha}_{p,K-p+1}\right)\left(p\mathcal{F}^{\alpha}_{p,K-p+1}\right)^2=1-\alpha+o\left(\frac{1}{K}\right),$$

as $K\to\infty$. It then follows that

$$\begin{aligned}
&P\left(\tilde{F}^*_T<p\mathcal{F}^{\alpha}_{p,K-p+1}\right)\\
&=G_p\left(p\mathcal{F}^{\alpha}_{p,K-p+1}\right)+\frac{K^2}{T^2}G_p'\left(p\mathcal{F}^{\alpha}_{p,K-p+1}\right)p\mathcal{F}^{\alpha}_{p,K-p+1}\bar{B}+\frac{1}{K}G_p''\left(p\mathcal{F}^{\alpha}_{p,K-p+1}\right)\left(p\mathcal{F}^{\alpha}_{p,K-p+1}\right)^2\\
&+o\left(\frac{1}{K}\right)+o\left(\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right)\\
&=1-\alpha+\frac{K^2}{T^2}G_p'\left(\mathcal{X}^{\alpha}_p\right)\mathcal{X}^{\alpha}_p\bar{B}+o\left(\frac{1}{K}\right)+o\left(\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right) \qquad (6)
\end{aligned}$$

where we have used $p\mathcal{F}^{\alpha}_{p,K-p+1}=\mathcal{X}^{\alpha}_p+o(1)$ as $K\to\infty$. So using the critical value $p\mathcal{F}^{\alpha}_{p,K-p+1}$ instead of the chi-square critical value $\mathcal{X}^{\alpha}_p$ removes the variance term in the high order expansion given by Theorem 2. When K is not too large, using the F critical value should lead to a test with more accurate size.

Theorem 2 gives an expansion of the distribution of $K^{-1}\left(K-p+1\right)\tilde{F}_T$. The factor $K^{-1}\left(K-p+1\right)$ is a finite sample correction that can be interpreted as a Bartlett type correction. See Sun (2011b) for more details. Alternatively, we can regard $\tilde{F}_T$ as the standard Wald statistic but using the following estimator for Ω :

$$\check{\Omega}=\frac{1}{K-p+1}\sum_{k=1}^{K}\tilde{\Lambda}_k\tilde{\Lambda}_k'.$$

So the finite sample correction factor $\left(K-p+1\right)/K$ can be viewed as a degree-of-freedom adjustment. This adjustment is small when p is small but it can be large when p is large or K is relatively small.

Theorem 2 applies to statistical inference that involves the whole parameter vector θ_0. Suppose we go back to the general testing problem with the null $\mathcal{H}_0 : R\theta_0 = r$ and the alternative $\mathcal{H}_1 : R\theta_0 \neq r$. A conservative approach to inference is to project the confidence region for the whole parameter vector θ_0 onto the subspace $R\theta_0$ to obtain

$$\Theta_R^\alpha = \left\{R\theta_0 : \text{for } \theta_0 \text{ such that } F_T^* \leq p\mathcal{F}_{p,K-p+1}^\alpha\right\}$$

and reject the null if the confidence set Θ_R^α does not contain r. Here we follow the more conventional approach by constructing the F_T statistic as in (1). It is easy to see that

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

So the F_T statistic can be viewed as a feasible Wald statistic for testing whether the mean of the multivariate process $RH^{-1}s_t(\theta_0)$ is zero.

In Appendix B, we show that, under some high level conditions, the high order expansion in (4) remains valid for

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

except that there is an additional approximation error of order $\log T/\sqrt{T}$ that does not depend on K. So the second-order correctness of the fixed-smoothing approximation holds in more general settings.

5 CPE-Optimal K Choice

Suppose we are interested in the whole parameter vector. If we construct the confidence region of the form $\left\{\theta_0 : F_T^* \leq \mathcal{X}_p^\alpha\right\}$, then up to small order terms its (absolute) coverage probability error (CPE) is

$$\begin{aligned} CPE &= \left|P\left(F_T^* \leq \mathcal{X}_p^\alpha\right) - (1-\alpha)\right| \\ &= \left|\frac{K^2}{T^2}G_p'\left(\mathcal{X}_p^\alpha\right)\mathcal{X}_p^\alpha\bar{B} + \frac{1}{K}G_p''\left(\mathcal{X}_p^\alpha\right)\left(\mathcal{X}_p^\alpha\right)^2\right| \\ &\leq \frac{K^2}{T^2}G_p'\left(\mathcal{X}_p^\alpha\right)\mathcal{X}_p^\alpha\left|\bar{B}\right| + \frac{1}{K}\left|G_p''\left(\mathcal{X}_p^\alpha\right)\right|\left(\mathcal{X}_p^\alpha\right)^2. \end{aligned}$$

To control the coverage probability error, we can choose K to minimize the above upper bound for the CPE, leading to

$$K_{CPE}^* = \left\lceil\left(\frac{\left|G_p''\left(\mathcal{X}_p^\alpha\right)\right|\mathcal{X}_p^\alpha}{2G_p'\left(\mathcal{X}_p^\alpha\right)\left|\bar{B}\right|}\right)^{1/3}T^{2/3}\right\rceil = \left\lceil\frac{\left|p-\mathcal{X}_p^\alpha-2\right|}{4\left|\bar{B}\right|}^{1/3}T^{2/3}\right\rceil, \tag{7}$$

where $\lceil\cdot\rceil$ is the ceiling function. This rate is slower than the MSE-optimal rate as derived in Phillips (2005):

$$K_{MSE}^* = \left\lceil\left(\frac{tr\left[\mathcal{W}\left(\mathbb{I}_{p^2}+\mathbb{K}_{pp}\right)(\Omega\otimes\Omega)\right]}{4vec(B)'\mathcal{W}vec(B)}\right)^{1/5}T^{4/5}\right\rceil$$

where $\mathbb{K}_{pp}$ is the $p^2 \times p^2$ commutation matrix, $\mathbb{I}_{p^2}$ is the $p^2 \times p^2$ identity matrix, and $\mathcal{W}$ is a $p^2 \times p^2$ weighting matrix. The constants in the optimal K formulae are also clearly different.

In the special case when $p = 1$, $\alpha = 5\%$, we have

$$K^*_{CPE} = \left\lceil 0.42293 \left|\bar{B}\right|^{-1/3} T^{2/3} \right\rceil, \tag{8}$$

and

$$K^*_{MSE} = \left\lceil 1.1487 \left|\bar{B}\right|^{-2/5} T^{4/5} \right\rceil.$$

Hence $K^*_{CPE}/K^*_{MSE} = \sqrt[15]{\left|\bar{B}\right|} T^{-\frac{2}{15}}$. So unless the relative bias $\bar{B}$ is very large, we expect K^*_{CPE} to be smaller than K^*_{MSE} in finite samples. However, this qualitative result does not hold for $p > 1$, in which case the finite sample comparison hinges on the specific values of B and Ω.

Since the high order expansion in (4) remains valid for testing a subvector of θ_0, the CPE-optimal K formula in (7) is applicable more generally.

To operationalize the CPE-optimal K formula, we can estimate the unknown parameter $\bar{B}$ 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 in the more general setting. It involves the following three steps. First, we obtain the estimator $\hat{\theta}_T$ and estimate the transformed score by $\hat{u}_t = RH_T^{-1}(\hat{\theta}_T)s_t(\hat{\theta}_T)$. 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 unobserved process $u_t = RH^{-1}s_t(\theta_0)$ and compute $\bar{B}$ as a function of the parameters of the parametric model. Plugging the estimate $\bar{B}$ into (7) gives the data-driven choice $\hat{K}^*_{CPE}$.

Suppose we use a VAR(1) as the approximating parametric model for $\hat{u}_t$ and use the following basis functions $\phi_{2k-1}(x) = \sqrt{2}\cos 2k\pi x$, $\phi_{2k}(x) = \sqrt{2}\sin 2k\pi x$ for $k = 1, ..., K/2$ in constructing $\hat{\Omega}$. Let $\hat{A}$ be the estimated autoregressive parameter matrix and $\hat{\Sigma}$ be the estimated innovation covariance matrix, then the plug-in estimates of Ω and B are

$$\hat{\Omega} = (\mathbb{I}_p - \hat{A})^{-1}\hat{\Sigma}(\mathbb{I}_p - \hat{A}')^{-1}, \tag{9}$$

$$\begin{aligned}\hat{B} = &-\frac{\pi^2}{6}(\mathbb{I}_p - \hat{A})^{-3}\Big(\hat{A}\hat{\Sigma} + \hat{A}^2\hat{\Sigma}\hat{A}' + \hat{A}^2\hat{\Sigma} - 6\hat{A}\hat{\Sigma}\hat{A}' \\ &+\hat{\Sigma}(\hat{A}')^2 + \hat{A}\hat{\Sigma}(\hat{A}')^2 + \hat{\Sigma}\hat{A}'\Big)(\mathbb{I}_p - \hat{A}')^{-3}.\end{aligned} \tag{10}$$

Here the constant $\pi^2/6$ in $\hat{B}$ is the same as that in Phillips (2005). We use the above basis functions and formulae in our simulation study. For the plug-in estimates under a general VAR(p) model, we refer to Andrews (1991) for the corresponding formulae.

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.

6 Simulation Evidence

This section provides some simulation evidence on the finite sample performance of the F approximation. We first consider the location model

$$Z_t = \theta + e_t$$

where $Z_t \in \mathbb{R}^6$ i.e. $d_z = d = 6$. The error e_t follows either a VAR(1) or VMA(1) process:

$$e_t = Ae_{t-1} + \sqrt{1-\psi^2}\varepsilon_t$$

$$e_t = A\varepsilon_{t-1} + \sqrt{1-\psi^2}\varepsilon_t$$

where $A = \psi \mathbb{I}_d$, $\varepsilon_t = (v_{1t} + \gamma f_t, v_{2t} + \gamma f_t, ..., v_{6t} + \gamma f_t)' / \sqrt{1+\gamma^2}$ 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

$$\Sigma = \frac{1}{1+\gamma^2}\mathbb{I}_d + \frac{\gamma^2}{1+\gamma^2}\mathbb{J}_d,$$

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

For the model parameters, we take $\psi = 0, 0.25, 0.50, 0.75$ and set $\gamma = 0$ or 1. We set the intercepts to zero as the tests we consider are invariant to them. For each test, we consider two significance levels $\alpha = 5\%$ and $\alpha = 10\%$, and three different sample sizes $T = 100, 200, 500$. The number of simulation replications is 10000.

We consider the following null hypotheses:

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

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 $\mathcal{H}_1\left(\delta^2\right)$ with $\tilde{c}$ distributed uniformly over the sphere in $\mathbb{R}^p$ with radius δ.

We examine the finite sample performance of seven different tests. The first four tests are based on the modified statistic F_T^*. They differ in terms of the bandwidth selection rule (CPE or MSE with $\mathcal{W} = \mathbb{I}$) and the critical value used ($p\mathcal{F}^{\alpha}_{p,K-p+1}$ or $\mathcal{X}^{\alpha}_p$). These four tests are labelled as F-CPE, χ^2-CPE, F-MSE, χ^2-MSE. The labels should be self-explanatory.

The fifth test is the conventional χ^2 test based on the unmodified Wald statistic F_T, MSE-based smoothing parameter, and critical value $\mathcal{X}_p^\alpha$. We refer to the test as the χ_c^2-MSE test with the subscript 'c' signifying a **c**onventional test. The next test is the 'F-MIN' test which is the same the F-CPE or F-MSE test except that K is set close to the minimal value which ensures that the series LRV estimator is not ill conditioned. In the simulation we use $K = p + 4$ in the F-MIN test. We also impose $K = p + 4$ to be the lower bound for all data-driven choices of K. The last test 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 F-MIN test and χ_c^2-MSE test can be regarded as the two ends of the power and size tradeoff. While the F-MIN test employs a small K in order to achieve size accuracy, the χ_c^2-MSE test uses a relatively large K for power improvement. Many other tests are available in the literature. We do not include all of them here as their performances are likely to be between the F-MIN test and the χ_c^2-MSE test.

Table 1: Empirical size of different 5% tests in a location model with VAR(1) error and sample size $T = 100$

	F-CPE	χ^2-CPE	χ_c^2-MSE	χ^2-MSE	F-MSE	F-MIN	KV
				$p = 1$			
$\psi = 0$	0.0504	0.0584	0.0548	0.0548	0.0489	0.0610	0.0513
$\psi = 0.25$	0.0549	0.0712	0.0760	0.0760	0.0646	0.0609	0.0557
$\psi = 0.50$	0.0618	0.0871	0.0921	0.0921	0.0706	0.0618	0.0637
$\psi = 0.75$	0.0736	0.1279	0.1286	0.1286	0.0813	0.0687	0.0807
				$p = 2$			
$\psi = 0$	0.0528	0.0638	0.0652	0.0621	0.0530	0.0501	0.0524
$\psi = 0.25$	0.0629	0.0861	0.0940	0.0875	0.0690	0.0509	0.0611
$\psi = 0.50$	0.0680	0.1118	0.1270	0.1128	0.0755	0.0540	0.0744
$\psi = 0.75$	0.0879	0.1784	0.2103	0.1765	0.0876	0.0695	0.1140
				$p = 3$			
$\psi = 0$	0.0542	0.0680	0.0747	0.0677	0.0534	0.0778	0.0551
$\psi = 0.25$	0.0677	0.1003	0.1202	0.1015	0.0753	0.0775	0.0684
$\psi = 0.50$	0.0751	0.1363	0.1778	0.1367	0.0808	0.0818	0.0917
$\psi = 0.75$	0.1020	0.2391	0.3286	0.2379	0.0995	0.0994	0.1613
				$p = 6$			
$\psi = 0$	0.0544	0.0799	0.1186	0.0822	0.0551	0.0517	0.0536
$\psi = 0.25$	0.0837	0.1401	0.2216	0.1321	0.0753	0.0533	0.0874
$\psi = 0.50$	0.1057	0.2131	0.4024	0.2130	0.0910	0.0597	0.1508
$\psi = 0.75$	0.1737	0.4040	0.6531	0.3563	0.1022	0.0988	0.3452

Table 1 gives the empirical size of the seven tests for the VAR(1) error with sample size $T = 100$ and $\gamma = 0$. The significance level is 5%, which is also the nominal size. We can make the following observations. First, as it is clear from the table, the conventional χ_c^2-MSE test 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 $\psi = 0.75$ and $p = 6$, the empirical size of the test is 0.6531, which is far from 0.05, the nominal size. Second, comparing the F-MSE test with the χ_c^2-MSE test, we find that using critical values from the F approximation 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 F approximation and chi-square approximation is large. Note that the hybrid F-MSE test is rigorously justified as the critical value from the F distribution is second order correct under the conventional increasing-smoothing asymptotics. Third, comparing the χ^2-MSE test with the χ_c^2-MSE test, we find that the finite sample scale correction is helpful in reducing the size distortion. However, the correction alone is not enough. The remaining size distortion calls for the use of an F approximation. Fourth, comparing the χ^2-CPE test with the χ^2-MSE tests, we find that the CPE-optimal K choice is a viable competitor to the MSE-optimal K choice. It is important to point out that unlike the standard practice, the MSE criterion we employ is the asymptotic MSE of $\hat{\Omega}$, which estimates the long run variance of $RH^{-1}s_t(\theta_0)$. So our MSE criterion is tailored toward the null hypothesis under consideration. Fifth, the size distortion of the F-CPE, F-MIN, and KV tests is substantially smaller than the conventional χ_c^2-MSE test. This is because these three tests employ asymptotic approximations that capture the estimation uncertainty of the LRV estimator. Finally, compared with the F-MIN and KV tests, the F-CPE test has only slightly larger size distortion. Since the bandwidth is set equal to the sample size, the KV test is designed to achieve the smallest possible size distortion. Similarly, the F-MIN test uses a small K value in order to achieve size accuracy. Given these observations, we can conclude that the F-CPE test succeeds in reducing the size distortion.

Table 2 presents the empirical size for the VMA(1) error process. The qualitative observations for the VAR(1) error remain valid. In fact, these qualitative observations hold for other parameter configurations such as different sample sizes, different values of γ and different error distributions.

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 (simulated) 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. Note that the F-CPE and the χ^2-CPE tests have the same size-adjusted power, as they are based on the same test statistic. For the same reason, the χ_c^2-MSE, χ^2-MSE and F-MSE tests have the same size-adjusted power. So we only report the power curves for the F-CPE, χ^2-MSE, F-MIN, and KV tests. Three observations can be drawn from these figures. First, the power of the F-CPE test is very close to the conventional Wald test. In some scenarios, the F-CPE test is more powerful. Second, the F-CPE test has higher power than the KV test in almost all cases. An exception is when the error dependence is very high and the number of restrictions is large. In this case, the two tests have almost the same power. When the error dependence is low, the selected K value is relatively large and the variance of the associated LRV estimator is

Table 2: Empirical size of different 5% tests in a location model with VMA(1) error and sample size $T = 100$

	F-CPE	χ^2-CPE	χ^2_c-MSE	χ^2-MSE	F-MSE	F-MIN	KV
				$p = 1$			
$\psi = 0$	0.0504	0.0584	0.0548	0.0548	0.0489	0.0610	0.0513
$\psi = 0.25$	0.0500	0.0639	0.0645	0.0645	0.0525	0.0616	0.0545
$\psi = 0.50$	0.0504	0.0706	0.0657	0.0657	0.0501	0.0612	0.0554
$\psi = 0.75$	0.0503	0.0745	0.0703	0.0703	0.0493	0.0606	0.0557
				$p = 2$			
$\psi = 0$	0.0528	0.0638	0.0652	0.0621	0.0530	0.0501	0.0524
$\psi = 0.25$	0.0570	0.0769	0.0819	0.0766	0.0584	0.0492	0.0576
$\psi = 0.50$	0.0528	0.0870	0.0905	0.0815	0.0551	0.0499	0.0603
$\psi = 0.75$	0.0534	0.0887	0.0929	0.0829	0.0536	0.0497	0.0609
				$p = 3$			
$\psi = 0$	0.0542	0.0680	0.0747	0.0677	0.0534	0.0778	0.0551
$\psi = 0.25$	0.0602	0.0875	0.1029	0.0876	0.0631	0.0777	0.0614
$\psi = 0.50$	0.0581	0.0971	0.1226	0.0944	0.0596	0.0780	0.0663
$\psi = 0.75$	0.0572	0.1009	0.1302	0.0978	0.0605	0.0787	0.0673
				$p = 6$			
$\psi = 0$	0.0544	0.0799	0.1186	0.0822	0.0551	0.0517	0.0536
$\psi = 0.25$	0.0678	0.1105	0.1878	0.1123	0.0621	0.0515	0.0741
$\psi = 0.50$	0.0669	0.1272	0.2571	0.1365	0.0639	0.0528	0.0860
$\psi = 0.75$	0.0668	0.1307	0.2953	0.1497	0.0657	0.0523	0.0884

small. In contrast, the LRV estimator used in the KV test has a large variance. As a result, the F-CPE 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 F-CPE test employ LRV estimators with large variance. The two tests can have comparable power. Finally, the F-CPE test is consistently more powerful than the F-MIN test. The power improvement is substantial in most of the cases. This is not surprising as the F-MIN test, like the KV test, 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 F-CPE test is as powerful as and sometimes more powerful than the conventional Wald test, the nonstandard KV test and the F-MIN test. This is true for all parameter combinations for the location model we considered.

We have also considered linear regression models as in Sun (2011b) in our simulation study:

$$y_t = \theta_0 + x_{t,1}\theta_1 + x_{t,2}\theta_2 + x_{t,3}\theta_3 + x_{t,4}\theta_4 + x_{t,0}$$

where the regressors $x_{t,j}$, $j = 1, 2, 3, 4$ and regression error $x_{t,0}$ follow mutually independent AR(1), MA(1) or MA(m) processes. For brevity, we only report the simulation results for the AR(1) case where $x_{t,j} = u_{t,j}/\sqrt{1-\psi^2}$, $j = 0, ..., 4$ for $u_{t,j} = \psi u_{t-1,j} + e_{t,j}$, $e_{t,j} \sim iidN(0,1)$. Results for other DGP's are qualitatively similar. We estimate $\theta = (\theta_0, \theta_1, ..., \theta_4)'$ by OLS and consider the null hypotheses $\mathcal{H}_{0p} : \theta_1 = ... = \theta_p = 0$ for $p = 1, 2, 3, 4$. Table 3 gives the empirical size of different 5% tests for sample size 200 and different values of ψ. It is clear that the qualitative observations for Table 1 remain valid. The qualitative conclusions for power comparisons in Figures 1-4 continue to apply in the regression case. For brevity, we omit the power figures.

7 Conclusion

Using the series LRV estimator, the paper proposes a new approach to autocorrelation robust inference in a time series setting. A great 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 having to simulate nonstandard critical values. Monte Carlo simulations show that our proposed F test has much more accurate size than the conventional Wald test. The size accuracy is achieved without power loss.

There are many extensions to the current paper. One possibility is to optimally select the smoothing parameter to minimize the type II error after controlling for the type I error. This is done in the working paper version of the present paper. See also Sun (2011a). We can also select the smoothing parameter to minimize the volume of a confidence 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)). It is straightforward to generalize of the series LRV estimation and inference to a first step GMM framework. It is more challenging to account for the estimation uncertainty of the optimal weighting matrix in a second-step efficient GMM framework.

Table 3: Empirical size of different 5% tests for linear regression models with AR(1) regressors and error with $T = 200$

	F-CPE	χ^2-CPE	χ^2-MSE	χ^2-MSE	F-MSE	K-MIN	KV
				$p = 1$			
$\psi = -0.5$	0.0675	0.0782	0.0807	0.0807	0.0750	0.0709	0.0618
$\psi = 0$	0.0574	0.0624	0.0615	0.0615	0.0580	0.0654	0.0534
$\psi = 0.3$	0.0621	0.0674	0.0687	0.0687	0.0651	0.0733	0.0565
$\psi = 0.5$	0.0679	0.0785	0.0815	0.0815	0.0746	0.0712	0.0603
$\psi = 0.7$	0.0815	0.0974	0.0989	0.0989	0.0893	0.0781	0.0691
$\psi = 0.7$	0.1400	0.1799	0.1851	0.1851	0.1544	0.1226	0.1200
				$p = 2$			
$\psi = -0.5$	0.0742	0.0900	0.0963	0.0924	0.0821	0.0626	0.0667
$\psi = 0$	0.0587	0.0644	0.0638	0.0620	0.0580	0.0534	0.0517
$\psi = 0.3$	0.0659	0.0750	0.0798	0.0768	0.0709	0.0603	0.0589
$\psi = 0.5$	0.0761	0.0933	0.0982	0.0944	0.0849	0.0643	0.0642
$\psi = 0.7$	0.1009	0.1291	0.1420	0.1339	0.1115	0.0745	0.0828
$\psi = 0.9$	0.1933	0.2642	0.2883	0.2668	0.2023	0.1311	0.1693
				$p = 3$			
$\psi = -0.5$	0.0827	0.1055	0.1177	0.1088	0.0903	0.0918	0.0775
$\psi = 0$	0.0647	0.0726	0.0771	0.0724	0.0641	0.0870	0.0609
$\psi = 0.3$	0.0748	0.0873	0.0960	0.0899	0.0777	0.0891	0.0653
$\psi = 0.5$	0.0903	0.1121	0.1268	0.1171	0.1014	0.0914	0.0770
$\psi = 0.7$	0.1193	0.1596	0.1840	0.1630	0.1307	0.1039	0.1020
$\psi = 0.9$	0.2444	0.3365	0.3936	0.3366	0.2327	0.1812	0.2222
				$p = 4$			
$\psi = -0.5$	0.0897	0.1176	0.1427	0.1222	0.1000	0.0652	0.0827
$\psi = 0$	0.0659	0.0769	0.0836	0.0749	0.0647	0.0575	0.0609
$\psi = 0.3$	0.0791	0.0938	0.1067	0.0938	0.0818	0.0593	0.0704
$\psi = 0.5$	0.0989	0.1256	0.1494	0.1285	0.1061	0.0632	0.0813
$\psi = 0.7$	0.1346	0.1884	0.2269	0.1935	0.1472	0.0817	0.1188
$\psi = 0.9$	0.2907	0.4112	0.5049	0.3963	0.2501	0.1637	0.2735

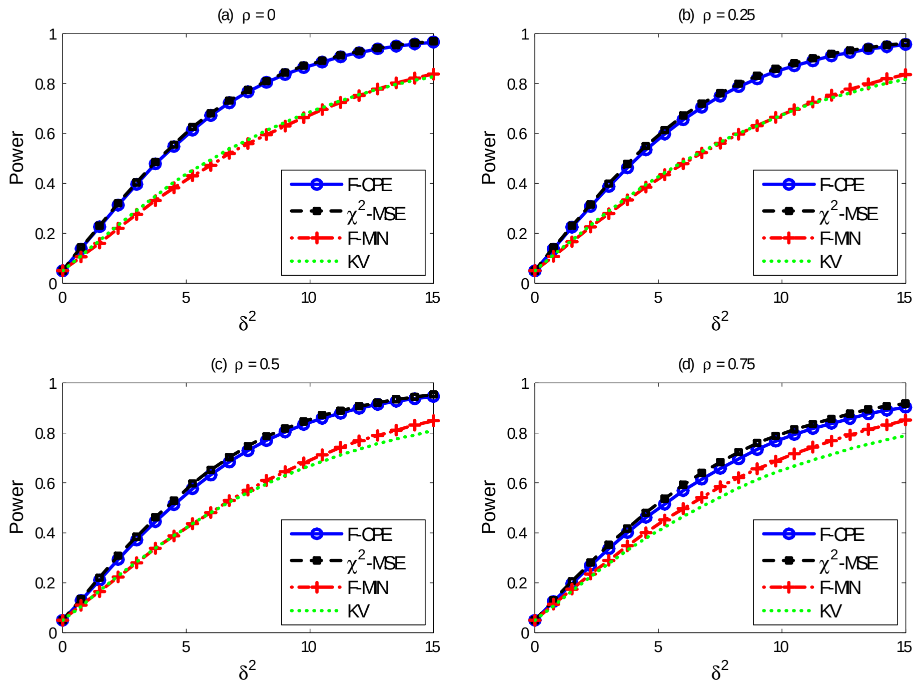

Figure 1: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with $T = 100$ and $p = 1$.

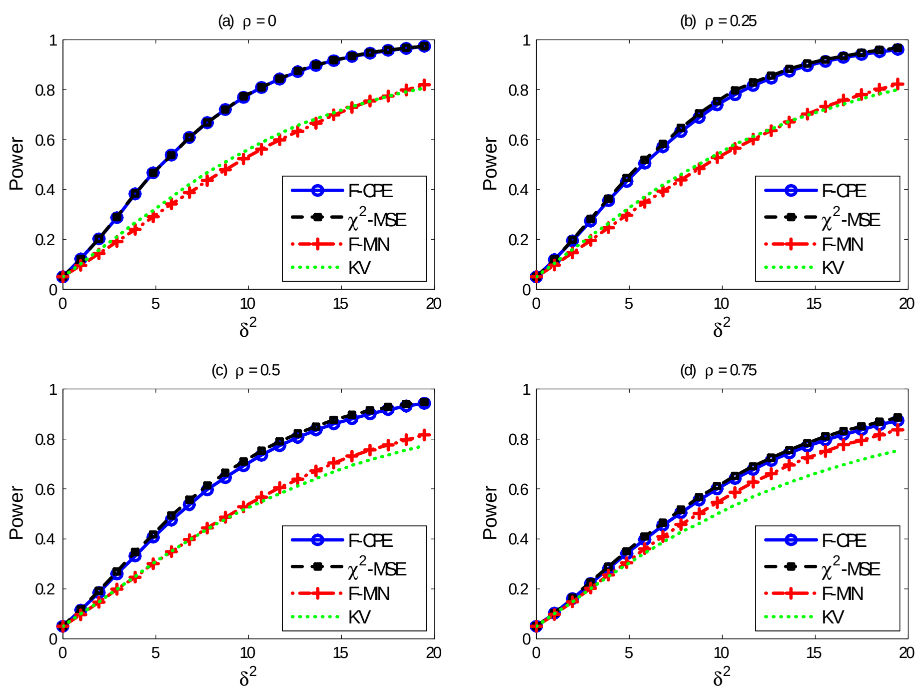

Figure 2: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with $T = 100$ and $p = 2$.

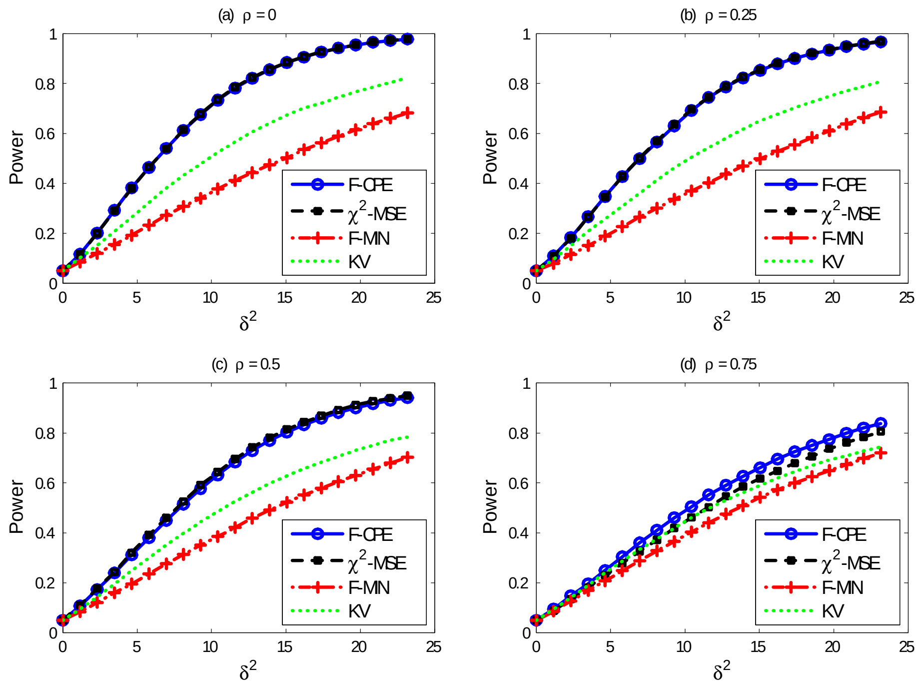

Figure 3: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with $T = 100$ and $p = 3$.

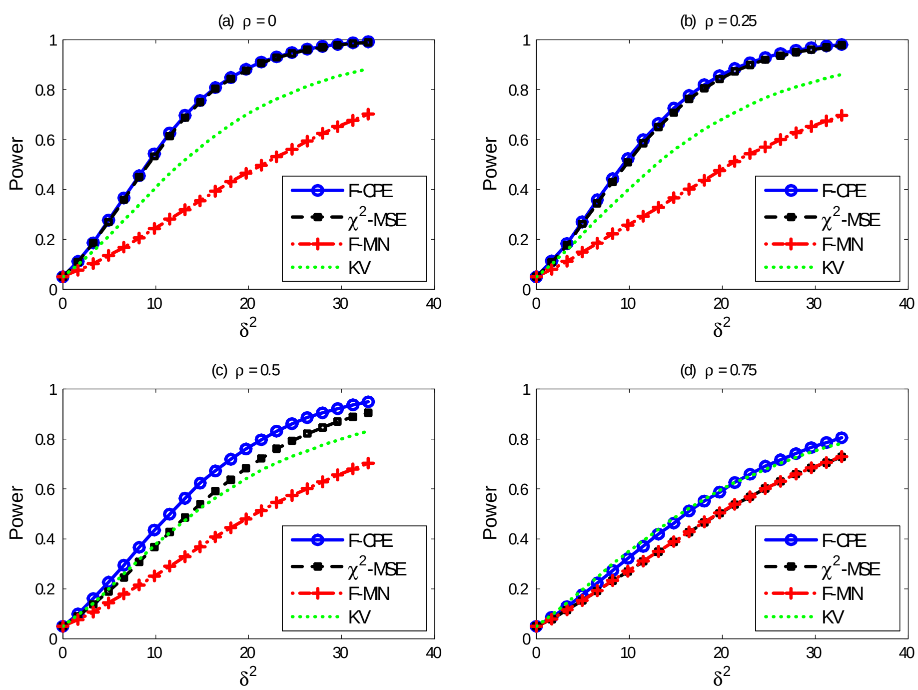

Figure 4: Size-adjusted Power of Different Testing Procedures for VAR(1) Error with $T = 100$ and $p = 6$.

8 Appendix of Proofs

8.1 Appendix A: Proof of Main Results

Proof of Theorem 1. Let

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

then under Assumption 1 (ii), we have

$$\begin{aligned}
\frac{1}{\sqrt{T}} S_t\left(\hat{\theta}_T\right) &= \frac{1}{\sqrt{T}} S_t(\theta_0) + \left(\frac{1}{T}\sum_{\tau=1}^{t} \frac{\partial s_\tau\left(\tilde{\theta}_T\right)}{\partial\theta'}\right)\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \\
&= \frac{1}{\sqrt{T}} S_t(\theta_0) - \left(\frac{1}{T}\sum_{\tau=1}^{t} \frac{\partial s_\tau\left(\tilde{\theta}_T\right)}{\partial\theta'}\right)\left[H^{-1}\frac{1}{\sqrt{T}} S_T(\theta_0) + o_p(1)\right] \\
&= \frac{1}{\sqrt{T}} S_t(\theta_0) - \frac{t}{T}\frac{1}{\sqrt{T}} S_T(\theta_0) + o_p(1), \qquad (11)
\end{aligned}$$

uniformly over t. As a result,

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

Here for convenience we have defined $\phi_k((T+1)/T) = 0$. Consequently,

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k(\frac{t}{T})\hat{u}_t &= RH_T^{-1}\left(\hat{\theta}_T\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k(\frac{t}{T})s_t\left(\hat{\theta}_T\right) \\
&= RH^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k(\frac{t}{T})\left[s_t(\theta_0) - \frac{1}{T}S_T(\theta_0)\right] + o_p(1) \\
&= RH^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_k(\frac{t}{T})s_t(\theta_0) + o_p(1) \\
&\xrightarrow{d} \Lambda\xi_k
\end{aligned}$$

jointly for $k = 0, ..., K$. Hence

$$K\Lambda^{-1}\hat{\Omega}\left(\Lambda'\right)^{-1} \xrightarrow{d} \sum_{k=1}^{K}\xi_k\xi_k' \overset{d}{=} \mathbb{W}_p(\mathbb{I}_p, K)$$

and

$$F_T \xrightarrow{d} (\Lambda\xi_0)' \left\{ \Lambda \frac{1}{K} \sum_{k=1}^{K} \xi_k \xi_k' \Lambda' \right\}^{-1} \Lambda\xi_0 = \xi_0' \left\{ \frac{1}{K} \sum_{k=1}^{K} \xi_k \xi_k' \right\}^{-1} \xi_0.$$

The limiting distribution of F_T is exactly the same as Hotelling's T-squared distribution (Hotelling (1931)). Using the well-known relationship between the T-squared distribution and the F-distribution, we have

$$\frac{(K-p+1)}{K} F_T \xrightarrow{d} F_{p,K-p+1}$$

as desired. ■

Proof of Theorem 2. To prove the theorem, we cast the problem in a location model

$$s_t = \mu + v_t$$

where $s_t = s_t(\theta_0)$ and $Ev_t = 0$. We want to test $\mathcal{H}_0 : \mu = 0$ against $\mathcal{H}_1 : \mu \neq 0$. The OLS estimator $\hat{\mu}_{OLS}$ of μ is $\hat{\mu}_{OLS} = T^{-1} \sum_{t=1}^{T} s_t$. Let $V = var([v_1', v_2', ..., v_T']')$ and $s = [s_1', s_2', ..., s_T']'$, then the GLS estimator $\hat{\mu}_{GLS}$ of μ is

$$\hat{\mu}_{GLS} = \left[(\ell_T \otimes \mathbb{I}_d)' V^{-1} (\ell_T \otimes \mathbb{I}_d) \right]^{-1} (\ell_T \otimes \mathbb{I}_d)' V^{-1} s,$$

where ℓ_T is a $T \times 1$ vector of ones. Let $\Omega = \sum_{j=-\infty}^{\infty} E s_t s_{t-j}'$ and $\Omega_{T,GLS} = var\left[T^{1/2} (\hat{\mu}_{GLS} - \mu) \right]$. Define

$$\begin{aligned} \Xi &= e_\mu' \Omega^{1/2} \tilde{\Omega}^{-1} \Omega^{1/2} e_\mu, \\ e_\mu &= \frac{(\Omega_{T,GLS})^{-1/2} \sqrt{T} (\hat{\mu}_{GLS} - \mu)}{\left\| (\Omega_{T,GLS})^{-1/2} \sqrt{T} (\hat{\mu}_{GLS} - \mu) \right\|}. \end{aligned}$$

Using the same argument as the proof of Lemma 3 in Sun (2011a), we can show that

$$P\left(\tilde{F}_T < z \right) = EG_p\left(z \Xi^{-1} \right) + O\left(\frac{1}{T} \right).$$

Writing $\Xi = \Xi(\tilde{\Omega})$ and taking a Taylor expansion of $\Xi(\tilde{\Omega})$ around $\Xi(\Omega) = 1$, we have

$$\left[\Xi\left(\tilde{\Omega} \right) \right]^{-1} = 1 + \mathcal{L} + \mathcal{Q} + remainder \tag{12}$$

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

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

and

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

It is not difficult to show that

$$E\mathcal{L}=\frac{K^2}{T^2}tr\left(B\Omega^{-1}\right)\frac{1}{p}\left(1+o\left(1\right)\right)+O\left(\frac{1}{T}\right),$$

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

and

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

Hence

$$\left[\Xi\left(\tilde{\Omega}\right)\right]^{-1}=1+\mathcal{L}+\mathcal{Q}+o_p\left(\frac{1}{K}+\frac{K^2}{T^2}\right)+o_p\left(\frac{1}{T}\right) \tag{13}$$

where the $o_p\left(\cdot\right)$ terms are also small in the root mean-square sense.

Note that

$$zG_p'\left(z\right)=\frac{1}{2^{k/2}\Gamma(k/2)}z^{p/2}\exp(-\frac{z}{2})\left\{z\geq 0\right\},$$

and

$$z^2G_p''\left(z\right)=-\frac{\left(z-p+2\right)}{2^{k/2+1}\Gamma(k/2)}z^{\frac{1}{2}p}\exp(-\frac{z}{2})\left\{z\geq 0\right\}$$

It is clear that there exists a constant $C>0$ such that $\left|zG_p'\left(z\right)\right|\leq C$ and $\left|z^2G_p''\left(z\right)\right|\leq C$ for all $z\in(0,\infty)$.

Using the asymptotic expansion in (13) and the boundedness of $G_p'\left(z\right)z$ and $G_p''\left(z\right)z^2$, we have

$$\begin{aligned}
&P\left(\tilde{F}_T^*<z\right) \\
&= EG_p\left(z\frac{K}{K-p+1}\left(1+\mathcal{L}+\mathcal{Q}\right)\right)+o\left(\frac{1}{K}+\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right) \\
&= EG_p\left(z\frac{K}{K-p+1}\right)+EG_p'\left(z\frac{K}{K-p+1}\right)z\frac{K}{K-p+1}\left(\mathcal{L}+\mathcal{Q}\right) \\
&+\frac{1}{2}EG_p''\left(z\frac{K}{K-p+1}\right)\left[z\frac{K}{K-p+1}\left(\mathcal{L}+\mathcal{Q}\right)\right]^2+o\left(\frac{1}{K}+\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right) \\
&= G_p\left(z\frac{K}{K-p+1}\right)+G_p'\left(z\right)zE\left(\mathcal{L}+\mathcal{Q}\right)+\frac{1}{2}EG_p''\left(z\right)z^2\left(E\mathcal{L}^2\right) \\
&+o\left(\frac{1}{K}+\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right) \\
&= G_p\left(z\right)+\frac{K^2}{T^2}G_p'\left(z\right)z\bar{B}+\frac{1}{K}G_p''\left(z\right)z^2+o\left(\frac{1}{K}\right)+o\left(\frac{K^2}{T^2}\right)+O\left(\frac{1}{T}\right)
\end{aligned}$$

as desired. ■

8.2 Appendix B: Expansion for General Wald Statistic

In this appendix, we make some simplifying and high-level assumptions and show that high order terms in Theorem 2 remain valid for a general hypothesis testing problem.

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

$$H_T(\tilde{\theta}_T) = H + O_p\left(\frac{1}{\sqrt{T}}\right) \tag{14}$$

for any $\tilde{\theta}_T = \theta_0 + O_p\left(1/\sqrt{T}\right)$. This assumption holds if the partial derivative of the (i,j)-th element $\partial \left[H_T(\theta)\right]_{ij} / \partial \theta'$ is stochastically bounded:

$$\frac{\partial \left[H_T(\tilde{\theta}_T)\right]_{ij}}{\partial \theta'} = O_p\left(1\right)$$

and $H_T\left(\theta_0\right) = H + O_p\left(1/\sqrt{T}\right)$. Under this assumption, we have:

$$\sqrt{T}\left(R\hat{\theta}_T - r\right) = -RH^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_t\left(\theta_0\right) + e_{1T}, \tag{15}$$

where $e_{1T} = O_p\left(1/\sqrt{T}\right)$.

We further assume that

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

uniformly over $r \in [0,1]$. This assumption implies (14). It holds if

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

and $\sqrt{T}\left[T^{-1}\sum_{t=1}^{[Tr]}\frac{\partial s_t(\theta_0)}{\partial \theta'} - rH\right]$ converges to a Gaussian process.

Using the above stronger assumption, we can improve the approximation error in (11):

$$\frac{1}{\sqrt{T}}S_t\left(\hat{\theta}_T\right) = \frac{1}{\sqrt{T}}S_t\left(\theta_0\right) - \frac{t}{T}\frac{1}{\sqrt{T}}S_T\left(\theta_0\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

uniformly over t. As a result, for $k \geq 1$,

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

So

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

where

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

Now

$$\begin{aligned}
\hat{\Omega}^{-1} &= \left\{\frac{1}{K}\sum_{k=1}^{K}\left[\Lambda_k+O_p\left(\frac{1}{\sqrt{T}}\right)\right]\left[\Lambda_k+O_p\left(\frac{1}{\sqrt{T}}\right)\right]'\right\}^{-1} \\
&= \left\{\frac{1}{K}\sum_{k=1}^{K}\Lambda_k\Lambda_k'+O_p\left[\left(\frac{1}{\sqrt{T}}\right)\left(\frac{1}{K}\sum_{k=1}^{K}\Lambda_k\right)\right]+O_p\left(\frac{1}{T}\right)\right\}^{-1} \\
&= \left[\frac{1}{K}\sum_{k=1}^{K}\Lambda_k\Lambda_k'+O_p\left(\frac{1}{\sqrt{TK}}\right)\right]^{-1} \\
&= \hat{\Omega}_L^{-1}+e_{2T}
\end{aligned}$$

where

$$\hat{\Omega}_L = \frac{1}{K}\sum_{k=1}^{K}\Lambda_k\Lambda_k' \text{ and } e_{2T} = O_p\left(\frac{1}{\sqrt{TK}}\right).$$

Let

$$\Lambda_0 = RH^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_t\left(\theta_0\right).$$

Then

$$\begin{aligned}
F_T &= \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}RH_T^{-1}(\tilde{\theta}_T)s_t\left(\theta_0\right)\right]'\hat{\Omega}^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}RH_T^{-1}(\tilde{\theta}_T)s_t\left(\theta_0\right)\right] \\
&= \left(-\Lambda_0+e_{1T}\right)'\left(\hat{\Omega}_L^{-1}+e_{2T}\right)\left(-\Lambda_0+e_{1T}\right) = F_{LT}+\eta_T+\eta_T^*
\end{aligned}$$

where

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

$$\eta_T = -2e_{1T}'\Omega^{-1}\Lambda_0 = O_p\left(\frac{1}{\sqrt{T}}\right) \tag{16}$$

and

$$
\begin{aligned}
\eta_T^* &= -2e_{1T}'\left(\hat{\Omega}_L^{-1}-\Omega^{-1}\right)\Lambda_0 - 2e_{1T}'e_{2T}\Lambda_0 + e_{1T}'\hat{\Omega}_L^{-1}e_{1T} + e_{1T}'e_{2T}e_{1T} \\
&= O_p\left(\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{K}}+\frac{K^2}{T^2}\right)\right) + O_p\left(\frac{1}{T\sqrt{K}}\right) + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{T\sqrt{TK}}\right) \\
&= O_p\left(\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{K}}+\frac{K^2}{T^2}\right)\right) + O_p\left(\frac{1}{T}\right). \qquad (17)
\end{aligned}
$$

Here η_T does not depend on K.

Note that F_{LT} is the Wald statistic for testing whether the mean of the process $\tilde{Y}_t = RH^{-1}s_t(\theta_0)$ is 0. In other words, F_{LT} is in exactly the same form as the statistic F_T defined in (3) but with a transformed score process. Therefore, if the transformed score $RH^{-1}s_t(\theta_0)$ satisfies the Assumptions in theorem 2, we can establish the following theorem.

Theorem 3 *Assume (i) $\rho(\theta, Z_t)$ is three times continuously differentiable in θ; (ii) $\hat{\theta}_T = \theta_0 + o_p(1)$ for an interior point θ_0 of Θ; (iii) for sufficiently large C, $P(|\eta_T| > \log T/\sqrt{T}) = O\left(1/\sqrt{T}\right)$ and $P(|\eta_T^*| > \delta_T/\log T) = o(\delta_T)$ for $\delta_T = 1/K + K^2/T^2$; (iv) $s_t(\theta_0)$ satisfies the Assumptions in Theorem 2, then*

$$
\begin{aligned}
&P\left(\frac{(K-p+1)}{K}F_T < z\right) \\
&= G_p(z) + \frac{K^2}{T^2}G_p'(z)z\bar{B} + \frac{1}{K}G_p''(z)z^2 + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{\log T}{\sqrt{T}}\right) \qquad (18)
\end{aligned}
$$

where

$$
\bar{B} = \frac{1}{p}tr\left\{B\Omega^{-1}\right\},\ B = E\hat{\Omega}_L - \Omega
$$

and the $O(\log T/\sqrt{T})$ term does not depend on K

Proof of Theorem 3. Since $F_T = F_{LT} + \eta_T + \eta_T^*$, we have

$$
\begin{aligned}
& P\left(\frac{(K-p+1)}{K}F_T < z\right) \\
=\ & P\left(\frac{(K-p+1)}{K}(F_{LT} + \eta_T + \eta_T^*) < z\right) \\
=\ & P\left(\frac{(K-p+1)}{K}(F_{LT} + \eta_T + \eta_T^*) < z, |\eta_T| < C/\sqrt{T}, |\eta_T^*| < \delta_T/\log T\right) \\
& + P\left(|\eta_T| \geq \log T/\sqrt{T}\right) + P\left(|\eta_T^*| \geq \delta_T/\log T\right) \\
=\ & P\left(\frac{(K-p+1)}{K}F_{LT} < z + \frac{\log T}{\sqrt{T}} + \frac{\delta_T}{\log T}\right) + o(\delta_T) + O\left(\frac{1}{\sqrt{T}}\right) \\
=\ & G_p(z) + \frac{K^2}{T^2}G_p'(z)z\bar{B} + \frac{1}{K}G_p''(z)z^2 + o\left(\frac{1}{K}\right) + o\left(\frac{K^2}{T^2}\right) + O\left(\frac{\log T}{\sqrt{T}}\right)
\end{aligned}
$$

where the last equality follows from Theorem 2 and the continuous differentiability of $G_p(\cdot), G_p'(\cdot)$, and $G_p''(\cdot)$ and boundedness of $G_p(z), zG_p'(z)$ and $z^2G_p''(z)$. ■

Theorem 3 shows that, if we ignore the terms that do not depend on K, the approximate CPE for the Wald confidence region based on the statistic F_T^* for a subvector of the parameter is exactly the same as that for the whole parameter vector. Hence the optimal K formula in (7) remains valid for inference on a subvector of the true parameter.

Assumption (iii) is a crucial high-level assumption. Given the probability orders in (16) and (17), the assumption holds under sufficient moment and mixing conditions. With more sophisticate arguments as in Sun and Phillips (2008), we can drop the logarithm factor in the expansion. Here we are content with the weaker result as our main interest is to capture the effect of K on the CPE.

References

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

[2] 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.

[3] Diebold, F. X. and R.S. Mariano (1995): “Comparing predictive accuracy.” *Journal of Business and Economic Statistics*, 13(3), 253–263.

[4] 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.

[5] Hotelling H. (1931): “The Generalization of Student’s Ratio.” *Annals of Mathematical Statistics* 2, 360–378.

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

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

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

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

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

[11] 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

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

[13] ——— (1994): "Automatic Lag Selection in Covariance Estimation." *Review of Economic Studies* 61, 631–654.

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

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

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

[17] 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.

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

[19] 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.

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

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

[22] Sun, Y. and M.S. Kim (2012): "Simple and Powerful GMM Over-identification Tests with Accurate Size." *Journal of Econometrics* 166, 267–281.

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

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