Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroskedasticity and Autocorrelation Robust Inference

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

June 15, 2013

Abstract

In the presence of heteroscedasticity and autocorrelation of unknown forms, the covariance matrix of the parameter estimator is often estimated using a nonparametric kernel method that involves a lag truncation parameter. Depending on whether this lag truncation parameter is specified to grow at a slower rate than or the same rate as the sample size, we obtain two types of asymptotic approximations: the small-b asymptotics and the fixed-b asymptotics. Using techniques for probability distribution approximation and high order expansions, this paper shows that the fixed-b asymptotic approximation provides a higher order refinement to the first order small-b asymptotics. This result provides a theoretical justification on the use of the fixed-b asymptotics in empirical applications. On the basis of the fixed-b asymptotics and higher order small-b asymptotics, the paper introduces a new and easy-to-use asymptotic F test that employs a finite sample corrected Wald statistic and uses an F-distribution as the reference distribution. Finally, the paper develops a bandwidth selection rule that is testing-optimal in that the bandwidth minimizes the type II error of the asymptotic F test while controlling for its type I error. Monte Carlo simulations show that the asymptotic F test with the testing-optimal bandwidth works very well in finite samples.

JEL Classification: C13; C14; C32; C51

Keywords: Asymptotic expansion, F-distribution, Heteroskedasticity and Autocorrelation Robust, long-run variance, robust standard error, testing-optimal smoothing parameter choice, type I and type II errors.

---

*Email: yisun@ucsd.edu. This paper replaces a 2010 working paper with the same title and a 2011 working paper entitled "Asymptotic F Test in a GMM Framework." The author gratefully acknowledges partial research support from NSF under Grant No. SES-0752443. The author thanks the coeditor Jianqing Fan, an associate editor, three anonymous referees and many seminar participants for helpful comments. Correspondence to: Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508.

1 Introduction

In linear and nonlinear models with moment restrictions, it is standard practice to employ the generalized method of moments (GMM) to estimate model parameters. Consistency of the GMM estimator in general does not depend on the dependence structure of the moment conditions. However, we often want not only point estimators of the model parameters, but also their covariance matrix in order to conduct inference. A popular covariance estimator that allows for general forms of dependence is the nonparametric kernel estimator. The underlying smoothing parameter is the truncation lag (or bandwidth parameter) or the ratio b of the truncation lag to the sample size; see for example See Newey and West (1987) and Andrews (1991). In econometrics, this covariance estimator is often referred to as the heteroskedasticity and autocorrelation robust (HAR) estimator. A major difficulty in using the HAR covariance estimator to perform hypothesis testing lies in how to select the smoothing parameter b and how to approximate the sampling distribution of the associated test statistic.

In terms of distributional approximations, both the conventional small-b asymptotics and nonstandard fixed-b asymptotics are considered in the literature. In the former case, b is assumed to be small in that it goes to zero at certain rate with the sample size. Under this asymptotic specification, the Wald statistic is asymptotically χ^2. In the latter case, b is assumed to be held fixed at a given value, and the Wald statistic has a nonstandard limiting distribution. See Kiefer and Vogelsang (2002a, 2002b, 2005, hereafter KV). KV (2005) show by simulation that the nonstandard fixed-b asymptotic approximation is more accurate than the conventional asymptotic χ^2 approximation. Jansson (2004) and Sun, Phillips, Jin (2008, hereafter SPJ) provide theoretical analyses for location models.

This paper has several objectives and makes several contributions. The first objective is to investigate the relationship between the small-b asymptotics and the fixed-b asymptotics. We show that the fixed-b asymptotic approximation provides a higher order refinement to the first order small-b asymptotics. This result is established via a high order expansion of the Wald statistic under the small-b asymptotics and an expansion of the fixed-b asymptotic distribution around $b = 0$. Our theoretical result establishes the asymptotic validity of the fixed-b critical values regardless of the asymptotic thought experiments we use.

The second objective is to approximate a modified Wald statistic by a standard F distribution. The modification corrects for the demeaning bias of the HAR estimator, which is due to the estimation uncertainty of model parameters, and the dimensionality bias of the Wald statistic, which is present when the number of joint hypotheses is greater than 1. We design an asymptotic F test that employs the modified Wald statistic and critical values from a standard F distribution. When b is not too large, more specifically,

$b \leq 0.4, 0.3, 0.2$ for the Bartlett, Parzen and QS kernels respectively, the asymptotic F test is as accurate in size as the nonstandard test of KV (2005) and yet is as easy to use as the standard Wald test, as both the correction factor and the critical values are easy to obtain.

The third objective is to provide a theoretical explanation on why the conventional Wald test has a severe size distortion when p, the number of restrictions being tested or the dimension of hypothesis space, is large. We show that the difference between the high-order corrected F critical value and the first-order χ^2 critical value depends on the bandwidth parameter b, the number of joint hypotheses, and the kernel function used in the HAR estimation. The conventional Wald test can be severely size distorted as it uses critical values that do not depend on b and the kernel function and do not adequately capture the effect of the dimension of the hypothesis space.

The fourth objective is to operationalize the asymptotic F test by determining suitable values of the bandwidth parameter b. At present it is standard practice to use the bandwidth parameter that is optimal for the point estimation of the covariance matrix of the parameter estimator. This choice may not be optimal from a testing point of view. In hypothesis testing, our ultimate goal is to minimize the type II error hence maximize the power of the test while controlling for the type I error. This goal is different from the minimization of the mean squared error of the covariance estimator. In this paper, we propose to select the bandwidth parameter that is optimal for hypothesis testing. More specifically, the testing-optimal bandwidth parameter minimizes the type II error subject to the constraint that the type I error is bounded by $\tau\alpha$ where α is the nominal type I error and $\tau > 1$ is the permitted tolerance towards the type I error. The type I and type II errors are approximately measured on the basis of higher order expansions.

The testing-optimal bandwidth is fundamentally different from the MSE-optimal bandwidth in terms of both the rate of convergence and the parameters on which they depend. The testing-optimal bandwidth is tailored to the testing problem at hand. As a result, it depends on every aspect of the testing problem under consideration. For example, it depends on the null and local alternative hypotheses, the significance level, and the number of restrictions being tested while the MSE-optimal bandwidth does not. When the permitted tolerance towards the type I error is small, the testing-optimal bandwidth is larger by an order of magnitude than the MSE-optimal bandwidth. In hypothesis testing, when the type I error is of greater concern, we should employ under-smoothing in order to achieve more bias reduction than that is required by the MSE criterion.

The final objective is to examine the finite sample performance of the asymptotic F test. We compare the asymptotic F test with testing-optimal bandwidth to the conventional Wald test with MSE-optimal bandwidth. We also include the KV type of test with $b = 1$

and using the nonstandard fixed-b asymptotics as the reference distribution. The simulation result shows the asymptotic F test is only slightly more size distorted than the KV test with $b = 1$. For all the kernels considered, the F test is much more accurate in size than the conventional Wald test. In terms of the power, the F test is as powerful as the Wald test. As expected, it is much more powerful than the KV test with $b = 1$. We can therefore conclude that the F test with testing-optimal bandwidth has good size and power properties.

The papers that are closely related to the present one are SPJ (2008) and Sun and Phillips (2009, unpublished). In SPJ (2008), a simple t test in a univariate location model is considered. Here we generalize their result to the Wald test in a general GMM framework. The generalization is far from trivial as it is much more difficult to obtain high order expansions in a GMM setting. In addition, the Wald test with multiple restrictions is also harder to analyze than a simple t test as considered in SPJ (2008). Sun and Phillips (2009) focuses only on the high order small-b asymptotics and optimal confidence interval construction under the small-b asymptotics.

A paper with conceptual ideas related to those presented here is Gao and Gijbels (2008). In their seminal work, Gao and Gijbels consider testing for a parametric function form in a nonparametric kernel regression. The test statistic depends on the kernel smoothing bandwidth. They propose choosing the bandwidth to maximize the power of the test over a set of bandwidth values under which the size is under control. This is conceptually similar to our approach, although the problems considered and the technical machinery used are fundamentally different.

The remainder of the paper is organized as follows. Section 2 describes the testing problem of concern and provides an overview of the fixed-b asymptotic theory. Section 3 expands the fixed-b asymptotic distribution around $b = 0$. The expansion and the representations that lead to it help deepen our understanding of the fixed-b approximation. Section 4 develops a high order expansion of the Wald statistic. Section 5 introduces the asymptotic F test. On basis of the high order expansion, the next section describes approximate measures of the type I and type II errors of the asymptotic F test. It also gives an explicit and closed-form expression for the testing-optimal bandwidth for the F test. Section 7 presents simulation evidence and last section concludes. Proofs are given in the Appendix.

2 Autocorrelation Robust Testing

We employ a standard GMM framework. We are interested in a $d \times 1$ vector of parameters $\theta \in \Theta \subseteq \mathbb{R}^d$. Let v_t denote a vector of observations. Let θ_0 be the true value and assume that θ_0 is an interior point of the compact parameter space Θ. The moment conditions

$$Ef\left(v_t,\theta\right)=0,\ \ t=1,2,...,T$$

hold if and only if $\theta=\theta_0$ where $f\left(\cdot\right)$ is an $m\times 1$ vector of twice continuously differentiable functions with $m\geq d$ and rank $E\left[\partial f\left(v_t,\theta_0\right)/\partial\theta'\right]=d$. Define

$$g_t\left(\theta\right)=T^{-1}\sum_{j=1}^{t}f(v_j,\theta),$$

the GMM estimator of θ_0 is then given by

$$\hat{\theta}_T=\arg\min_{\theta\in\Theta}g_T\left(\theta\right)'\mathcal{W}_T g_T\left(\theta\right)$$

where $\mathcal{W}_T$ is an $m\times m$ positive semidefinite weighting matrix.

Let

$$G_t(\theta)=\frac{\partial g_t\left(\theta\right)}{\partial\theta'}=\frac{1}{T}\sum_{j=1}^{t}\frac{\partial f(v_j,\theta)}{\partial\theta'}\text{ and }G_0=E\frac{\partial f(v_j,\theta_0)}{\partial\theta'}.$$

The following high level assumptions are standard in the literature on the fixed-b asymptotics; See for example KV (2005), Lee and Kuan (2009), Bester, Conley, Hansen (2011), Zhang and Shao (2012) and references therein.

Assumption 1 *$plim_{T\to\infty}\hat{\theta}_T=\theta_0$ and θ_0 is an interior point of Θ.*

Assumption 2 *$plim_{T\to\infty}G_{[rT]}(\tilde{\theta}_T)=rG_0$ uniformly in r for any $\tilde{\theta}_T$ whose elements are between the corresponding elements of $\hat{\theta}_T$ and θ_0.*

Assumption 3 *$\mathcal{W}_T$ is positive semidefinite, $plim_{T\to\infty}\mathcal{W}_T=\mathcal{W}_\infty$, and $G_0'\mathcal{W}_\infty G_0$ is positive definite.*

Under the above assumptions, we have, using element-by-element mean value expansions:

$$\sqrt{T}\left(\hat{\theta}_T-\theta_0\right)=-\left(G_0'\mathcal{W}_\infty G_0\right)^{-1}G_0'\mathcal{W}_\infty\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f(v_t,\theta_0)+o_p\left(1\right).\tag{1}$$

Consider the null hypothesis $H_0: r(\theta_0)=0$ and the alternative hypothesis $H_1: r\left(\theta_0\right)\neq 0$ where $r\left(\theta\right)$ is a $p\times 1$ vector of twice continuously differentiable functions with first order derivative matrix $R(\theta)=\partial r(\theta)/\partial\theta'$. The Wald statistic is based on the difference $r(\hat{\theta}_T)-r\left(\theta_0\right)$. Under Assumptions 1–3, we have, using (1):

$$\sqrt{T}\left[r(\hat{\theta}_T)-r\left(\theta_0\right)\right]=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t+o_p\left(1\right)$$

where

$$u_t := \phi(v_t,\theta_0) = -R_0\left(G_0'\mathcal{W}_\infty G_0\right)^{-1} G_0'\mathcal{W}_\infty f(v_t,\theta_0),$$

and $R_0 = R(\theta_0)$. $\phi(v_t,\theta_0)$ can be regarded as an influence function representation of $\sqrt{T}[r(\hat{\theta}_T) - r(\theta_0)]$.

Assumption 4 *$T^{-1/2}\sum_{t=1}^{[Tr]} u_t \to^d \Lambda W_p(r)$ where $\Lambda\Lambda' = \Omega = \sum_{j=-\infty}^{\infty} Eu_t u_{t-j}'$ is the long run variance (LRV) of u_t and $W_p(r)$ is the p-dimensional standard Brownian motion.*

The FCLT in Assumption 4 holds for serially correlated and heterogeneously distributed data that satisfy certain regularity conditions on moments and the dependence structure over time. These primitive regularity conditions can be found in Davidson (1994), Tanaka (1996), White (2001), among others. For example, Theorem 3.14 in Tanaka (1996, p. 98) considers a linear process of the form: $u_t = \sum_{\ell=0}^{\infty} A_\ell \varepsilon_{t-\ell}$ where $\varepsilon_t \sim iid(0, I_p)$ and $\{A_\ell\}_{\ell=0}^{\infty}$ is a sequence of $p\times p$ matrices. Sufficient conditions for the FCLT are (i) $\sum_{\ell=0}^{\infty} \ell \|A_\ell\| < \infty$ for $\|A_\ell\| = [tr(A_\ell' A_\ell)]^{1/2}$; (ii) $A = \sum_{\ell=0}^{\infty} A_\ell$ and A_0 are nonsingular. Tanaka (1996) proves the theorem using the linear process approach of Phillips and Solo (1992). The iid assumption on ε_t can be replaced by the martingale difference sequence assumption with additional moment requirements. See Theorem 3.15 in Phillips and Solo (1992).

For processes that are not necessarily linear, mixing conditions are usually imposed to obtain the FCLT. For example, Phillips and Durlauf (1986, Corollary 2.2) consider a mean zero and weakly stationary sequence $\{u_t = (u_{t1}, ..., u_{tp})'\}$. They show that the following three conditions are sufficient for the FCLT:

(i) $E|u_{1i}|^\beta > \infty$ $(i = 1, ..., p)$ for $\beta \geq 2$;

(ii) u_t is φ-mixing with mixing coefficients satisfying $\sum_{\ell=1}^{\infty} \varphi_\ell^{1-1/\beta} < \infty$;

(iii) $\Omega = \sum_{j=-\infty}^{\infty} Eu_t u_{t-j}'$ is positive definite.

Conditions (i) and (ii) can be replaced by

(i)' $E|u_{1i}|^\beta > \infty$ $(i = 1, ..., p)$ for $\beta > 2$;

(ii)' u_t is α-mixing with mixing coefficients satisfying $\sum_{\ell=1}^{\infty} \alpha_\ell^{1-2/\beta} < \infty$.

Phillips and Durlauf (1986, Theorem 2.1) give another set of sufficient conditions without assuming weak stationarity. See also Theorem 7.30 of White (2001) for a FCLT for heterogenous mixing processes.

Note that the mixing properties are hereditary in the sense that, for any measurable function $m(\cdot)$, the process $m(v_t)$ possesses the mixing property of v_t. Hence, in our setting it suffices to verify whether the mixing conditions hold for $\{v_t\}$. Doukhan (1994, Section 1.3.2 and Chapter 2) provides examples of time series with various mixing properties. In particular, the commonly-used ARMA process satisfies the α-mixing condition if it is stationary and the innovations are iid and have a probability density with respect to the Lebesgue measure.

While Assumption 4 holds for processes that may not be weakly stationary, it rules out the case that u_t is strongly persistent, i.e. each component of u_t follows a unit root or near unit root process. While it is possible to develop fixed-b asymptotics for these two cases, the mechanics and details are sufficiently different to warrant a separate investigation.

Under Assumptions 1-4, we now have

$$\sqrt{T}\left[r(\hat{\theta}_T) - r(\theta_0)\right] \to^d \Lambda W_p(1) \sim N(0, \Omega),$$

which provides the usual basis for robust testing. The F-test version of the Wald statistic for testing H_0 against H_1 is

$$F_T = \left[\sqrt{T} r(\hat{\theta}_T)\right]' \hat{\Omega}_T^{-1} \left[\sqrt{T} r(\hat{\theta}_T)\right] / p, \tag{2}$$

where $\hat{\Omega}_T$ is an estimator of Ω. The kernel estimator $\hat{\Omega}_T$ of Ω takes the form of

$$\hat{\Omega}_T = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} k_b\left(\frac{t-\tau}{T}\right)\hat{u}_t\hat{u}_\tau' \tag{3}$$

where $\hat{u}_t$ is a plug-in estimator of u_t given by

$$\hat{u}_t = -R(\hat{\theta}_T)\left(G_T'(\hat{\theta}_T)\mathcal{W}_T G_T(\hat{\theta}_T)\right)^{-1} G_T'(\hat{\theta}_T)\mathcal{W}_T f(v_t, \hat{\theta}_T), \tag{4}$$

$k(\cdot)$ is a kernel function, and $k_b(x) = k(x/b)$ for $x \in [-1, 1]$. Here b is the smoothing parameter that affects the asymptotic properties of $\hat{\Omega}_T$ and the associated test statistic.

Following KV (2005) and using integration by parts, we can show that under the assumptions given above:

$$F_T \to^d F_\infty(p, b)$$

for any fixed value of b, where

$$F_\infty(p,b) = W_p'(1)\left[\int_0^1\int_0^1 k_b(r-s)dV_p(r)dV_p'(s)\right]^{-1} W_p(1)/p, \tag{5}$$

and $V_p(r) = W_p(r) - rW_p(1)$ is the p-dimensional Brownian bridge process. Note that $cov\left[W_p(1), V_p(r)\right] = cov\left[W_p(1), W_p(r)\right] - r cov\left[W_p(1), W_p(1)\right] = r - r = 0$, so $W_p(1)$ is independent of the Brownian bridge process $V_p(r)$.

The distribution $F_\infty(p, b)$ is the so-called fixed-b limiting distribution of F_T. When there is no possibility of confusion, we use $F_\infty(p, b)$ to denote a random variable with distribution $F_\infty(p, b)$ and the distribution itself. Similarly, we use $F_{p,K}$ to denote a random variable with F distribution $F_{p,K}$ and the distribution itself.

3 Representation and Expansion of the Nonstandard Asymptotic Distribution

This section presents alternative representations and develops an asymptotic expansion of the limit distribution given in (5) as the bandwidth parameter $b \to 0$. The asymptotic expansion and later developments in the paper make use of the following kernel conditions:

Assumption 5 *(i) $k(x) : \mathbb{R} \to [0,1]$ is symmetric, piecewise smooth with $k(0) = 1$ and $\int_0^\infty k(x)xdx < \infty$. (ii) The Parzen characteristic exponent defined by*

$$q = \max\{q_0 : q_0 \in \mathbb{Z}^+,\ g_{q_0} = \lim_{x \to 0} \frac{1 - k(x)}{|x|^{q_0}} < \infty\} \tag{6}$$

is greater than or equal to 1.

Assumption 5 imposes only mild conditions on the kernel function. All the commonly used kernels satisfy (i) and (ii). The assumption $\int_0^\infty k(x)xdx < \infty$ ensures the integrals that appear in our proofs are finite. It also enables us to use the Riemann-Lebesgue lemma. For the Bartlett kernel, the Parzen characteristic exponent is 1. For the Parzen and QS kernels, the Parzen characteristic exponent is 2. We focus on these three kernels as they are positive semidefinite, a condition that ensures the positive semidefiniteness of the associated LRV estimator.

Define

$$k_b^*(r,s) = k_b(r-s) - \int_0^1 k_b(r-t)dt - \int_0^1 k_b(\tau - s)d\tau + \int_0^1 \int_0^1 k_b(t-\tau)dtd\tau,$$

which is the "centered" version of the kernel function in the sense that

$$\int_0^1 k_b^*(r,s)dr = \int_0^1 k_b^*(r,s)ds = \int_0^1 \int_0^1 k_b^*(r,s)drds = 0 \text{ for any } r \text{ and } s.$$

Then it is easy to show that

$$\int_0^1 \int_0^1 k_b(r-s)dV_p(r)dV_p'(s) = \int_0^1 \int_0^1 k_b^*(r,s)dW_p(r)dW_p'(s).$$

Note that while $k(x)$ may be defined on $\mathbb{R}$, $k_b(r-s)$ and hence $k_b^*(r,s)$ are defined on $[0,1] \times [0,1]$ for any given b. Under Assumption 5, $k_b^*(r,s)$ is a symmetric and integrable function in $L^2([0,1] \times [0,1])$. So the Fredholm integral operator with kernel $k_b^*(r,s)$ is self-adjoint and compact. By the spectral theorem, e.g. Promislow (2008, p. 199), we can expand $k_b^*(r,s)$ as

$$k_b^*(r,s) = \sum_{n=1}^\infty \lambda_n^* f_n^*(r) f_n^*(s), \tag{7}$$

where the right hand side converges in $L^2([0,1]\times[0,1])$. Here λ_n^* is an eigenvalue of the centered kernel and $f_n^*(r)$ is the corresponding eigenfunction, i.e. $\lambda_n^* f_n^*(s) = \int_0^1 k_b^*(r,s) f_n^*(r)dr$. Since $k_b^*(r,s)$ is centered, we have $\int_0^1 f_n^*(r)dr = 0$. It follows from (7) that

$$\int_0^1 \int_0^1 k_b^*(r,s)dW_p(r)dW_p'(s) = \sum_{n=1}^{\infty} \lambda_n^* \zeta_n \zeta_n' \tag{8}$$

where $\zeta_n = \int_0^1 f_n^*(r)dW_p(r)$. Since $f_n^*(s)$ is an orthonormal sequence of functions in $L^2[0,1]$, $\zeta_n \sim iidN(0,\mathbb{I}_p)$ and $\zeta_n\zeta_n'$ follows $\mathbb{W}_p(\mathbb{I}_p,1)$, a simple Wishart distribution. Hence the double stochastic integral is equal in distribution to a weighted sum of independent Wishart distributions.

Using (8), we obtain our first representation of $pF_\infty(p,b)$ as

$$pF_\infty(p,b) \overset{d}{=} \eta' \left[\sum_{n=1}^{\infty} \lambda_n^* \zeta_n \zeta_n'\right]^{-1} \eta, \tag{9}$$

where $\zeta_n \sim iidN(0,\mathbb{I}_p)$, $\eta \sim N(0,\mathbb{I}_p)$ and ζ_n is independent of η for all n. The independence holds because $cov(\zeta_n,\eta) = cov\left[\int_0^1 f_n^*(r)dW_p(r), \int_0^1 dW_p(r)\right] = \int_0^1 f_n^*(r)dr = 0$ and both ζ_n and η are normal. That is, $pF_\infty(p,b)$ is equal in distribution to a quadratic form of standard normals with an independent and random weighting matrix.

Let $\mathcal{H}$ be an orthonormal matrix such that $\mathcal{H} = (\eta/\|\eta\|,\Pi)'$ where Π is a $p\times(p-1)$ matrix, then by definition $\mathcal{H}\eta = \|\eta\| e_1$ and

$$\begin{aligned} pF_\infty(p,b) &\overset{d}{=} (\mathcal{H}\eta)' \left(\sum_{n=1}^{\infty} \lambda_n^* (\mathcal{H}\zeta_n)(\mathcal{H}\zeta_n)'\right)^{-1} \mathcal{H}\eta \\ &\overset{d}{=} \|\eta\|^2 e_1' \left(\sum_{n=1}^{\infty} \lambda_n^* (\mathcal{H}\zeta_n)(\mathcal{H}\zeta_n)'\right)^{-1} e_1 \end{aligned}$$

where $e_1 = (1,0,0,\ldots,0,0)'$. Note that $\|\eta\|^2$ is independent of $\mathcal{H}$ and $\mathcal{H}\zeta_n$ has the same distribution as ζ_n, so we can write

$$pF_\infty(p,b) \overset{d}{=} \|\eta\|^2 e_1' \left(\sum_{n=1}^{\infty} \lambda_n^* \zeta_n \zeta_n'\right)^{-1} e_1.$$

Let

$$\sum_{n=1}^{\infty} \lambda_n^* \zeta_n \zeta_n' = \begin{pmatrix} \nu_{11} & \nu_{12} \\ \nu_{21} & \nu_{22} \end{pmatrix}$$

where $\nu_{11} \in \mathbb{R}$ and $\nu_{22} \in \mathbb{R}^{(p-1)\times(p-1)}$. Then

$$pF_\infty(p,b) \overset{d}{=} \frac{\|\eta\|^2}{\nu_{11\cdot 2}} \text{ for } \nu_{11\cdot 2} = \nu_{11} - \nu_{12}\nu_{22}^{-1}\nu_{21}. \tag{10}$$

This is our second representation of $pF_\infty(p,b)$. It shows that $pF_\infty(p,b)$ is equal in distribution to a chi-square variate scaled by an independent and almost surely positive random variable. So $F_\infty(p,b)$ is similar to an F distribution.

As $b \to 0$, we expect $\nu_{11.2}$ to be concentrated around 1. By taking a Taylor expansion $G_p\left(z\nu_{11.2}\right)$ around $G_p\left(z\right)$ and computing the moments of $\nu_{11.2}$, we can prove the following theorem.

Theorem 1 *As $b \to 0$, we have*

$$P\left\{pF_\infty(p,b) \le z\right\} = G_p(z) + A(z)b + o(b) \tag{11}$$

where

$$A(z) = G_p''(z)z^2c_2 - G_p'(z)z\left[c_1 + c_2\left(p-1\right)\right],$$
$$c_1 = \int_{-\infty}^{\infty} k(x)dx, c_2 = \int_{-\infty}^{\infty} k^2(x)dx.$$

There are two terms in $A(z)b$. The term $G_p''(z)z^2c_2b$ arises from the asymptotic mean square error $E\left(\nu_{11.2}-1\right)^2$ of $\nu_{11.2}$ while the term $-G_p'(z)z\left[c_1 + c_2\left(p-1\right)\right]b$ arises from the asymptotic bias $E\left(\nu_{11.2}-1\right)$ of $\nu_{11.2}$. The bias term comes from two sources. The first is the estimating uncertainty of model parameters. This is reflected in the dependence of $\nu_{11.2}$ on the transformed kernel function $k_b^*(\cdot,\cdot)$ rather than the original kernel function $k_b(\cdot)$. This type of bias may be referred to as the demeaning bias as $k_b^*(\cdot,\cdot)$ can be regarded as a demeaned version of $k_b(\cdot)$. The second comes from a dimension adjustment. When $p>1$, $\nu_{11.2}$ is not equal to ν_{11} but its projected version, viz $\nu_{11} - \nu_{12}\nu_{22}^{-1}\nu_{21}$. In contrast, when $p=1$, $\nu_{11.2}$ is equal to ν_{11} and there is no dimension adjustment. Given that this type of bias depends on the dimension of the hypothesis space, we may refer to it as the dimensionality bias.

When $p=1$, Theorem 1 reduces to Theorem 1 in SPJ (2008). The main difference between the scalar case and the multivariate case is the presence of the dimensionality bias. This bias depends on p, the number of restrictions being tested or the dimension of the hypothesis space. As we show later, one of the reasons that the fixed-b approximation is more accurate than the chi-square approximation is that it captures the dimensionality bias.

4 Second-order Correctness of the Fixed-b Approximation under the Small-b Asymptotics

In this section, we show that the fixed-b approximation is high-order correct under the small-b asymptotics where $b \to 0$ and $T \to \infty$ jointly. We first employ a Gaussian location model to illustrate the basic point. We then extend the result to a general GMM setting.

4.1 Gaussian Location Model

Consider a vector time series y_t:

$$y_t = \theta_0 + v_t, t = 1, 2, ..., T, \tag{12}$$

where $y_t = (y_{1t}, ..., y_{dt})'$, $\theta_0 = (\theta_{10}, ..., \theta_{d0})'$, $v_t = (v_{1t}, ..., v_{dt})'$ is a stochastic process with zero mean.

The OLS estimator of θ_0 is the average of $\{y_t\}$, viz $\hat{\theta}_{OLS} = T^{-1} \sum_{t=1}^{T} y_t$. To simplify the presentation, we consider testing linear restrictions $H_0 : R_0\theta_0 = r_0$ against $H_1 : R_0\theta_0 \neq r_0$ for some $p \times d$ matrix R_0. It is easy to generalize the result to nonlinear restrictions. Under the null hypothesis, we have

$$\sqrt{T}\left(R_0\hat{\theta}_{OLS} - r_0\right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_t \text{ for } u_t = R_0 v_t.$$

Let $F_{T,OLS}$ be the F-test version of the Wald statistic based on the OLS estimator:

$$F_{T,OLS} = \left[\sqrt{T}(R_0\hat{\theta}_{OLS} - r_0)\right]' \hat{\Omega}_T^{-1} \left[\sqrt{T}(R_0\hat{\theta}_{OLS} - r_0)\right] /p \tag{13}$$

where $\hat{\Omega}_T$ is defined as in (3) with $\hat{u}_t = R_0(y_t - \hat{\theta}_{OLS})$.

The Gaussian location model is a special case in the GMM setting. The underlying moment condition is $f(y_t, \theta) = y_t - \theta$. The model is exactly identified so $m = d$. The OLS estimator is a GMM estimator with $G_T = -I_d$ and any weighting matrix $\mathcal{W}_T$, say $\mathcal{W}_T = I_d$.

We maintain the following assumption.

Assumption 6 *(i) u_t is a stationary Gaussian process. (ii) For any $c \in \mathbb{R}^d$, the spectral density of $c'u_t$ is bounded above and away from zero in a neighborhood around the origin. (iii) The FCLT holds: $T^{-1/2} \sum_{t=1}^{[Tr]} u_t \to^d \Lambda W_p(r)$.*

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

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

where $\Omega_v = var([v_1', v_2', ..., v_T']')$, $y = [y_1', y_2', ..., y_T']'$ and ℓ_T is a vector of ones. Replacing Ω_v by $\mathbb{I}_T \otimes \mathbb{I}_d$ leads to the OLS estimator $\hat{\theta}_{OLS}$ introduced earlier. Define

$$\Delta = \hat{\theta}_{OLS} - \theta_0 - (\hat{\theta}_{GLS} - \theta_0).$$

Under Assumption 6(i) and (ii), it follows from Grenander and Rosenblatt (1957) that $\hat{\theta}_{OLS}$ and $\hat{\theta}_{GLS}$ are asymptotically equivalent. In addition, simple calculations show that $E[(\hat{\theta}_{GLS} - \theta_0)\Delta'] = \mathbf{0}$ and

$$\begin{aligned}
&E[(\hat{\theta}_{GLS} - \theta)\hat{u}_t'] \\
&= cov\left\{\left[(\ell_T \otimes \mathbb{I}_d)' \Omega_v^{-1} (\ell_T \otimes \mathbb{I}_d)\right]^{-1} (\ell_T \otimes \mathbb{I}_d)' \Omega_v^{-1} v, R_0 \left[\left(\mathbb{I}_T - \ell_T \ell_T'/T\right) \otimes \mathbb{I}_d\right] v\right\} \\
&= E\left[(\ell_T \otimes \mathbb{I}_d)' \Omega_v^{-1} (\ell_T \otimes \mathbb{I}_d)\right]^{-1} (\ell_T \otimes \mathbb{I}_d)' \Omega_v^{-1} v v' \left[\left(\mathbb{I}_T - \ell_T \ell_T'/T\right) \otimes \mathbb{I}_d\right]\} R_0 \\
&= E\left[(\ell_T \otimes \mathbb{I}_d)' \Omega_v^{-1} (\ell_T \otimes \mathbb{I}_d)\right]^{-1} (\ell_T \otimes \mathbb{I}_d)' \left[\left(\mathbb{I}_T - \ell_T \ell_T'/T\right) \otimes \mathbb{I}_d\right]\} R_0 \\
&= \mathbf{0}
\end{aligned}$$

for all t. So $\hat{\theta}_{GLS} - \theta_0$ is independent of both Δ and $\hat{u}_t$.

Let $F_{T,GLS}$ be the F-test version of the Wald statistic based on the GLS estimator:

$$F_{T,GLS} = \left[\sqrt{T}(R_0\hat{\theta}_{GLS} - r_0)\right]' \hat{\Omega}_T^{-1} \left[\sqrt{T}(R_0\hat{\theta}_{GLS} - r_0)\right]/p$$

where $\hat{\Omega}_T$ is the same estimator as in $F_{T,OLS}$ given by (13).

Using the asymptotic equivalence of the OLS and GLS estimators and independence of $\hat{\theta}_{GLS} - \theta_0$ from Δ and $\hat{u}_t$, we can prove the following lemma.

Lemma 1 *Let Assumption 6 hold. Then*

(a) $P\left(pF_{T,GLS} \le z\right) = EG_p\left(z\Xi_T^{-1}\right) + O\left(T^{-1}\right),$

(b) $P\left(pF_{T,OLS} \le z\right) = P\left(pF_{T,GLS} \le z\right) + O\left(T^{-1}\right),$

where

$$\Xi_T = e_T' \left[\Omega^{1/2}\hat{\Omega}_T^{-1}\Omega^{1/2}\right] e_T,\ e_T = \frac{\Omega_{T,GLS}^{-1/2}\sqrt{T}(R_0\hat{\theta}_{GLS} - r_0)}{\left\|\Omega_{T,GLS}^{-1/2}\sqrt{T}(R_0\hat{\theta}_{GLS} - r_0)\right\|}$$

and $\Omega_{T,GLS}$ *is the variance of* $\sqrt{T}(R_0\hat{\theta}_{GLS} - r_0)$.

Lemma 1 shows that the estimation uncertainty of $\hat{\Omega}_T$ affects the distribution of the Wald statistic only through Ξ_T. Taking a Taylor expansion, we have $\Xi_T^{-1} = 1 + L + Q + err$, where err is the approximation error, L is linear in $\hat{\Omega}_T - \Omega$ and Q is quadratic in $\hat{\Omega}_T - \Omega$. The exact expressions for L and Q are not important here but are given in the proof of Theorem 2. Using this stochastic expansion and Lemma 1, we can establish a high-order expansion of the finite sample distribution for the case where $b \to 0$ as $T \to \infty$.

Theorem 2 *Let Assumptions 5 and 6 hold. Assume that* $\sum_{h=-\infty}^{\infty} |h|^q Eu_t u_{t-h}' < \infty$*. If* $b \to 0$ *such that* $bT \to \infty$*, then*

$$P\left(pF_{T,OLS} \leq z\right) = G_p\left(z\right) + A(z)b + (bT)^{-q} G_p'\left(z\right) z\bar{B} + o\left(b\right) + o\left((bT)^{-q}\right) \tag{14}$$

where

$$\bar{B} = tr\left\{B\Omega^{-1}\right\}/p, \;\; B = -g_q \sum_{h=-\infty}^{\infty} |h|^q Eu_t u_{t-h}'$$

and q *and* g_q *are given in Assumption 5 (ii).*

The first term in (14) comes from the standard chi-square approximation of the Wald statistic. The second term captures the demeaning bias, the dimensionality bias, and the variance of the LRV estimator. The third term reflects the usual nonparametric bias of the LRV estimator. In view of its representation, $\bar{B}$ can be regarded as a measurement of the relative nonparametric bias.

Let $\mathcal{X}_p^\alpha$ be the $1-\alpha$ quantile from the χ_p^2 distribution, then, up to smaller order terms,

$$P\left(pF_{T,OLS} > \mathcal{X}_p^\alpha\right) = \alpha - A(\mathcal{X}_p^\alpha)b - (bT)^{-q} G_p'\left(\mathcal{X}_p^\alpha\right)\mathcal{X}_p^\alpha \bar{B}. \tag{15}$$

Since $G_p''\left(\mathcal{X}_p^\alpha\right) < 0$ and $G_p'(\mathcal{X}_p^\alpha) > 0$, all terms in $-A(\mathcal{X}_p^\alpha)b$ are positive. First, the variance term $-G_p''(\mathcal{X}_p^\alpha)\left(\mathcal{X}_p^\alpha\right)^2 c_2$ is positive. This is expected. Using χ_p^2 as the reference distribution does not take into account the randomness of the LRV estimator and the critical values from it tend to be smaller than they should be. As a result, the rejection region is larger, leading to over rejection. Second, the bias term $-G_p'(\mathcal{X}_p^\alpha)\mathcal{X}_p^\alpha c_1$ from demeaning is positive. This type of bias is easier to understand in the scalar case where the LRV is positive. In this case, demeaning effectively dampens the low frequency components and introduces a downward bias into the LRV estimator. The downward bias translates into an increase in the test statistic and leads to over rejection. Finally, the bias term $-G_p'(\mathcal{X}_p^\alpha)\mathcal{X}_p^\alpha c_2(p-1)$ from the dimension adjustment is positive. Intuitively, when $p > 1$, the $p \times p$ matrix $\hat{\Omega}_T$ may become singular in $p-1$ different directions. When that happens, the Wald statistic will blow up and we reject the null hypothesis. So the dimensionality bias also tends to give arise over-rejection. On the other hand, the nonparametric bias term $-(bT)^{-q} G_p'\left(\mathcal{X}_p^\alpha\right)\mathcal{X}_p^\alpha \bar{B}$ may be positive or negative, leading to over-rejection or under-rejection.

Comparing Theorem 1 with Theorem 2, we find that the fixed-b asymptotics captures some terms in the high order expansion of the small-b asymptotics. Let $\mathcal{F}_\infty^\alpha(p,b)$ be the $1-\alpha$ quantile from the distribution $F_\infty(p,b)$, i.e. $P\left(F_\infty(p,b) > \mathcal{F}_\infty^\alpha(p,b)\right) = \alpha$. According to Theorem 1, we have $1 - G_p(p\mathcal{F}_\infty^\alpha(p,b)) - A\left(p\mathcal{F}_\infty^\alpha(p,b)\right)b + o(b) = \alpha$. So as $b \to 0$,

$$P\left(F_{T,OLS} > \mathcal{F}_\infty^\alpha(p,b)\right) = \alpha - (bT)^{-q} G_p'\left(p\mathcal{F}_\infty^\alpha(p,b)\right)p\mathcal{F}_\infty^\alpha(p,b)\bar{B} + o\left(b\right) + o\left((bT)^{-q}\right). \tag{16}$$

Therefore, use of the nonstandard critical value $\mathcal{F}_\infty^\alpha(p,b)$ removes the demeaning bias, dimensionality bias and variance term from the higher order expansion. The size distortion is then of order $O\left((bT)^{-q}\right)$. In contrast, if $\mathcal{X}_p^\alpha/p$ is used as the critical value, the size distortion is of order $O\left((bT)^{-q}\right)+O\left(b\right)$. So when $(bT)^{-q}b^{-1}\rightarrow 0$, using critical value $\mathcal{F}_\infty^\alpha(p,b)$ should lead to size improvements. We have thus shown that critical values from the fixed-b asymptotics are second order correct under the small-b asymptotics.

4.2 General GMM Setting

To establish a high order expansion in the general GMM setting, we establish a stochastic approximation in the appendix:

$$pF_T \equiv pF_{T,L}+\psi_T+\psi_T^*$$

where

$$pF_{T,L}=\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t\right]'\tilde{\Omega}_T^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t\right]$$

is the dominated linear term in the approximation and

$$\tilde{\Omega}_T=\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}k\left(\frac{t-\tau}{bT}\right)\left[u_t-\frac{1}{T}\sum_{s=1}^{T}u_s\right]\left[u_\tau-\frac{1}{T}\sum_{s=1}^{T}u_s\right]'$$

is the kernel estimator of the long run variance of u_t when the mean is assumed to be unknown. In this stochastic approximation, $\psi_T=O_p(1/\sqrt{T})$ does not depend on b and $\psi_T^*=O_p([\sqrt{b}+(bT)^{-q}]/\sqrt{T}+1/T)$.

Define

$$y_t\equiv \mu+u_t$$

for some $\mu\in\mathbb{R}^p$. Then $pF_{T,L}$ is exactly the same as the Wald statistic for testing whether the mean of y_t satisfies $E(y_t)=\mu=0$. Using Theorem 2, we can prove Theorem 3 below.

Theorem 3 *Assume (i)* $plim_{T\rightarrow\infty}\hat{\theta}_T=\theta_0$, *(ii) for sufficiently large* C, $P(|\psi_T|>(\log T)/\sqrt{T})=O(1/\sqrt{T})$ *and* $P\left(|\psi_T^*|>\delta_T/\log T\right)=o\left(\delta_T\right)$ *for* $\delta_T=b+(bT)^{-q}$, *(iii)* $u_t=\phi(v_t,\theta_0)$ *satisfies Assumption 6, (iv)* $\mathcal{W}_T=\mathcal{W}_\infty+O_p(1/\sqrt{T})$ *and* $G_0'\mathcal{W}_\infty G_0$ *is positive definite, (v)* $G_{[rT]}(\tilde{\theta}_T)=rG_0+O_p(1/\sqrt{T})$ *uniformly in* r *for any* $\tilde{\theta}_T$ *between* $\hat{\theta}_T$ *and* θ_0, *(vi) Assumption 5 holds. If* $b\rightarrow 0$ *such that* $bT\rightarrow\infty$, *then*

$$P\left(pF_T\leq z\right)=G_p\left(z\right)+A(z)b+(bT)^{-q}G_p'\left(z\right)z\bar{B}+O(T^{-1/2}\log T)+o\left(b\right)+o\left((bT)^{-q}\right) \quad (17)$$

where the $O\left(T^{-1/2}\log T\right)$ *term does not depend on* b.

Some comments on the assumptions are in order. Assumption (ii) is a high level assumption. It holds with sufficient mixing and moment conditions. This type of result can often be proved using results in Doukhan (1995, Section 1.2.2). The Gaussianity assumption in (iii) is made for convenience. It greatly simplifies the proof and makes our arguments much more transparent. There is no need to use general Edgeworth expansion techniques to establish the high order expansion. The Gaussianity assumption be relaxed but at the cost of much greater complexity in terms of both proof strategies and technical conditions. See Sun and Phillips (2009) for the expansion without the Gaussianity assumption. The expansion there contains more terms but reduces to the above expansion when the Gaussian assumption holds. Assumption (iv) requires the weighting matrix to converge at the parametric rate. It rules out the two step GMM estimator with the weighting matrix $\mathcal{W}_T$ converging only at a nonparametric rate. Again, we make this assumption in order to greatly simplify our proof. In practice, we may ignore the convergence rate of $\mathcal{W}_T$ even if it is a nonparametric rate and use the formula derived here. Assumption (v) is stronger than Assumption 2. It holds if $\sqrt{T}[G_{[rT]}(\theta_0) - rG_0]$ satisfies a functional central limit theorem and $\partial G_T(\theta)/\partial\theta'$ satisfies a uniform law of large numbers.

Theorem 3 shows that the high order expansion for the location model remains valid for GMM estimators. The only difference is the term of order $O\left(T^{-1/2}\log T\right)$. This term reflects, among others, the higher order terms in the linear representation of the GMM estimator and the nonlinearity of the restrictions being tested. With more sophisticate and tedious arguments as in Sun and Phillips (2009), the term $O(T^{-1/2}\log T)$ can be reduced to $O(T^{-1/2})$. Here we are content with the weaker result as our main interest is to capture the effect of b on the sampling distribution of F_T.

Given the similarity of the two expansions, the qualitative results for the location model in the previous subsection apply to the GMM setting. In particular, the fixed-b critical values are high order correct under the conventional small-b asymptotics.

5 Standard F Approximation

In this section, we establish a standard F approximation that is second order correct under the conventional small-b asymptotics.

For some constant κ to be determined, define $F_T^* = F_T/\kappa$ to be a modified Wald statistic. We want to approximate the distribution of F_T^* by a standard F distribution $F_{p,K}$. Like in the conventional χ_p^2 approximation, the first degree of freedom of the approximating F distribution is the number of joint hypotheses p. The second degree of freedom K together with κ will be chosen to capture the high order term in the expansion given in (14).

Let $\mathcal{F}^{\alpha}_{p,K}$ be the $1-\alpha$ quantile of the F distribution $F_{p,K}$, i.e. $P(F_{p,K} > \mathcal{F}^{\alpha}_{p,K}) = \alpha$, we choose κ and K such that

$$1 - G_p\left(\kappa \mathcal{F}^{\alpha}_{p,K}\right) - A(\kappa \mathcal{F}^{\alpha}_{p,K})b = \alpha + o(b). \tag{18}$$

That is, we choose κ and K to remove the term of order $O(b)$ in the expansion of $P(F_T^* > \mathcal{F}^{\alpha}_{p,K})$.

To derive explicit expressions for κ and K, we first expand the F critical value $\mathcal{F}^{\alpha}_{p,K}$ around the corresponding χ^2_p critical value $\mathcal{X}^{\alpha}_p/p$. When $K \propto 1/b$, we have, by definition:

$$\alpha = 1 - EG_p\left(p\mathcal{F}^{\alpha}_{p,K}\frac{\chi^2_K}{K}\right).$$

Taking a second order Taylor expansion, we immediately obtain

$$\mathcal{F}^{\alpha}_{p,K} = \frac{\mathcal{X}^{\alpha}_p}{p} - \frac{G''_p\left(\mathcal{X}^{\alpha}_p\right)\left(\mathcal{X}^{\alpha}_p\right)^2}{pG'_p\left(\mathcal{X}^{\alpha}_p\right)}\frac{1}{K} + o(b).$$

Using this and expanding (18) around $\mathcal{X}^{\alpha}_p$, we have:

$$G''_p\left(\mathcal{X}^{\alpha}_p\right)\left(\mathcal{X}^{\alpha}_p\right)^2\left(\frac{\kappa}{K} - c_2 b\right) - G'\left(\mathcal{X}^{\alpha}_p\right)\mathcal{X}^{\alpha}_p\left\{\kappa - 1 - \left[c_1 + c_2\left(p-1\right)\right]b\right\} = o(b).$$

So we can choose any κ and K combination that satisfies

$$\kappa = 1 + \left[c_1 + c_2\left(p-1\right)\right]b + o(b) \text{ and } K = \frac{1}{bc_2}\left(1 + o\left(b\right)\right).$$

For the Bartlett kernel, $c_1 = 1$, $c_2 = 2/3$; For the Parzen kernel, $c_1 = 3/4$, $c_2 = 0.539285$; For the quadratic spectral (QS) kernel, $c_1 = 1.25$, $c_2 = 1$.

Theorem 4 *Let assumptions in Theorem 2 or 3 hold. Let $K = K^* - p + 1$ or K^* for*

$$K^* = \max(\lceil 1/\left(bc_2\right)\rceil, p)$$

and

$$\kappa = \frac{\exp\left(b\left[c_1 + \left(p-1\right)c_2\right]\right) + \left(1 + b\left[c_1 + \left(p-1\right)c_2\right]\right)}{2}$$

where $\lceil\cdot\rceil$ is the ceiling function. As $b \to 0$, we have

$$P\left(F_T > \kappa\mathcal{F}^{\alpha}_{p,K}\right) = \alpha + o(b) + O\left((bT)^{-q}\right) + O\left(\log T/\sqrt{T}\right). \tag{19}$$

The parameter κ corrects for the demeaning bias and the dimensionality bias. It can be motivated from a Bartlett type correction. See Bartlett (1937, 1954) for the original papers and Cribari-Neto and Cordeiro (1999) for a more recent survey. The argument goes as follows. Suppose that $F_{T,L} \to^d \chi^2_p/p$ and $EF_{T,L} = C$ for some constant C, then as $b \to 0$,

$F_{T,L}/C$ is closer to the χ_p^2/p distribution than the original $F_{T,L}$. When $u_t \sim iidN(0,\sigma^2)$, we can show that $C = \kappa + o(b)$. So we can choose $C = \kappa$ to make the correction and $F_T^* = F_T/\kappa$ becomes a Bartlett-corrected Wald statistic.

Theorem 4 goes one more step beyond the Bartlett correction. Instead of approximating the distribution of F_T^* by the normalized chi-squared distribution χ_p^2/p, Theorem 4 approximates it by the standard F distribution with degrees of freedom p and K.

Theorem 4 adjusts the value of K^* to ensure that $K = K^* - p + 1 \geq 1$. The parameter K is asymptotically equivalent to the inverse of the asymptotic variance of the LRV estimator. It can be called the "equivalent degree of freedom" of the LRV estimator. As b decreases, i.e. as the degree of smoothing increases, the variance decreases and K increases. In other words, the higher the degree of freedom is, the larger the degree of smoothing is, and the smaller the variance is.

The scaled F critical value $\kappa\mathcal{F}_{p,K}^{\alpha}$ is larger than the standard critical value from χ_p^2/p for two reasons. First, $\mathcal{F}_{p,K}^{\alpha}$ is larger than $\mathcal{X}_p^{\alpha}/p$, the corresponding critical value from χ_p^2/p due to the presence of a random denominator in the F distribution. Second, the correction factor κ is larger than 1. As b increases, both the correction factor and F critical value $\mathcal{F}_{p,K}^{\alpha}$ increase. As a result, the second-order correct critical value $\kappa\mathcal{F}_{p,K}^{\alpha}$ is an increasing function of b.

In view of (16) and (19), both the nonstandard critical value $\mathcal{F}_{\infty}^{\alpha}(p,b)$ and the corrected F critical value $\kappa\mathcal{F}_{p,K}^{\alpha}$ are second order correct under the conventional small-b asymptotics. Unreported numerical work shows that when b is small, the two critical values $\mathcal{F}_{\infty}^{\alpha}(p,b)$ and $\kappa\mathcal{F}_{p,K}^{\alpha}$ are very close to each other. See also the simulation study in Section 7.1.

For the adjusted F critical value $\kappa\mathcal{F}_{p,K}^{\alpha}$, the choices of $K = K^* - p + 1$ and K^* are asymptotically equivalent but make a difference when b is not small. For the Bartlett kernel, using $K = K^*$ brings the F critical value $\kappa\mathcal{F}_{p,K}^{\alpha}$ closer to the nonstandard critical value $\mathcal{F}_{\infty}^{\alpha}(p,b)$. For the Parzen and QS kernels, $K = K^* - p + 1$ leads to a smaller difference between $\kappa\mathcal{F}_{p,K}^{\alpha}$ and $\mathcal{F}_{\infty}^{\alpha}(p,b)$. We will use $K = K^*$ for the Bartlett kernel and $K = K^* - p + 1$ for the Parzen and QS kernels in our simulation study.

For convenience, we refer to the test based on the test statistic F_T^* and the F critical value $\mathcal{F}_{p,K}^{\alpha}$ as the asymptotic F test. This of course is the same as the test based on the original Wald statistic and the scaled F critical value $\kappa\mathcal{F}_{p,K}^{\alpha}$.

As $b \to 0$, some simple calculations show that

$$\begin{aligned}\frac{\kappa \mathcal{F}_{p,K}^{\alpha} - \mathcal{X}_p^{\alpha}/p}{\mathcal{X}_p^{\alpha}/p} &= \kappa \left(1 - \frac{G_p''\left(\mathcal{X}_p^{\alpha}\right)\mathcal{X}_p^{\alpha}}{G_p'\left(\mathcal{X}_p^{\alpha}\right)} bc_2\right) - 1 + o(b) \\ &= \left[c_1 + c_2\left(p-1\right)\right] b - \left(\frac{1}{2}p - \mathcal{X}_p^{\alpha} - 1\right) c_2 b + o(b) \\ &= \left[c_1 + \frac{1}{2}c_2\left(\frac{p}{2} + \mathcal{X}_p^{\alpha}\right)\right] b + o(b).\end{aligned}$$

Similarly,

$$\frac{\mathcal{F}_{\infty}^{\alpha}(p,b) - \mathcal{X}_p^{\alpha}/p}{\mathcal{X}_p^{\alpha}/p} = \left[c_1 + \frac{1}{2}c_2\left(\frac{p}{2} + \mathcal{X}_p^{\alpha}\right)\right] b + o(b).$$

So to the order of $O(b)$, the percentage adjustment of the critical value increases with p. This is true for both the scaled F critical value $\kappa \mathcal{F}_{p,K}^{\alpha}$ and the nonstandard critical value $\mathcal{F}_{\infty}^{\alpha}(p,b)$. Our result is especially interesting when the number of restrictions is large. In this case, the size distortion of the usual Wald test is large. This is due to the presence of the dimensionality bias. The nonstandard critical value and F critical value automatically correct for the dimensionality problem. Our results provide an explanation of the finite sample results reported by Ravikumar, Ray and Savin (2004) who find that the fixed-b asymptotic approximation can substantially reduce size distortion in tests of joint hypotheses especially when the number of hypotheses being tested is large. See also Ray and Savin (2008) and Ray, Savin and Tiwari (2009).

6 Testing-Optimal Bandwidth Choice

In this section, we consider selecting the bandwidth parameter b for the asymptotic F test in the GMM setting. It is standard practice to select b to minimize the MSE of the LRV estimator. However, the MSE-optimal b is not optimal for hypothesis testing. We propose to select b to minimize the type II error while controlling for the type I error. Our testing-orientated criterion addresses the central concern of hypothesis testing.

It follows from (17) that the type I error of the α-level F test can be approximated by

$$e_I(b) = \alpha - (bT)^{-q} G_p'\left(\mathcal{X}_p^{\alpha}\right)\mathcal{X}_p^{\alpha}\bar{B}.$$

For Gaussian location models, this approximation has an error of order $o(b)$ and $o\left((bT)^{-q}\right)$ as $b \to 0$ such that $bT \to \infty$. For Gaussian GMM models, there is an additional error term of order $O(\log T/\sqrt{T})$. But this term does not depend on b and can thus be ignored for the purpose of optimal b selection.

To obtain the type II error of the F test, we have to first specify the alternative hypothesis. We consider the standard local alternative hypothesis of the form:

$$H_1\left(\delta^2\right): r\left(\theta_0\right) = \Lambda \tilde{c}/\sqrt{T}$$

for some vector $\tilde{c} \in \mathbb{R}^p$ such that $\|\tilde{c}\|^2 = \delta^2$. Under $H_1\left(\delta^2\right)$ and the small-b asymptotics, we have $F_T \to^d \chi_p^2\left(\delta^2\right)/p$, the normalized noncentral chi-square distribution with noncentrality parameter δ^2. Hence the type II error of the usual Wald test depends on the local alternative parameter $\tilde{c}$ only through its squared length $\|\tilde{c}\|^2$. When we do not know the direction of the local alternative, it is reasonable to assume that $\tilde{c}$ is uniformly distributed on the sphere $\mathcal{S}_p\left(\delta^2\right) = \{\tilde{c} \in \mathbb{R}^p : \|\tilde{c}\|^2 = \delta^2\}$. We will maintain this specification. In other words, the type II error we will obtain is the average of the type II error associated with each point $\tilde{c}$ on the sphere $\mathcal{S}_p\left(\delta^2\right)$.

It remains to specify the noncentrality parameter δ^2. Since it can not be consistently estimated from the data, we choose δ^2 such that the local power of the standard Wald test is 75% under the first order asymptotics. More specifically, δ^2 satisfies $P(\varkappa \leq \mathcal{X}_p^\alpha) = 75\%$ where $\varkappa \sim \chi_p^2\left(\delta^2\right)$. This strategy is similar to that used in the optimal testing literature. In the absence of a uniformly most powerful test, it is often recommended to pick a reasonable point under the alternative and construct an optimal test against this particular point alternative. It is hoped that the resulting test, although not uniformly most powerful, is reasonably close to the power envelope. Here we use the same idea and select the radius of the sphere according to the power requirement. We hope that the smoothing parameter that is optimal for the chosen radius also works well for other points under the alternative hypothesis. This is confirmed by our Monte Carlo study.

Theorem 5 *Let the assumptions in Theorem 3 hold. Consider the local alternative hypothesis* $H_1\left(\delta^2\right): r\left(\theta_0\right) = \Lambda \tilde{c}/\sqrt{T}$ *where* $\tilde{c}$ *is uniformly distributed on* $\mathcal{S}_p\left(\delta^2\right) = \{\tilde{c} \in \mathbb{R}^p : \|\tilde{c}\|^2 = \delta^2\}$. *Under the small-b asymptotics, the type II error of the asymptotic* F *test is*

$$P(F_T \leq \kappa \mathcal{F}_{p,K}^\alpha) = e_{II}(b) + O(\log T/\sqrt{T}) + o(b) + o\left((bT)^{-q}\right)$$

where the $O(\log T/\sqrt{T})$ *term does not depend on* b,

$$e_{II}(b) = G_{p,\delta^2}\left(\mathcal{X}_p^\alpha\right) + (bT)^{-q} G'_{p,\delta^2}\left(\mathcal{X}_p^\alpha\right)\mathcal{X}_p^\alpha \bar{B} + \frac{\delta^2}{2} G'_{(p+2),\delta^2}\left(\mathcal{X}_p^\alpha\right)\mathcal{X}_p^\alpha c_2 b$$

and $G'_{l,\delta^2}(z)$ *is the pdf of noncentral* χ^2 *distribution with degrees of freedom* l *and noncentrality parameter* δ^2.

There are three terms in the type II error. The first term in e_{II} reflects the usual first order approximation to the type II error. The second term is due to the nonparametric bias of the LRV estimator. This bias has opposite effects on the type I and type II errors. The third term reflects the difference in curvature of the null distribution and alternative distribution at the critical value $\mathcal{X}_p^\alpha$.

Given the approximate measures of the type I and type II errors, we can select the bandwidth parameter b to solve the constrained minimization problem:

$$b^* = \arg\min e_{II}(b) \text{ s.t. } e_I(b) \leq \tau\alpha$$

for some parameter $\tau > 1$. The presence of τ allows the approximate type I error to be different from the nominal type I error. Depending on our tolerance towards this difference, we may choose τ to be small or large. Hence we can call τ the permitted tolerance. Sun (2011) imposes a similar upper bound when selecting the smoothing parameter in nonparametric series LRV estimation.

The constrained minimization problem is easy to solve. The testing-optimal bandwidth is

$$b^* = \begin{cases} \left[\dfrac{2qG'_{p,\delta^2}(\mathcal{X}_p^\alpha)|\bar{B}|}{\delta^2 G'_{(p+2),\delta^2}(\mathcal{X}_p^\alpha)c_2}\right]^{\frac{1}{q+1}} T^{-\frac{q}{q+1}}, & \bar{B} > 0 \\ \left[\dfrac{G'_p(\mathcal{X}_p^\alpha)\mathcal{X}_p^\alpha|\bar{B}|}{(\tau-1)\alpha}\right]^{1/q} \dfrac{1}{T}, & \bar{B} \leq 0 \end{cases}$$

The testing-optimal b is fundamentally different from the MSE-optimal b. First, the testing-optimal b depends on the direction of $\bar{B}$ where the MSE-optimal b does not. The direction of the bias has a different impact on the test statistic and as a result on the type I and type II errors of the test. The testing-optimal b reflects this. By construction, the MSE-optimal b does not depend on the direction of the nonparametric bias. Second, the testing-optimal b has a different decaying rate from the MSE-optimal b. The MSE-optimal b is of order $O(T^{-2q/(2q+1)})$. When $\bar{B} > 0$, the testing-optimal b is of larger order than the MSE-optimal b. This is also true when $\bar{B} \leq 0$, provided that τ is close enough to 1. More specifically, when $\tau - 1 = o(T^{-q/(2q+1)})$, the testing-optimal b is larger than the MSE-optimal b by an order of magnitude regardless of whether $\bar{B} > 0$ or not. So when the permitted tolerance on the type I error is low, under-smoothing is required for hypothesis testing, compared to the point estimation of the LRV matrix. Third, the testing-optimal b depends on the null hypothesis and alternative hypothesis being considered. The dependence factors in via the relative bias $\bar{B}$, which depends on the direction of the restriction matrix R_0, and the noncentrality parameter δ^2, which captures the departure of the alternative from the null. By definition, the MSE-optimal b does not depend on the hypotheses being considered.

The testing-optimal bandwidth can be written as $b^* = b^*(\bar{B})$ where $\bar{B} = tr\left(B\Omega^{-1}\right)/p$. The parameter $\bar{B}$ is unknown but could be estimated by a standard plug-in procedure based on a simple model like a VAR(1) or univariate AR(1). See Andrews (1991). This method achieves a valid order of magnitude and the procedure is obviously analogous to conventional data-driven methods for HAR estimation.

7 Simulation Study

This section provides some simulation evidence on the finite sample performance of the asymptotic F test using the smoothing parameter that minimizes the approximate type II error while controlling for the approximate type I error.

We consider the following data generating process:

$$y_t = \gamma + x_t'\beta + \varepsilon_t$$

where x_t is a 4×1 vector process and x_t and ε_t follow either an AR (1) process

$$x_{t,j} = \rho x_{t-1,j} + \sqrt{1-\rho^2}e_{t,j},\ \varepsilon_t = \rho\varepsilon_{t-1} + \sqrt{1-\rho^2}e_{t,0}$$

or an MA(1) process

$$x_{t,j} = \rho e_{t-1,j} + \sqrt{1-\rho^2}e_{t,j},\ \varepsilon_t = \rho e_{t-1,0} + \sqrt{1-\rho^2}e_{t,0}.$$

The error term $e_{t,j} \sim iidN(0,1)$ across t and j. Throughout we are concerned with testing for the regression parameter β and set $\gamma = 0$ without the loss of generality. We take $\rho = 0.0, 0.25$, 0.50 and 0.75.

Let $\theta = (\gamma', \beta')'$. We estimate θ by the OLS estimator. Since the model is exactly identified, the weighted matrix $\mathcal{W}_T$ becomes irrelevant. Let $\tilde{x}_t' = [1, x_t']$ and $\tilde{X} = [\tilde{x}_1, ..., \tilde{x}_T]'$, then the OLS estimator is $\hat{\theta}_T - \theta_0 = -G_T^{-1}g_T(\theta_0)$ where $G_T = -\tilde{X}'\tilde{X}/T$, $G_0 = E(G_T)$, $g_T(\theta_0) = T^{-1}\sum_{t=1}^T \tilde{x}_t\varepsilon_t$.

We consider the following null hypotheses:

$$H_{0p} : \beta_1 = ... = \beta_p = 0$$

for $p = 1, 2, 3, 4$. The corresponding restriction matrix $R_{0p} = \mathbb{I}_5(2 : p+1, :)$, i.e., row 2 to row $p+1$ of the identity matrix $\mathbb{I}_5$. The local alternative hypothesis is $H_{1p}\left(\delta^2\right) : R_{0p}\theta = c_p/\sqrt{T}$ where $c_p = \Lambda_{0p}\tilde{c}$, Λ_{0p} is the matrix square root of the LRV of $R_{0p}G_0^{-1}\tilde{x}_t\varepsilon_t$, and $\tilde{c}$ is uniformly distributed over the sphere $\mathcal{S}_p\left(\delta^2\right)$, that is, $\tilde{c} = \delta\xi/\left\|\xi\right\|, \xi \sim N(0, \mathbb{I}_p)$. More specifically,

$$H_{1p}\left(\delta^2\right) : (\beta_1, ..., \beta_p) = c_p/\sqrt{T},\ \beta_{p+1} = ... = \beta_4 = 0.$$

To explore the finite sample size of the tests, we generate data under the null hypothesis. To compare the power of the tests, we generate data under the local alternative. Since our test statistic is invariant to the value of θ, we can impose the null hypothesis given above by setting $\theta_0 = 0$. Let

$$F_T(c_p) = \left[\sqrt{T} R_{0p} \hat{\theta}_T + c_p\right]' \hat{\Omega}_T^{-1} \left[\sqrt{T} R_{0p} \hat{\theta}_T + c_p\right] / p$$

where $\hat{\Omega}_T$ is defined in (3) and all the estimates are computed under the null DGP. Then $F_T(0)$ is the Wald statistic under the null H_{0p}. Note that $\hat{\Omega}_T$ is invariant to the value of θ_0, $F_T(c_p)$ has the same distribution as the Wald statistic under local alternative H_{1p}. So for the linear regression model, there is no need to simulate the data generating process under both the null and the local alternative.

We consider two significance levels $\alpha = 5\%$ and $\alpha = 10\%$ and three different sample sizes $T = 100, 200, 500$. We employ three commonly-used positive semi-definite kernels: Bartlett, Parzen, and QS kernels. The number of simulation replications is 10000.

7.1 A fixed sequence of smoothing parameters

In order to disentangle the effect of smoothing parameter choices from that of distribution approximations, we consider a sequence of fixed b values between 0.01 and 1 with increment 0.02. We compute $F_T(0)$ and employ three different critical values: $\mathcal{X}_p^\alpha / p$, $\kappa \mathcal{F}_{p,K}^\alpha$ and $\mathcal{F}_\infty^\alpha \left(\equiv \mathcal{F}_\infty^\alpha (p, b)\right)$, which come from the standard χ^2 distribution, the standard F distribution and the nonstandard fixed-b asymptotic distribution, respectively.

We focus on the size properties in this subsection. Figures 1 and 2 graph the empirical type I errors of the three tests against the bandwidth parameter b when $\rho = 0$. The nominal type I error is $\alpha = 0.05$. Figure 1 reports the results for the Bartlett kernel. It is clear that the empirical type I error of the F test is very close to that of the nonstandard test when $b \leq 0.4$ and for all values of p considered. A direct implication is that the F critical value $\kappa \mathcal{F}_{p,K}^\alpha$ is very close to the nonstandard critical value $\mathcal{F}_\infty^\alpha$ when $b \leq 0.4$ and $p = 1, 2, 3, 4$. When $b > 0.4$ and $p = 1, 2$, the F critical value $\kappa \mathcal{F}_{p,K}^\alpha$ is too large compared to the nonstandard critical value $\mathcal{F}_\infty^\alpha$. When $p = 3$ and 4, $\kappa \mathcal{F}_{p,K}^\alpha$ and $\mathcal{F}_\infty^\alpha$ are close to each other even when $b > 0.4$.

Figure 2 reports the results for the Parzen kernel. The same pattern for the Bartlett kernel holds when $b \leq 0.3$. The F critical value and the nonstandard critical value are very close to each other for all p values considered. This remains to be true for the QS kernel when $b \leq 0.2$. The figure for the QS kernel is similar to that for the Parzen kernel and is omitted. By simulating the nonstandard critical value and comparing it directly with the corresponding $\kappa \mathcal{F}_{p,K}^\alpha$, we have also found direct evidence that these two sets of critical values are close to each other when b is not large.

Both Figure 1 and Figure 2 clearly demonstrate that the chi-square critical value is too small. This is especially the case when the number of joint hypotheses is larger than 1.

To sum up, we find that the F critical value and the nonstandard critical value are close to each other when $b \leq 0.4, 0.3, 0.2$ for the Bartlett kernel, the Parzen kernel and the QS kernel, respectively. This is true for both $\alpha = 5\%$ and 10% and for up to four joint hypotheses. The difference in the upper bounds of b reflects the different shapes of three kernels.

7.2 Data-driven smoothing parameter choice

In this subsection, we follow the recommendation in the previous section and consider data-driven choices of b. We consider three different values of the tolerance parameter: $\tau = 110\%, 115\%$ and 120%, and set δ_o such that the power of the test when the LRV is known is 75%. This choice of δ_o may not coincide with the true noncentrality parameters δ. We consider a sequence of δ's in order to obtain the power curve. In effect, our procedure aims at a particular local alternative $H_{1p}\left(\delta_o^2\right)$. We hope the smoothing parameter obtained under $H_{1p}\left(\delta_o^2\right)$ works well for other local alternatives $H_{1p}\left(\delta^2\right)$. This is confirmed by our simulation study.

We examine the finite sample performance of the Wald type tests for different smoothing parameter and reference distribution combinations. The first one is the asymptotic F test, which is based on Wald statistic and uses the testing-optimal b and the F critical value $\kappa\mathcal{F}_{p,K}^{\alpha}$. The testing-optimal b is implemented via the VAR(1) plug-in procedure. The second one is the conventional Wald test, which is based on the Wald statistic and uses the MSE-optimal b and the chi-square critical value $\mathcal{X}_p^{\alpha}/p$. The MSE-optimal b is implemented using the VAR(1) plug-in procedure in Andrews (1991). The last one sets $b = 1$ and uses the nonstandard fixed-b critical value $\mathcal{F}_{\infty}^{\alpha}$. We also consider a hybrid testing procedure, which is based on the Wald statistic F_T and uses the MSE-optimal b and the critical value $\kappa\mathcal{F}_{p,K}^{\alpha}$. The difference between the hybrid test and the asymptotic F test lies in the bandwidth parameter b used. The difference between the hybrid test and the standard Wald test lies in the critical values used. The four methods are referred to as 'b-opt', 'b-mse', 'b-max', 'b-mix' respectively in the tables.

Table 1 gives the type I errors of the four testing methods for the AR(1) regressors and error with sample size $T = 100$, tolerance parameter $\tau = 115\%$. The significance level is 5%, which is also the nominal type I error. Several patterns emerge. First, as it is clear from the table, the conventional method has a large size distortion. The size distortion increases with both the error dependence and the number of restrictions being tested. The size distortion can be very severe. Second, the size distortion of the b-opt, b-max and b-mix 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. Third, compared with the b-max test, the b-opt test has somewhat larger size distortion than the b-max test, especially when the process has strong persistence and the number of restrictions is relatively large. This is well expected as the b-max test is designed to achieve the smallest possible size distortion at the cost of power loss. Nevertheless, when the Bartlett kernel is used and p is large, the b-opt test is more accurate in size than the b-max test.

Table 2 reports the average and standard deviation of the selected b values with the same parameter configurations as in Table 1. The average and standard deviation are computed over 10000 simulation replications. It is clear that for both the testing-oriented criterion and the MSE criterion, the average of selected b values increases with the error dependence. In general, the testing-optimal b is larger than the MSE-optimal b. This is consistent with our theoretical prediction. It also partly explains why the type I error of the b-opt test is smaller than that of the b-mix test. Furthermore, the range of selected b values is in general within the range that the F critical value is very close to the corresponding nonstandard fixed-b critical value.

In deriving the testing-optimal choice b, we impose an upper bound on the approximate type I error. Due to the presence of approximation errors, this may not translate into the same upper bound on the empirical type I error. This is demonstrated in Table 1, as the asymptotic F test can still have some size distortion. The quality of approximation depends on the persistence of the time series. When the time series is highly persistent, the first order asymptotic bias of the LRV estimator may not approximate the finite sample bias very well. As a result, the approximate type I error, which is based on the first order asymptotic bias, may not fully capture the empirical type I error. So it is important to keep in mind that the empirical type I error may still be larger than the nominal type I error even if we exert some control over the approximate type I error.

Table 3 presents the simulated type I errors for MA(1) regressors and error. The qualitative observations for the AR(1) case remain valid. In fact, these qualitative observations hold for other parameter configurations such as different sample sizes and significance levels. All else being equal, the size distortion of the b-opt test for $\tau = 120\%$ is slightly larger than that for $\tau = 115\%$. This is expected as we have a higher tolerance for the type I error when the value of τ is larger. Similarly, the size distortion of the b-opt test for $\tau = 110\%$ is slightly smaller than that for $\tau = 115\%$.

To save space, we do not report the figures that compare the size-adjusted power of different tests, but make a few brief comments here. Since the test statistics differ only in terms of the bandwidths used and the size of each test is adjusted, the power comparison effectively compares the impact of bandwidth choice on power. Since the b-max test employs the largest bandwidth, it is less powerful than the b-opt and b-mse tests. The power loss can be substantial. However, there are cases where a larger bandwidth is called for to reduce over-rejection. The power of the b-opt test is more or less the same as the conventional Wald test, i.e. the b-mse test, which in turn has the same size-adjusted power as the b-mix test. This can be explained by comparable bandwidths selected by the b-opt and b-mse tests.

8 Conclusion

On the basis of the fixed-b asymptotics and higher order small-b asymptotics, the paper proposes a new asymptotic F test in the GMM framework where the moment conditions may exhibit general forms of serial dependence. The asymptotic F test employs a finite sample corrected Wald statistic and uses an F distribution as the reference distribution. It is as easy to implement as the standard Wald test. There is no extra computing cost.

To make the F test operational, the paper develops a method for bandwidth choice that addresses the central concern of hypothesis testing. The testing-optimal bandwidth minimizes the asymptotic type II error while controlling for the asymptotic type I error. Simulations show that the F test with data-driven testing-optimal bandwidth performs very well in finite samples. It has a much smaller size distortion than the conventional Wald test while retaining the power of the latter test.

We recommend using the asymptotic F test with testing-optimal bandwidth parameter in practical situations. At a minimum, when the MSE-optimal bandwidth is used, the Wald statistic should be corrected and an F-distribution should be used as the reference distribution.

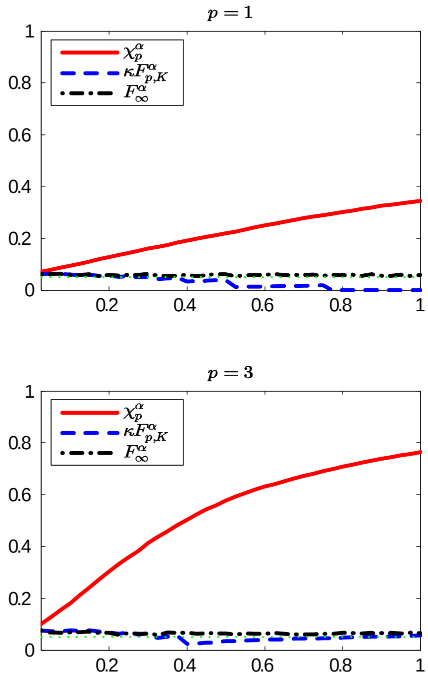

Figure 1: Empirical Type I errors of different 5% tests against the bandwidth parameter b with the Bartlett kernel and for $T = 100$ and $\rho = 0$.

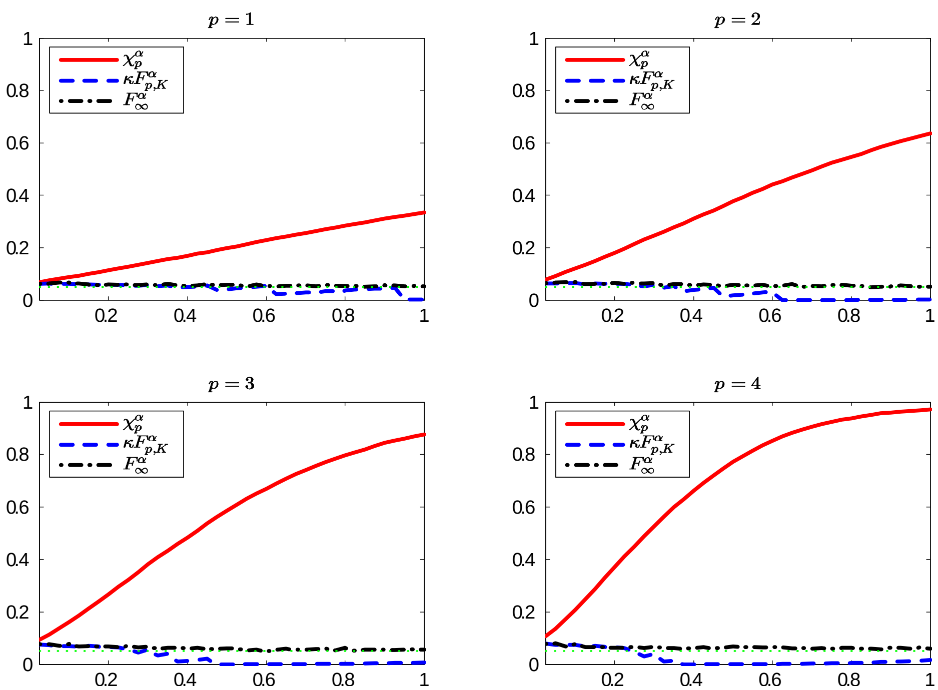

Figure 2: Empirical Type I errors of different 5% tests against the bandwidth parameter b with the Parzen kernel and for $T = 100$ and $\rho = 0$.

Table 1: Empirical Type I error of different tests for AR(1) regressors and error with $T = 100, \tau = 1.15$

	b-OPT	**b-MAX**	**b-MSE**	**b-MIX**	**b-OPT**	**b-MAX**	**b-MSE**	**b-MIX**
Bartlett								
	p=1				p=2			
ρ=0	0.064	0.058	0.066	0.064	0.073	0.062	0.075	0.073
ρ=0.25	0.074	0.063	0.077	0.074	0.093	0.066	0.099	0.090
ρ=0.50	0.106	0.075	0.113	0.094	0.131	0.090	0.159	0.121
ρ=0.75	0.113	0.108	0.219	0.116	0.120	0.147	0.357	0.158
	p=3				p=4			
ρ=0	0.085	0.069	0.090	0.086	0.095	0.070	0.103	0.099
ρ=0.25	0.109	0.076	0.119	0.106	0.127	0.082	0.143	0.123
ρ=0.50	0.148	0.107	0.207	0.141	0.169	0.115	0.261	0.169
ρ=0.75	0.133	0.186	0.502	0.202	0.162	0.226	0.640	0.253
Parzen								
	p=1				p=2			
ρ=0	0.061	0.059	0.068	0.061	0.068	0.054	0.081	0.066
ρ=0.25	0.067	0.058	0.075	0.066	0.079	0.058	0.102	0.078
ρ=0.50	0.082	0.066	0.106	0.082	0.098	0.066	0.151	0.101
ρ=0.75	0.119	0.088	0.173	0.121	0.139	0.096	0.271	0.153
	p=3				p=4			
ρ=0	0.074	0.054	0.098	0.074	0.079	0.057	0.112	0.080
ρ=0.25	0.087	0.056	0.129	0.087	0.095	0.063	0.154	0.094
ρ=0.50	0.111	0.069	0.200	0.116	0.119	0.078	0.258	0.130
ρ=0.75	0.150	0.105	0.374	0.179	0.149	0.124	0.484	0.204
QS								
	p=1				p=2			
ρ=0	0.062	0.060	0.067	0.060	0.068	0.052	0.080	0.065
ρ=0.25	0.067	0.060	0.075	0.066	0.079	0.055	0.100	0.079
ρ=0.50	0.083	0.067	0.105	0.084	0.100	0.060	0.147	0.103
ρ=0.75	0.121	0.086	0.169	0.123	0.144	0.076	0.264	0.156
	p=3				p=4			
ρ=0	0.076	0.060	0.095	0.074	0.080	0.053	0.109	0.080
ρ=0.25	0.089	0.056	0.126	0.087	0.096	0.059	0.148	0.095
ρ=0.50	0.114	0.067	0.193	0.118	0.125	0.058	0.249	0.132
ρ=0.75	0.162	0.086	0.361	0.185	0.174	0.083	0.467	0.214

Note: b-OPT: the asymptotic F test with testing-optimal b implemented by a VAR(1) plug-in procedure; b-MSE: the standard Wald test with MSE-optimal b implemented by a VAR(1) plug-in procedure; b-MAX: the KV test with b=1; b-MIX: the asymptotic F test with MSE-optimal b.

Table 2: Sample mean and standard derivation of the testing-optimal and MSE-optimal bandwidths over simulation replications under AR(1) regressors and error with $T = 100, \tau = 1.15$

	$\bar{b}_{OPT}$	$std\,(b_{OPT})$	$\bar{b}_{MSE}$	$std\,(b_{MSE})$	$\bar{b}_{OPT}$	$std\,(b_{OPT})$	$\bar{b}_{MSE}$	$std\,(b_{MSE})$
Bartlett								
	p=1				p=2			
ρ=0	0.003	0.006	0.004	0.002	0.004	0.005	0.004	0.002
ρ=0.25	0.005	0.011	0.011	0.008	0.007	0.010	0.011	0.008
ρ=0.50	0.029	0.049	0.046	0.029	0.039	0.049	0.046	0.029
ρ=0.75	0.262	0.265	0.217	0.154	0.356	0.277	0.216	0.154
	p=3				p=4			
ρ=0	0.005	0.005	0.004	0.002	0.006	0.005	0.004	0.002
ρ=0.25	0.008	0.010	0.011	0.008	0.009	0.010	0.011	0.008
ρ=0.50	0.046	0.048	0.046	0.029	0.052	0.047	0.046	0.030
ρ=0.75	0.422	0.284	0.216	0.153	0.474	0.285	0.216	0.154
Parzen								
	p=1				p=2			
ρ=0	0.042	0.016	0.028	0.006	0.037	0.014	0.028	0.006
ρ=0.25	0.044	0.019	0.043	0.011	0.042	0.019	0.043	0.011
ρ=0.50	0.070	0.032	0.075	0.018	0.082	0.030	0.075	0.018
ρ=0.75	0.151	0.057	0.138	0.037	0.177	0.053	0.138	0.038
	p=3				p=4			
ρ=0	0.034	0.014	0.028	0.006	0.033	0.014	0.028	0.006
ρ=0.25	0.043	0.019	0.043	0.011	0.043	0.019	0.043	0.011
ρ=0.50	0.090	0.028	0.075	0.018	0.095	0.027	0.075	0.018
ρ=0.75	0.193	0.050	0.138	0.037	0.204	0.049	0.138	0.038
QS								
	p=1				p=2			
ρ=0	0.021	0.008	0.014	0.003	0.018	0.007	0.014	0.003
ρ=0.25	0.022	0.009	0.021	0.006	0.021	0.009	0.021	0.006
ρ=0.50	0.034	0.015	0.037	0.009	0.040	0.015	0.037	0.009
ρ=0.75	0.073	0.028	0.069	0.018	0.086	0.027	0.069	0.019
	p=3				p=4			
ρ=0	0.017	0.007	0.014	0.003	0.016	0.007	0.014	0.003
ρ=0.25	0.021	0.009	0.021	0.006	0.021	0.009	0.021	0.006
ρ=0.50	0.044	0.014	0.037	0.009	0.046	0.013	0.037	0.009
ρ=0.75	0.094	0.024	0.069	0.018	0.099	0.024	0.069	0.019

Note: $\bar{b}_{OPT}$ and $std\,(b_{OPT})$ are the average and standard deviation of testing-optimal b's across 10000 simulation replications. $\bar{b}_{MSE}$ and $std\,(b_{MSE})$ are defined similarly. Parameter configuration is the same as in Table 1.

Table 3: Empirical Type I error of different tests for MA(1) regressors and error with $T = 100, \tau = 1.15$

	b-OPT	**b-MAX**	**b-MSE**	**b-MIX**	**b-OPT**	**b-MAX**	**b-MSE**	**b-MIX**
Bartlett								
	p=1				p=2			
ρ=0	0.064	0.058	0.066	0.064	0.073	0.062	0.075	0.073
ρ=0.25	0.073	0.063	0.076	0.072	0.091	0.067	0.096	0.088
ρ=0.50	0.092	0.068	0.094	0.080	0.114	0.075	0.131	0.101
ρ=0.75	0.092	0.069	0.100	0.080	0.110	0.080	0.145	0.101
	p=3				p=4			
ρ=0	0.085	0.069	0.090	0.086	0.095	0.069	0.103	0.099
ρ=0.25	0.108	0.073	0.116	0.105	0.123	0.080	0.138	0.121
ρ=0.50	0.127	0.087	0.163	0.117	0.142	0.092	0.200	0.134
ρ=0.75	0.118	0.092	0.181	0.114	0.132	0.099	0.230	0.133
Parzen								
	p=1				p=2			
ρ=0	0.061	0.059	0.068	0.061	0.068	0.053	0.081	0.066
ρ=0.25	0.064	0.057	0.075	0.063	0.076	0.055	0.101	0.076
ρ=0.50	0.069	0.060	0.091	0.069	0.081	0.059	0.128	0.085
ρ=0.75	0.071	0.063	0.095	0.072	0.081	0.060	0.138	0.086
	p=3				p=4			
ρ=0	0.074	0.055	0.098	0.074	0.079	0.058	0.112	0.080
ρ=0.25	0.085	0.058	0.128	0.085	0.094	0.064	0.151	0.093
ρ=0.50	0.090	0.065	0.164	0.091	0.098	0.069	0.205	0.102
ρ=0.75	0.089	0.068	0.177	0.092	0.097	0.070	0.228	0.106
QS								
	p=1				p=2			
ρ=0	0.062	0.060	0.067	0.060	0.068	0.052	0.080	0.065
ρ=0.25	0.065	0.060	0.074	0.064	0.077	0.052	0.098	0.076
ρ=0.50	0.070	0.062	0.089	0.070	0.083	0.055	0.124	0.086
ρ=0.75	0.072	0.064	0.092	0.073	0.084	0.057	0.132	0.087
	p=3				p=4			
ρ=0	0.076	0.059	0.095	0.074	0.080	0.053	0.109	0.080
ρ=0.25	0.087	0.058	0.123	0.085	0.095	0.056	0.146	0.093
ρ=0.50	0.091	0.064	0.158	0.093	0.101	0.057	0.199	0.105
ρ=0.75	0.091	0.063	0.172	0.095	0.102	0.059	0.218	0.108

Note: b-OPT: the asymptotic F test with testing-optimal b implemented by a VAR(1) plug-in procedure; b-MSE: the standard Wald test with MSE-optimal b implemented by a VAR(1) plug-in procedure; b-MAX: the KV test with b=1; b-MIX: the asymptotic F test with MSE-optimal b.

9 Appendix of Proofs

9.1 Additional Technical Results

Lemma 2 *Let Assumption 5(i) hold. As $b \to 0$, we have*
(a) $\mu_1 = \sum_{n=1}^{\infty} \lambda_n^* = 1 - bc_1 + O(b^2)$,
(b) $\mu_2 = \sum_{n=1}^{\infty} (\lambda_n^*)^2 = bc_2 + O(b^2)$.

Proof of Lemma 2. Note that

$$\mu_1 = \sum_{n=1}^{\infty} \lambda_n^* = \int_0^1 k_b^*(r, r)dr = 1 - \int_0^1 \int_0^1 k_b(r - s)drds$$

and

$$\begin{aligned}
\mu_2 &= \sum_{m=1}^{\infty} (\lambda_m^*)^2 = \int_0^1 \int_0^1 [k_b^*(r, s)]^2 \, drds \\
&= \left(\int_0^1 \int_0^1 k_b(r - s)drds \right)^2 + \int_0^1 \int_0^1 k_b^2(r - s)drds \\
&- 2 \int_0^1 \int_0^1 \int_0^1 k_b(r - p)k_b(r - q)drdpdq.
\end{aligned}$$

To evaluate μ_1 and μ_2, we let

$$\mathcal{K}_1 (\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} k(x) \exp(-i\lambda x)dx, \; \mathcal{K}_2 (\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} k^2(x) \exp(-i\lambda x)dx. \tag{20}$$

Then

$$k(x) = \int_{-\infty}^{\infty} \mathcal{K}_1 (\lambda) \exp(i\lambda x)d\lambda, \; k^2(x) = \int_{-\infty}^{\infty} \mathcal{K}_2 (\lambda) \exp(i\lambda x)d\lambda. \tag{21}$$

For the integral that appears in both μ_1 and μ_2, we have

$$\begin{aligned}
&\int_0^1 \int_0^1 k_b(r - s)drds \\
&= \int_{-\infty}^{\infty} \mathcal{K}_1 (\lambda) \left[\int_0^1 \exp\left(\frac{i\lambda r}{b}\right) dr \right] \left[\int_0^1 \exp\left(-\frac{i\lambda s}{b}\right) ds \right] d\lambda \\
&= \int_{-\infty}^{\infty} \mathcal{K}_1 (\lambda) \frac{b^2}{\lambda^2} \left[\left(1 - \cos\left(\frac{\lambda}{b}\right)\right)^2 + \left(\sin\left(\frac{\lambda}{b}\right)\right)^2 \right] d\lambda \\
&= b \int_{-\infty}^{\infty} \mathcal{K}_1 (\lambda) \, b \left(\frac{\sin \frac{\lambda}{2b}}{\frac{\lambda}{2}} \right)^2 d\lambda \\
&= 2\pi b \mathcal{K}_1(0) + 4b^2 \int_{-\infty}^{\infty} \frac{\mathcal{K}_1 (\lambda) - \mathcal{K}_1 (0)}{\lambda^2} \left(\sin \frac{\lambda}{2b} \right)^2 d\lambda,
\end{aligned} \tag{22}$$

where the last equality holds because

$$\int_{-\infty}^{\infty}\left(\frac{\lambda}{2b}\right)^{-2}\left(\sin\frac{\lambda}{2b}\right)^{2}d\lambda = 2b\int_{-\infty}^{\infty}x^{-2}\sin^{2}x dx = 2\pi b. \tag{23}$$

Now,

$$\begin{aligned}
&\int_{-\infty}^{\infty}\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\left(\sin\frac{\lambda}{2b}\right)^{2}d\lambda \\
&= \int_{-\infty}^{\infty}\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\left(\left(\sin\frac{\lambda}{2b}\right)^{2}-\frac{1}{2}\right)d\lambda + \frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}d\lambda \\
&= -\frac{1}{2}\int_{-\infty}^{\infty}\left(\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\right)\left(\cos\frac{1}{b}\lambda\right)d\lambda + \frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}d\lambda \\
&= \frac{1}{2}\int_{-\infty}^{\infty}\left(\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\right)d\lambda + o(1)
\end{aligned} \tag{24}$$

as $b \to 0$, where we have used the Riemann-Lebesgue lemma. In view of the symmetry of $k(x)$, $\mathcal{K}_1(\lambda) = (2\pi)^{-1}\int_{-\infty}^{\infty}k(x)\cos(\lambda x)dx$, we have, using (22) and (24):

$$\begin{aligned}
&\int_0^1\int_0^1 k_b(r-s)drds \\
&= 2\pi b\mathcal{K}_1(0) + 2b^2\int_{-\infty}^{\infty}\left(\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\right)d\lambda + o(b^2) \\
&= 2\pi b\mathcal{K}_1(0) + b^2\frac{1}{\pi}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}k(x)\frac{\cos\lambda x-1}{\lambda^2}dx\right)d\lambda + o(b^2) \\
&= 2\pi b\mathcal{K}_1(0) - 2b^2\frac{1}{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}k(x)\frac{\sin^2(\lambda x/2)}{\lambda^2}dxd\lambda + o(b^2) \\
&= 2\pi b\mathcal{K}_1(0) - b^2\int_{-\infty}^{\infty}k(x)\left|x\right|dx + o\left(b^2\right) \\
&= bc_1 + O(b^2).
\end{aligned} \tag{25}$$

Similarly, under the assumption that $\int_{-\infty}^{\infty}k^2(x)x^2dx < \infty$, we have

$$\int_0^1\int_0^1 k_b^2(r-s)drds = bc_2 + O(b^2). \tag{26}$$

Next,

$$\begin{aligned}
&\int_0^1 k_b(r-s)ds \\
&= \frac{1}{2}\int_{-\infty}^{\infty}\mathcal{K}_1(\lambda)\int_0^1\left\{\exp\left(\frac{i\lambda(r-s)}{b}\right)+\exp\left(-\frac{i\lambda(r-s)}{b}\right)\right\}dsd\lambda \\
&= \int_{-\infty}^{\infty}\mathcal{K}_1(\lambda)\int_0^1\cos\left(\frac{\lambda(r-s)}{b}\right)dsd\lambda \qquad (27) \\
&= -b\int_{-\infty}^{\infty}\mathcal{K}_1(\lambda)\frac{1}{\lambda}\left[\sin\left(\frac{\lambda(r-1)}{b}\right)-\sin\left(\frac{\lambda r}{b}\right)\right]d\lambda \\
&= -b\int_{-\infty}^{\infty}\mathcal{K}_1(xb)\frac{1}{x}\left[\sin\left(x(r-1)\right)-\sin\left(xr\right)\right]dx,
\end{aligned}$$

so

$$
\begin{aligned}
&\int_0^1\int_0^1\int_0^1 k_b(r-p)k_b(r-q)drdpdq\\
&=b^2\int_0^1\left[\int_{-\infty}^{\infty}\mathcal{K}_1(xb)\frac{1}{x}\left[\sin(x(r-1))-\sin(xr)\right]dx\right]^2dr\\
&=b^2\mathcal{K}_1^2(0)\int_0^1\left(\int_{-\infty}^{\infty}\frac{1}{x}\sin(x(r-1))dx-\int_{-\infty}^{\infty}\frac{1}{x}\sin(xr)dx\right)^2dr\,(1+o(1))\\
&=b^2\mathcal{K}_1^2(0)\int_0^1\left(-\int_{-\infty}^{\infty}\frac{\sin(x(r-1))}{x(r-1)}d(x(r-1))-\int_{-\infty}^{\infty}\frac{1}{xr}\sin(xr)d(xr)\right)^2dr\,(1+o(1))\\
&=b^2\mathcal{K}_1^2(0)\int_0^1\left(2\int_{-\infty}^{\infty}\frac{1}{y}\sin(y)dy\right)^2dr\,(1+o(1))=c_1^2b^2+o(b^2).
\end{aligned}\tag{28}
$$

Combining (25), (26), and (28) yields the lemma. ■

Lemma 3 *Let Assumption 5(i) hold. As $b\to0$, we have*

(a) $E\left(\nu_{11}-\nu_{12}\nu_{22}^{-1}\nu_{21}\right)=1-bc_1-bc_2(p-1)+o(b)$,

(b) $E\left[\left(\nu_{11}-\nu_{12}\nu_{22}^{-1}\nu_{21}\right)^2\right]=1-2b(c_1-c_2)-2(p-1)bc_2+o(b)$,

(c) $E\left[\left(\nu_{11}-\nu_{12}\nu_{22}^{-1}\nu_{21}\right)-1\right]^2=2bc_2+o(b)$.

Proof of Lemma 3. (a) Let $W_p(r)=\left[W_1'(r),W_{p-1}'(r)\right]'$, then

$$E\nu_{11}=E\int_0^1\int_0^1 k_b^*(r,s)dW_1(r)dW_1'(s)=\sum_{n=1}^{\infty}\lambda_n^*=1-bc_1+o(b)$$

by Lemma 2, and

$$
\begin{aligned}
&E\left[\nu_{12}\nu_{22}^{-1}\nu_{21}\right]\\
&=E\left(\int_0^1\int_0^1 k_b^*(r,s)dW_1(r)dW_{p-1}'(s)\right)\left(\int_0^1\int_0^1 k_b^*(r,s)dW_{p-1}(r)dW_{p-1}'(s)\right)^{-1}\\
&\times\int_0^1\int_0^1 k_b^*(r,s)dW_1'(r)dW_{p-1}(s)\\
&=Etr\left(\int_0^1\int_0^1 k_b^*(r,s)dW_{p-1}(r)dW_{p-1}'(s)\right)^{-1}\\
&\times\int_0^1\int_0^1\left(\int_0^1 k_b^*(r,\tau_1)k_b^*(r,\tau_2)dr\right)dW_{p-1}(\tau_2)dW_{p-1}'(\tau_1).
\end{aligned}
$$

Let $\xi_n=\int_0^1 f_n^*(r)dW_{p-1}(r)\in\mathbb{R}^{p-1}$, then

$$
\begin{aligned}
&\int_0^1\int_0^1 k_b^*(r,s)dW_{p-1}(r)dW_{p-1}'(s)\\
&=\int_0^1\int_0^1\sum_{n=1}^{\infty}\lambda_n^*f_n^*(r)f_n^*(s)dW_{p-1}(r)dW_{p-1}'(s)\\
&=\sum_{n=1}^{\infty}\lambda_n^*\left(\int_0^1 f_n^*(r)dW_{p-1}(r)\right)\left(\int_0^1 f_n^*(r)dW_{p-1}(r)\right)'=\sum_{n=1}^{\infty}\lambda_n^*\xi_n\xi_n'.
\end{aligned}
$$

Since

$$\begin{aligned}
&\int_0^1 k_b^*(r,\tau_1)k_b^*(r,\tau_2)\,dr \\
&= \int_0^1 \sum_{m=1}^{\infty} \lambda_m^* f_m^*(r) f_m^*(\tau_1) \sum_{n=1}^{\infty} \lambda_n^* f_n^*(r) f_n^*(\tau_2)\,dr \\
&= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \lambda_m^*\lambda_n^* \left(\int_0^1 f_m^*(r) f_n^*(r) dr\right) f_m^*(\tau_1) f_n^*(\tau_2) \\
&= \sum_{n=1}^{\infty} (\lambda_n^*)^2 f_n^*(\tau_1) f_n^*(\tau_2),
\end{aligned}$$

we have

$$\int_0^1\int_0^1 \left(\int_0^1 k_b^*(r,\tau_1)k_b^*(r,\tau_2)\,dr\right) dW_{p-1}(\tau_2)dW_{p-1}'(\tau_1) = \sum_{m=1}^{\infty} (\lambda_m^*)^2 \xi_m\xi_m'.$$

Therefore

$$\begin{aligned}
E\nu_{12}\nu_{22}^{-1}\nu_{21} &= Etr\left[\left(\sum_{n=1}^{\infty}\lambda_n^*\xi_n\xi_n'\right)^{-1}\left(\sum_{m=1}^{\infty}(\lambda_m^*)^2\xi_m\xi_m'\right)\right] \\
&= \frac{\mu_2}{\mu_1}(p-1)(1+o(1)) = \frac{bc_2+o(b)}{1-bc_1+o(b)}(p-1)(1+o(1)) \\
&= bc_2(p-1)+o(b),
\end{aligned}$$

using Lemma 2.

(b) Note that

$$E\left(\nu_{11}-\nu_{12}\nu_{22}^{-1}\nu_{21}\right)^2 = E\nu_{11}^2 + E\nu_{12}\nu_{22}^{-1}\nu_{21}\nu_{12}\nu_{22}^{-1}\nu_{21} - 2E\nu_{11}\nu_{12}\nu_{22}^{-1}\nu_{21}.$$

We consider each term in turn. First,

$$\begin{aligned}
E\nu_{11}^2 &= E\left(\int_0^1\int_0^1 k_b^*(r_1,s_1)dW_1(r_1)dW_1'(s_1)\right)\left(\int_0^1\int_0^1 k_b^*(r_2,s_2)dW_1(r_2)dW_1'(s_2)\right) \\
&= \left(\int_0^1 k_b^*(r,r)dr\right)^2 + 2\int_0^1\int_0^1 (k_b^*(r,s))^2 drds \\
&= \left(\sum_{n=1}^{\infty}\lambda_n^*\right)^2 + 2\sum_{n=1}^{\infty}(\lambda_n^*)^2 = (1-bc_1+o(b))^2 + 2bc_2 + o(b), \\
&= 1-2b(c_1-c_2)+o(b).
\end{aligned}$$

Second,

$$
\begin{aligned}
&E\nu_{11}\nu_{12}\nu_{22}^{-1}\nu_{21} \\
&= E\int_0^1\int_0^1 k_b^*(r,s)dW_1(r)dW_1'(s)\int_0^1\int_0^1 k_b^*(r,\tau_1)dW_1(r)dW_{p-1}'(\tau_1) \\
&\times\left[\int_0^1\int_0^1 k_b^*(r,s)dW_{p-1}(r)dW_{p-1}'(s)\right]^{-1}\int_0^1\int_0^1 k_b^*(r,\tau_2)dW_1'(r)dW_{p-1}(\tau_2) \\
&= E\left(\int_0^1 k_b^*(r,r)dr\right)\int_0^1\int_0^1\left(\int_0^1 k_b^*(r,\tau_1)k_b^*(r,\tau_2)dr\right)dW_{p-1}'(\tau_1) \\
&\times\left[\int_0^1\int_0^1 k_b^*(r,s)dW_{p-1}(r)dW_{p-1}'(s)\right]^{-1}dW_{p-1}(\tau_2) \\
&+2E\int_0^1\int_0^1\left[\int_0^1\int_0^1 k_b^*(r,s)k_b^*(r,\tau_1)k_b^*(s,\tau_2)drds\right]dW_{p-1}'(\tau_1) \\
&\times\left[\int_0^1\int_0^1 k_b^*(r,s)dW_{p-1}(r)dW_{p-1}'(s)\right]^{-1}dW_{p-1}(\tau_2) \\
&= \left(\sum_{n=1}^{\infty}\lambda_n^*\right)Etr\left(\sum_{k=1}^{\infty}\lambda_n^*\xi_n\xi_n'\right)^{-1}\left(\sum_{n=1}^{\infty}\left(\lambda_n^*\right)^2\xi_n\xi_n'\right) \\
&+2Etr\left(\sum_{n=1}^{\infty}\lambda_n^*\xi_n\xi_n'\right)^{-1}\left(\sum_{n=1}^{\infty}\left(\lambda_n^*\right)^3\xi_n\xi_n'\right)
\end{aligned}
$$

where the last line follows because

$$
\begin{aligned}
&\int_0^1\int_0^1 k_b^*(r,s)k_b^*(r,\tau_1)k_b^*(s,\tau_2)drds \\
&=\int_0^1\int_0^1\sum_{k_1=1}^{\infty}\lambda_{k_1}^*f_{k_1}^*(r)f_{k_1}^*(s)\sum_{k_2=1}^{\infty}\lambda_{k_2}^*f_{k_2}^*(r)f_{k_2}^*(\tau_1)\sum_{k_3=1}^{\infty}\lambda_{k_3}^*f_{k_3}^*(s)f_{k_3}^*(\tau_2)\,drds \\
&=\sum_{k=1}^{\infty}\left(\lambda_k^*\right)^3 f_k^*(\tau_1)\,f_k^*(\tau_2)\,.
\end{aligned}
$$

Using Lemma 2 and the fact that

$$
\sum_{n=1}^{\infty}\left(\lambda_n^*\right)^3 = o\left(\sum_{n=1}^{\infty}\left(\lambda_n^*\right)^2\right) = o(b),
$$

we have

$$
E\nu_{11}\nu_{12}\nu_{22}^{-1}\nu_{21} = (p-1)\,bc_2 + o(b).
$$

Finally,

$$
\begin{aligned}
&E\nu_{12}\nu_{22}^{-1}\nu_{21}\nu_{12}\nu_{22}^{-1}\nu_{21}\\
&= E\left[\int_0^1\int_0^1 k_b^*(r,s)dW_1(r)dW_{p-1}'(s)\right]\left[\int_0^1\int_0^1 k_b^*(r,s)dW_{p-1}(r)dW_{p-1}'(s)\right]^{-1}\\
&\times\left[\int_0^1\int_0^1 k_b^*(r,s)dW_1'(r)dW_{p-1}(s)\right]\left[\int_0^1\int_0^1 k_b^*(r,s)dW_1(r)dW_{p-1}'(s)\right]\\
&\times\left[\int_0^1\int_0^1 k_b^*(r,s)dW_{p-1}(r)dW_{p-1}'(s)\right]^{-1}\left[\int_0^1\int_0^1 k_b^*(r,s)dW_1'(r)dW_{p-1}(s)\right]\\
&= E\int_0^1\int_0^1\left[\int_0^1 k_b^*(r,s_1)k_b^*(r,s_2)dr\right]dW_{p-1}'(s_1)\left[\int_0^1\int_0^1 k_b^*(r,s)dW_{p-1}(r)dW_{p-1}'(s)\right]^{-1}dW_{p-1}(s_2)\\
&\times\left[\int_0^1\int_0^1\left[\int_0^1 k_b^*(\tau,s_3)k_b^*(\tau,s_4)d\tau\right]dW_{p-1}'(s_3)\right]\left[\int_0^1\int_0^1 k_b^*(r,s)dW_{p-1}(r)dW_{p-1}'(s)\right]^{-1}dW_{p-1}(s_4)\\
&+2E\left[\int_0^1\int_0^1\left[\int_0^1 k_b^*(r,s_1)k_b^*(r,s_2)dr\right]dW_{p-1}'(s_1)\right]\left[\int_0^1\int_0^1 k_b^*(r,s)dW_{p-1}(r)dW_{p-1}'(s)\right]^{-1}\\
&\times\int_0^1\int_0^1\left[\int_0^1 k_b^*(r,s_3)k_b^*(r,s_4)dr\right]dW_{p-1}(s_3)dW_{p-1}'(s_4)\\
&\times\left[\int_0^1\int_0^1 k_b^*(r,s)dW_{p-1}(r)dW_{p-1}'(s)\right]^{-1}dW_{p-1}(s_2)\\
&= Etr\left(\left[\sum_{n=1}^\infty\lambda_n^*\xi_n\xi_n'\right]^{-1}\left[\sum_{m=1}^\infty\left(\lambda_m^*\right)^2\xi_m\xi_m'\right]\right)tr\left(\left[\sum_{n=1}^\infty\lambda_n^*\xi_n\xi_n'\right]^{-1}\left[\sum_{m=1}^\infty\left(\lambda_m^*\right)^2\xi_m\xi_m'\right]\right)\\
&+2Etr\left\{\left[\sum_{n=1}^\infty\lambda_n^*\xi_n\xi_n'\right]^{-1}\left[\sum_{m=1}^\infty\left(\lambda_m^*\right)^2\xi_m\xi_m'\right]\left[\sum_{n=1}^\infty\lambda_n^*\xi_n\xi_n'\right]^{-1}\left[\sum_{m=1}^\infty\left(\lambda_m^*\right)^2\xi_m\xi_m'\right]\right\}\\
&=\left[(p-1)^2\left(\sum_{n=1}^\infty\lambda_n^*\right)^{-2}\left(\sum_{m=1}^\infty\left(\lambda_m^*\right)^2\right)^2+2\left(p-1\right)\left(\sum_{n=1}^\infty\lambda_n^*\right)^{-2}\left(\sum_{m=1}^\infty\left(\lambda_m^*\right)^2\right)^2\right](1+o(1))\\
&= o(b)
\end{aligned}
$$

using Lemma 2.

Hence

$$E\left(\nu_{11}-\nu_{12}\nu_{22}^{-1}\nu_{21}\right)^2 = 1-2b\left(c_1-c_2\right)-2\left(p-1\right)bc_2+o(b)$$

Part (c) follows from parts (a) and (b). ■

9.2 Proof of the Main Results

Proof of Theorem 1. Taking a Taylor expansion, we have

$$
\begin{aligned}
&P\left\{pF_\infty(p,b)\le z\right\} = EG_p\left(z\left(\nu_{11}-\nu_{12}\nu_{22}^{-1}\nu_{21}\right)\right)\\
&= G_p(z) + G_p'(z)zE\left[\left(\nu_{11}-\nu_{12}\nu_{22}^{-1}\nu_{21}\right)-1\right]\\
&+\frac{1}{2}G_p''(z)z^2E\left[\left(\nu_{11}-\nu_{12}\nu_{22}^{-1}\nu_{21}\right)-1\right]^2\\
&+\frac{1}{2}E\left[G_p''(\tilde z)-G_p''(z)\right]z^2\left[\left(\nu_{11}-\nu_{12}\nu_{22}^{-1}\nu_{21}\right)-1\right]^2
\end{aligned}
$$

where $\tilde z$ is between z and $z\left(\nu_{11}-\nu_{12}\nu_{22}^{-1}\nu_{21}\right)$. Using Lemma 3, we have

$$
\begin{aligned}
&P\left\{pF_\infty(p,b)\le z\right\}\\
&= G_p(z) - G_p'(z)z\left[c_1+c_2\left(p-1\right)\right]b\\
&+\frac{1}{2}G_p''(z)z^2\left[2-2b\left(c_1-c_2\right)-2\left(p-1\right)bc_2-2\left(1-bc_1-bc_2\left(p-1\right)\right)\right]+o(b)\\
&= G_p(z)+\left\{G_p''(z)z^2c_2-G_p'(z)z\left[c_1+c_2\left(p-1\right)\right]\right\}b+o(b)\\
&= G_p(z)+A(z)b+o(b)
\end{aligned}
$$

as stated. ■

Proof of Lemma 1. Part (a). We write the statistic $pF_{T,GLS}$ as

$$
\begin{aligned}
pF_{T,GLS} &= \left\{\left[R_0T^{1/2}(\hat\theta_{GLS}-\theta_0)\right]'\Omega_{T,GLS}^{-1/2}\right\}\Omega_{T,GLS}^{1/2}\hat\Omega_T^{-1}\Omega_{T,GLS}^{1/2}\left\{\Omega_{T,GLS}^{-1/2}\left[R_0T^{1/2}(\hat\theta_{GLS}-\theta_0)\right]\right\}\\
&= \left\|\Omega_{T,GLS}^{-1/2}\left[R_0T^{1/2}(\hat\theta_{GLS}-\theta_0)\right]\right\|^2\times e_T'\Omega_{T,GLS}^{1/2}\hat\Omega_T^{-1}\Omega_{T,GLS}^{1/2}e_T\\
&\equiv \Upsilon_T\Xi_T^*
\end{aligned}
$$

where

$$
\Upsilon_T = \left\|\Omega_{T,GLS}^{-1/2}\left[R_0T^{1/2}(\hat\theta_{GLS}-\theta_0)\right]\right\|^2,
$$

and

$$
\Xi_T^* = e_T'\Omega_{T,GLS}^{1/2}\hat\Omega_T^{-1}\Omega_{T,GLS}^{1/2}e_T.
$$

Note that Υ_T is independent of Ξ_T because (i) $(\hat\theta_{GLS}-\theta_0)$ is independent of $\hat\Omega_T$, which follows from the facts that $\hat\Omega_T$ is a function of $\{\hat u_t\}$ and that $\hat u_t$ is independent of $\hat\theta_{GLS}-\theta$. (ii) Υ_T is the squared length of a standard normal vector and e_T is the direction of this vector. The length is independent of the direction. Hence

$$
\begin{aligned}
P\left[pF_{T,GLS}\le z\right] &= P\left[\Upsilon_T\Xi_T^*\le z\right] = EG_p\left[z\left(\Xi_T^*\right)^{-1}\right]\\
&= EG_p\left(e_T'\Omega_{T,GLS}^{1/2}\hat\Omega_T^{-1}\Omega_{T,GLS}^{1/2}e_T\right).
\end{aligned}
$$

It is not hard to show that $e_T'\Omega_{T,GLS}^{1/2}\hat\Omega_T^{-1}\Omega_{T,GLS}^{1/2}e_T = \left\|e_T'\Omega_{T,GLS}^{1/2}\hat\Omega_T^{-1/2}\right\|^2 \ge c > 0$ for some c with probability one. Let $F_{T,\Omega}(V)$ be the CDF of $\hat\Omega_T$. Using $\Omega_{T,GLS} =$

$\Omega\left(1+O\left(T^{-1}\right)\right)$ and the boundedness of $G_p'(x)$ for $x \geq \delta > 0$, we have

$$
\begin{aligned}
P\left[pF_{T,GLS} \leq z\right] &= \int G_p\left(e_T' \Omega_{T,GLS}^{1/2} V^{-1} \Omega_{T,GLS}^{1/2} e_T\right) dF_{T,\Omega}(V) \\
&= \int \left[G_p\left(e_T' \Omega^{1/2} V^{-1} \Omega^{1/2} e_T\right) + O\left(T^{-1}\right)\right] dF_{T,\Omega}(V) \\
&= \int \left[G_p\left(e_T' \Omega^{1/2} V^{-1} \Omega^{1/2} e_T\right)\right] dF_{T,\Omega}(V) + O\left(T^{-1}\right) \\
&= EG_p\left[z \Xi_T^{-1}\right] + O\left(T^{-1}\right)
\end{aligned}
$$

as stated.

Part (b). Let

$$
\begin{aligned}
\zeta_{1T} &= 2(R_0 T^{1/2}\Delta)' \hat{\Omega}_T^{-1} \Omega_{T,GLS}^{1/2} e_T \\
\zeta_{2T} &= (R_0 T^{1/2}\Delta)' \hat{\Omega}_T^{-1} \left(R_0 T^{1/2}\Delta\right)
\end{aligned}
$$

and $\zeta_T = \sqrt{\Upsilon_T}\zeta_{1T} + \zeta_{2T}$. Then

$$
\begin{aligned}
pF_{T,OLS} &= \left[R_0 T^{1/2}(\hat{\theta}_{GLS} - \theta_0) + R_0 T^{1/2}\Delta\right]' \hat{\Omega}_T^{-1} \left[R_0 T^{1/2}(\hat{\theta}_{GLS} - \theta_0) + R_0 T^{1/2}\Delta\right] \\
&= \left[R_0 T^{1/2}(\hat{\theta}_{GLS} - \theta_0)\right]' \hat{\Omega}_T^{-1} \left[R_0 T^{1/2}(\hat{\theta}_{GLS} - \theta_0)\right] \\
&+ \left[R_0 T^{1/2}\Delta\right]' \hat{\Omega}_T^{-1} \left[R_0 T^{1/2}\Delta\right] + 2(R_0 T^{1/2}\Delta)' \hat{\Omega}_T^{-1} \left[R_0 T^{1/2}(\hat{\theta}_{GLS} - \theta_0)\right] \\
&= pF_{T,GLS} + \zeta_{2T} + 2(R_0 T^{1/2}\Delta)' \hat{\Omega}_T^{-1} \Omega_{T,GLS}^{1/2} \frac{\Omega_{T,GLS}^{-1/2} R_0 T^{1/2}(\hat{\theta}_{GLS} - \theta_0)}{\left\| \Omega_{T,GLS}^{-1/2} R_0 T^{1/2}(\hat{\theta}_{GLS} - \theta_0)\right\|} \sqrt{\Upsilon_T} \\
&= pF_{T,GLS} + \zeta_{2T} + \sqrt{\Upsilon_T}\zeta_{1T} \\
&= pF_{T,GLS} + \zeta_T.
\end{aligned}
$$

Note that Υ_T is independent of ζ_{1T}, ζ_{2T} and Ξ_T, we have

$$
\begin{aligned}
P\left[pF_{T,OLS} \leq z\right] &= P\left[(pF_{T,GLS} + \zeta_T) \leq z\right] \\
&= P\left\{\left[\Upsilon_T \Xi_T^* + \sqrt{\Upsilon_T}\zeta_{1T} + \zeta_{2T}\right] \leq z\right\} \\
&= P\left\{\left[\Upsilon_T \Xi_T + \sqrt{\Upsilon_T}\zeta_{1T} + \zeta_{2T}\right] \leq z\right\} + O\left(T^{-1}\right) \\
&\equiv EF\left(\zeta_{1T}, \zeta_{2T}, \Xi_T\right) + O\left(T^{-1}\right),
\end{aligned}
$$

where

$$
F(a,b,c) = P\left\{\left[\Upsilon_T c + \sqrt{\Upsilon_T} a + b\right] \leq z\right\}.
$$

But

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

where $F_1'(a,b,c) = \partial F(a,b,c)/\partial a$. 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 $var(c'\Delta\Delta'c) = O\left(1/T\right)$ for any constant c. Next, let $f_e(x)$ be the pdf of e_T. Since e_T is independent of $\hat{\Omega}_T$ and Δ, we have

$$\begin{aligned}
& EF_1'\left(0,0,\Xi_T\right)\zeta_{1T} \\
&= \int E\left[F_1'\left(0,0,\Xi_T\right)\zeta_{1T}\left|e_T = x\right.\right] f_e(x)dx \\
&= \int EF_1'\left(0,0,x'\Omega^{1/2}\hat{\Omega}_T^{-1}\Omega^{1/2}x'\right) 2\left(R_0T^{1/2}\Delta\right)'\hat{\Omega}_T^{-1}\Omega_{T,GLS}^{1/2} x f_e(x)dx
\end{aligned}$$

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

$$EF_1'\left(0,0,x'\Omega^{1/2}\hat{\Omega}_T^{-1}\Omega^{1/2}x'\right) 2\left(R_0T^{1/2}\Delta\right)'\hat{\Omega}_T^{-1}\Omega_{T,GLS}^{1/2} x = 0 \text{ for all } x.$$

As a result,

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

So

$$EF\left(\zeta_{1T},\zeta_{2T},\Xi_T\right) = EF\left(0,0,\Xi_T\right) + O\left(T^{-1}\right).$$

We have therefore shown that

$$\begin{aligned}
P\left[pF_{T,OLS} \le z\right] &= EF\left(\zeta_{1T},\zeta_{2T},\Xi_T\right) + O\left(T^{-1}\right) \\
&= EF\left(0,0,\Xi_T\right) + O\left(T^{-1}\right) \\
&= P\left[pF_{T,GLS} \le z\right] + O\left(T^{-1}\right)
\end{aligned}$$

as desired. ■

Proof of Theorem 2. Writing $\Xi_T = \Xi_T(\hat{\Omega}_T)$ and taking a Taylor expansion of $\Xi_T(\hat{\Omega}_T)$ around $\Xi_T(\Omega) = 1$, we have

$$\left[\Xi_T\left(\hat{\Omega}_T\right)\right]^{-1} = 1 + L + Q + remainder \tag{29}$$

where

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

and

$$\begin{aligned}
D &= \left(\left[e_T'\Omega^{-1/2}\right]\otimes\left[e_T'\Omega^{-1/2}\right]\right), \\
J_1 &= \left[2\Omega^{-1/2}\left(e_Te_T'\right)\Omega^{-1/2}\right]\otimes\left[\Omega^{-1/2}\left(e_Te_T'\right)\Omega^{-1/2}\right], \\
J_2 &= -\left[\Omega^{-1/2}e_Te_T'\Omega^{-1/2}\otimes\Omega^{-1}\right]\mathbb{K}_{dd}\left(\mathbb{I}_{d^2} + \mathbb{K}_{dd}\right),
\end{aligned}$$

and *remainder* is the remainder term of the Taylor expansion. It can be shown that the remainder term is of smaller order than Q.

We proceed to compute the moments of L and Q. First, extending Lemma 6 in Velasco and Robinson (2001) to the vector case, we have

$$E\hat{\Omega}_T - \Omega = -bc_1\Omega + (bT)^{-q}B(1 + o(1)) + o(b).$$

So

$$\begin{aligned}
EL &= E\left(\left[e_T'\Omega^{-1/2}\right] \otimes \left[e_T'\Omega^{-1/2}\right]\right) vec\left(\hat{\Omega}_T - \Omega\right) \\
&= Ee_T'\Omega^{-1/2}\left(\hat{\Omega}_T - \Omega\right)\Omega^{-1/2}e_T \\
&= (bT)^{-q}Ee_T'\Omega^{-1/2}B\Omega^{-1/2}e_T(1 + o(1)) - bc_1Ee_T'\Omega^{-1/2}\Omega\Omega^{-1/2}e_T + o(b) \\
&= (bT)^{-q}Etr\left(\Omega^{-1/2}B\Omega^{-1/2}e_Te_T'\right)(1 + o(1)) - bc_1 + o(b) \\
&= (bT)^{-q}tr\left[\Omega^{-1/2}B\Omega^{-1/2}\right]\frac{1}{p}(1 + o(1)) - bc_1 + o(b) \\
&= (bT)^{-q}\bar{B}(1 + o(1)) - bc_1 + o(b)
\end{aligned}$$

where we have used the independence of e_T from $\hat{\Omega}_T$ and $Ee_Te_T' = \mathbb{I}_p/p$. Following Sun (2011), we can show that

$$EL^2 = 2c_2b + o\left(b + (bT)^{-q}\right),$$

and

$$EQ = -bc_2(p-1) + o\left(b + (bT)^{-q}\right).$$

Hence

$$\left[\Xi_T\left(\hat{\Omega}_T\right)\right]^{-1} = 1 + L + Q + o_p\left(b + (bT)^{-q}\right). \tag{30}$$

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

Note that

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

and

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

It is clear that there exists a constant $C > 0$ such that $\left|zG_p'(z)\right| \leq C$ and $\left|z^2G_p''(z)\right| \leq C$ for all $z \in (0, \infty)$. Using the asymptotic expansion in (30) and the boundedness of $G_p'(z)z$ and $G_p''(z)z^2$, we have

$$\begin{aligned}
P(pF_{T,OLS} \leq z) &= P\left(\Upsilon_T \leq z\Xi_T^{-1}\right) + O(T^{-1}) \\
&= EG_p(z(1 + L + Q)) + o\left(b + (bT)^{-q}\right) \\
&= G_p(z) + G_p'(z)zE(L + Q) + \frac{1}{2}EG_p''(z)z^2\left(EL^2\right) + o\left(b + (bT)^{-q}\right) \\
&= G_p(z) + (bT)^{-q}G_p'(z)z\bar{B} - bc_1G_p'(z)z \\
&\quad - bc_2G_p'(z)z(p-1) + bc_2G_p''(z)z^2 + o(b) + o\left((bT)^{-q}\right) \\
&= G_p(z) + A(z)b + (bT)^{-q}G_p'(z)z\bar{B} + o(b) + o\left((bT)^{-q}\right)
\end{aligned}$$

as desired. ■

Proof of Theorem 3. For notational economy, let $H_T(\theta) = -R(\theta)\left(G_T'(\theta)\mathcal{W}_T G_T(\theta)\right)^{-1} G_T'(\theta)\mathcal{W}_T$ and $H_0 = -R_0\left[G_0'\mathcal{W}_\infty G_0\right]^{-1} G_0\mathcal{W}_\infty$. For any $\tilde{\theta}_T$ and $\check{\theta}_T$ between $\hat{\theta}_T$ and θ_0, it is easy to show that when $R(\theta)$ and $f(v_t,\theta)$ are twice continuously differentiable in θ on Θ,

$$R(\tilde{\theta}_T) = R_0 + O_p(1/\sqrt{T}),\ G_T\left(\check{\theta}_T\right) = G_0 + O_p(1/\sqrt{T}).$$

Combining this with $\mathcal{W}_T = \mathcal{W}_\infty + O_p(1/\sqrt{T})$, we have

$$H_T(\theta) = H_0 + O_p(1/\sqrt{T})$$

and

$$\begin{aligned}
&\sqrt{T}\left[r(\hat{\theta}_T) - r(\theta_0)\right] \\
&= \sqrt{T}R(\tilde{\theta}_T)\left(\hat{\theta}_T - \theta_0\right) \equiv R_0\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + 1/\sqrt{T}\tilde{\psi}_{0T} \\
&= -R_0\left(G_T'(\hat{\theta}_T)\mathcal{W}_T G_T(\check{\theta}_T)\right)^{-1} G_T'(\hat{\theta}_T)\mathcal{W}_T\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(v_t,\theta_0) + 1/\sqrt{T}\tilde{\psi}_{0T} \\
&= -R_0\left(G_0'\mathcal{W}_\infty G_0\right)^{-1} G_0'\mathcal{W}_\infty\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(v_t,\theta_0) + \frac{1}{\sqrt{T}}\psi_{1T} + \frac{1}{T}\psi_{2T} \\
&= H_0\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(v_t,\theta_0) + \frac{1}{\sqrt{T}}\psi_{1T} + \frac{1}{T}\psi_{2T}
\end{aligned}$$

where $(1/\sqrt{T})\psi_{1T}$ captures the terms of order $O_p(1/\sqrt{T})$ that do not depend on the smoothing parameter b and $(1/T)\,\psi_{2T}$ collects the higher order terms of order $O_p(1/T)$.

Next, define $S_t = \sum_{j=1}^{t} f(v_j,\hat{\theta}_T)$, $S_0 = 0$, then we can show that

$$\begin{aligned}
\hat{\Omega}_T &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} D(t,\tau)\left(H_T(\hat{\theta}_T)S_t\right)\left(H_T(\hat{\theta}_T)S_\tau\right)' \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} D(t,\tau)\left(\left[H_0 + O_p(1/\sqrt{T})\right]S_t\right)\left(\left[H_0 + O_p(1/\sqrt{T})\right]S_\tau\right)' \\
&\equiv \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} D(t,\tau)\left(H_0 S_t\right)\left(H_0 S_\tau\right)' + \frac{1}{\sqrt{T}}\psi_{3T} + \frac{1}{\sqrt{T}}\left[\sqrt{b} + (bT)^{-q}\right]\psi_{4T}
\end{aligned}$$

where the term containing ψ_{3T} captures terms of order $O_p\left(1/\sqrt{T}\right)$ that are independent of b, the term containing ψ_{4T} captures high order terms, and

$$D(t,\tau) = k(\frac{t-\tau}{bT}) - k(\frac{t+1-\tau}{bT}) - k(\frac{t-\tau-1}{bT}) + k(\frac{t-\tau}{bT}).$$

Under the assumption that $G_{[rT]}(\tilde{\theta}_T) = rG_0 + O_p(1/\sqrt{T})$, we have

$$\begin{aligned}
H_0 S_t &= H_0 \sum_{j=1}^{t} f(v_j, \theta_0) + H_0 \sum_{j=1}^{t} \frac{\partial f(v_j, \tilde{\theta}_T)}{\partial \theta'} \left(\hat{\theta}_T - \theta_0 \right) \\
&= H_0 \sum_{j=1}^{t} f(v_j, \theta_0) + t \times H_0 G_0 \left(\hat{\theta}_T - \theta_0 \right) + O_p\left(\frac{1}{\sqrt{T}} \right) \\
&= H_0 \sum_{j=1}^{t} f(v_j, \theta_0) - t \times R_0 \left(G_T'(\hat{\theta}_T) \mathcal{W}_T G_T(\check{\theta}_T) \right)^{-1} G_T'(\hat{\theta}_T) \mathcal{W}_T \frac{1}{T} \sum_{t=1}^{T} f(v_t, \theta_0) + O_p\left(\frac{1}{\sqrt{T}} \right) \\
&= H_0 \sum_{j=1}^{t} f(v_j, \theta_0) - H_0 \frac{t}{T} \sum_{t=1}^{T} f(v_t, \theta_0) + O_p\left(\frac{1}{\sqrt{T}} \right) + O_p\left(\frac{1}{T} \right) \\
&\equiv H_0 \sum_{j=1}^{t} \left[f(v_j, \theta_0) - \frac{1}{T} \sum_{\tau=1}^{T} f(v_\tau, \theta_0) \right] + \frac{1}{\sqrt{T}} \psi_{5T} + \frac{1}{T} \psi_{6T}
\end{aligned}$$

where $\psi_{5T} = O_p(1)$ and $\psi_{6T} = O_p(1)$. Hence

$$\hat{\Omega}_T = \tilde{\Omega}_T + \frac{1}{\sqrt{T}} \psi_{7,T} + \frac{1}{\sqrt{T}} \left[\sqrt{b} + (bT)^{-q} \right] \psi_{8,T}.$$

Combining the above analyses yields:

$$\begin{aligned}
pF_{T,L} = & \left[H_0 \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(v_t, \theta_0) + \frac{1}{\sqrt{T}} \psi_{1T} + \frac{1}{T} \psi_{2T} \right]' \\
& \times \left(\left[\tilde{\Omega}_T + \frac{1}{\sqrt{T}} \psi_{7,T} + \frac{1}{\sqrt{T}} \left[\sqrt{b} + (bT)^{-q} \right] \psi_{8,T} \right] \right)^{-1} \\
& \times \left[H_0 \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(v_t, \theta_0) + \frac{1}{\sqrt{T}} \psi_{1T} + \frac{1}{T} \psi_{2T} \right] \\
& \equiv pF_{T,L} + \psi_T + \psi_T^*
\end{aligned}$$

where $\psi_T = O_p(1/\sqrt{T})$ does not depend on b and $\psi_T^* = O_p([\sqrt{b} + (bT)^{-q}]/\sqrt{T} + 1/T)$.

Note that $pF_{T,L}^d$ is the exactly the same as the Wald statistic for testing whether the mean of process $\phi(v_t, \theta_0)$ satisfies $E\phi(v_t, \theta_0) = 0$. So we can invoke the asymptotic approximation established for the Gaussian location model to complete our proof. Under assumption (ii) of the Theorem, we have

$$\begin{aligned}
P\left(pF_T \le z\right) &= P\left(pF_{T,L} + \psi_T + \psi_T^* \le z\right) \\
&= P\left(pF_{T,L} + \psi_T + \psi_T^* \le z, |\psi_T| < C/\sqrt{T}, |\psi_T^*| < \delta_T / \log T\right) \\
&+ P\left(|\psi_T| \ge \log T/\sqrt{T}\right) + P\left(|\psi_T^*| \ge \delta_T / \log T\right) \\
&= P\left(pF_{T,L} \le z + \frac{\log T}{\sqrt{T}} + \frac{\delta_T}{\log T}\right) + o\left(\delta_T\right) + O\left(\frac{1}{\sqrt{T}}\right) \\
&= G_p\left(z\right) + A(z) b + (bT)^{-q} G_p'\left(z\right) z \bar{B} + O\left(T^{-1/2} \log T\right) + o\left(b\right) + o\left((bT)^{-q}\right)
\end{aligned}$$

where the $O\left(T^{-1/2}\log T\right)$ term does not depend on b. ■

Proof of Theorem 5. We first prove the theorem for the Gaussian location model. It is easy to see that Lemma 1 holds for any given $\tilde{c}$. So

$$\begin{aligned}P(F_{T,OLS}^{*}\leq\mathcal{F}_{p,K}^{\alpha}|H_{1}(\tilde{c})) &= P(F_{T,OLS}\leq\kappa\mathcal{F}_{p,K}^{\alpha}|H_{1}(\tilde{c}))\\ &= P(F_{T,GLS}\leq\kappa\mathcal{F}_{p,K}^{\alpha}|H_{1}(\tilde{c}))+O\left(T^{-1}\right).\end{aligned}$$

But

$$\begin{aligned}pF_{T,GLS} &= \left[R_{0}T^{1/2}(\hat{\theta}_{GLS}-\theta_{0})\right]^{\prime}\Omega_{T,GLS}^{-1/2}\left[\Omega_{T,GLS}^{1/2}\hat{\Omega}_{T}^{-1}\Omega_{T,GLS}^{1/2}\right]\Omega_{T,GLS}^{-1/2}\left[R_{0}T^{1/2}(\hat{\theta}_{GLS}-\theta_{0})\right]\\ &= \left\Vert\Omega_{T,GLS}^{-1/2}\left[R_{0}T^{1/2}(\hat{\theta}_{GLS}-\theta_{0})\right]+\tilde{c}\right\Vert^{2}e_{Tc}^{\prime}\Omega^{1/2}\hat{\Omega}_{T}^{-1}\Omega^{1/2}e_{Tc}+O_{p}\left(T^{-1}\right)\\ &\equiv\Upsilon_{c}\Xi_{c}+O_{p}\left(T^{-1}\right),\end{aligned}$$

where

$$e_{Tc}=\frac{\Omega_{T,GLS}^{-1/2}\left[R_{0}T^{1/2}(\hat{\theta}_{GLS}-\theta_{0})\right]+\tilde{c}}{\left\Vert\Omega_{T,GLS}^{-1/2}\left[R_{0}T^{1/2}(\hat{\theta}_{GLS}-\theta_{0})\right]+\tilde{c}\right\Vert},$$

and

$$\begin{aligned}\Upsilon_{c} &= \left\Vert\Omega_{T,GLS}^{-1/2}\left[R_{0}T^{1/2}(\hat{\theta}_{GLS}-\theta_{0})\right]+\tilde{c}\right\Vert^{2},\\ \Xi_{c} &= e_{c}^{\prime}\Omega^{1/2}\hat{\Omega}_{T}^{-1}\Omega^{1/2}e_{c}.\end{aligned}$$

To compute the average type II error, we can expand the probability space so that $\tilde{c}$, $\hat{\theta}_{GLS}$ and $\hat{\Omega}_{T}$ all live in this expanded space. In addition, $\tilde{c}$ is a random vector uniformly distributed on the sphere $\mathcal{S}_{p}\left(\delta^{2}\right)$, and c is independent of $\hat{\theta}_{GLS}$ and $\hat{\Omega}_{T}$. Hence $\Upsilon_{c}\sim\chi_{p}^{2}\left(\delta^{2}\right)$ and e_{c} is uniformly distributed on the unit sphere $\mathcal{S}_{p}\left(1\right)$. Using the same calculation as in the proof of Theorem 2, we have,

$$\begin{aligned}&P(F_{T,GLS}\leq\kappa\mathcal{F}_{p,K}^{\alpha}|H_{1}(\delta^{2}))=P(pF_{T,GLS}\leq\kappa\mathcal{F}_{p,K}^{\alpha}|H_{1}(\delta^{2}))\\ &=EG_{p,\delta^{2}}\left(p\kappa\mathcal{F}_{p,K}^{\alpha}\Xi_{c}^{-1}\right)+O\left(T^{-1}\right)\\ &=G_{p,\delta^{2}}\left(p\kappa\mathcal{F}_{p,K}^{\alpha}\right)+\left(bT\right)^{-q}G_{p,\delta^{2}}^{\prime}\left(p\kappa\mathcal{F}_{p,K}^{\alpha}\right)p\kappa\mathcal{F}_{p,K}^{\alpha}\bar{B}+A_{\delta^{2}}(p\kappa\mathcal{F}_{p,K}^{\alpha})b+o\left(b\right)+o\left(\left(bT\right)^{-q}\right)\end{aligned}$$

where

$$A_{\delta^{2}}\left(z\right)=G_{p,\delta^{2}}^{\prime\prime}(z)z^{2}c_{2}-G_{p,\delta^{2}}^{\prime}(z)z\left[c_{1}+c_{2}\left(p-1\right)\right].$$

Invoking a Cornish-Fisher type expansion, we can show that

$$p\mathcal{F}_{p,K}^{\alpha}=\mathcal{X}_{p}^{\alpha}-\frac{c_{2}G_{p}^{\prime\prime}\left(\mathcal{X}_{p}^{\alpha}\right)\left(\mathcal{X}_{p}^{\alpha}\right)^{2}}{G_{p}^{\prime}\left(\mathcal{X}_{p}^{\alpha}\right)}b+o\left(b\right).$$

Using this, we have

$$
\begin{aligned}
& G_{p,\delta^2}\left(p\kappa\mathcal{F}_{p,K}^{\alpha}\right) \\
& = G_{p,\delta^2}\left\{\kappa\left(\mathcal{X}_p^{\alpha} - \frac{c_2 G_p''\left(\mathcal{X}_p^{\alpha}\right)\left(\mathcal{X}_p^{\alpha}\right)^2}{G_p'\left(\mathcal{X}_p^{\alpha}\right)} b + o\left(b\right)\right)\right\} \\
& = G_{p,\delta^2}\left\{\kappa\left(\mathcal{X}_p^{\alpha} - \frac{c_2 G_p''\left(\mathcal{X}_p^{\alpha}\right)\left(\mathcal{X}_p^{\alpha}\right)^2}{G_p'\left(\mathcal{X}_p^{\alpha}\right)} b\right)\right\} + o(b) \\
& = G_{p,\delta^2}\left(\mathcal{X}_p^{\alpha}\right) + G_{p,\delta^2}'\left(\mathcal{X}_p^{\alpha}\right)\left\{\mathcal{X}_p^{\alpha}\left[c_1 + c_2\left(p-1\right)\right] - \frac{c_2 G_p''\left(\mathcal{X}_p^{\alpha}\right)\left(\mathcal{X}_p^{\alpha}\right)^2}{G_p'\left(\mathcal{X}_p^{\alpha}\right)}\right\} b + o(b),
\end{aligned}
$$

and so

$$
\begin{aligned}
& G_{p,\delta^2}\left(p\kappa\mathcal{F}_{p,K}^{\alpha}\right) + A_{\delta^2}(p\kappa\mathcal{F}_{p,K}^{\alpha})b \\
& = G_{p,\delta^2}\left(\mathcal{X}_p^{\alpha}\right) + G_{p,\delta^2}'\left(\mathcal{X}_p^{\alpha}\right)\left\{\mathcal{X}_p^{\alpha}\left[c_1 + \left(p-1\right)c_2\right] - \frac{c_2 G_p''\left(\mathcal{X}_p^{\alpha}\right)\left(\mathcal{X}_p^{\alpha}\right)^2}{G_p'\left(\mathcal{X}_p^{\alpha}\right)}\right\} b \\
& + G_{p,\delta^2}''(\mathcal{X}_p^{\alpha})\left(\mathcal{X}_p^{\alpha}\right)^2 c_2 b - G_{p,\delta^2}'(\mathcal{X}_p^{\alpha})\mathcal{X}_p^{\alpha}\left[c_1 + c_2\left(p-1\right)\right] b \\
& = G_{p,\delta^2}\left(\mathcal{X}_p^{\alpha}\right) + \left[G_{p,\delta^2}''(\mathcal{X}_p^{\alpha})\left(\mathcal{X}_p^{\alpha}\right)^2 c_2 - G_{p,\delta^2}'\left(\mathcal{X}_p^{\alpha}\right)\frac{G_p''\left(\mathcal{X}_p^{\alpha}\right)\left(\mathcal{X}_p^{\alpha}\right)^2 c_2}{G_p'\left(\mathcal{X}_p^{\alpha}\right)}\right] b \\
& = G_{p,\delta^2}\left(\mathcal{X}_p^{\alpha}\right) + \left[G_{p,\delta^2}''(\mathcal{X}_p^{\alpha}) - G_{p,\delta^2}'\left(\mathcal{X}_p^{\alpha}\right)\frac{G_p''\left(\mathcal{X}_p^{\alpha}\right)}{G_p'\left(\mathcal{X}_p^{\alpha}\right)}\right]\left(\mathcal{X}_p^{\alpha}\right)^2 c_2 b.
\end{aligned}
$$

Some simple calculation shows that

$$
G_{p,\delta^2}''\left(\mathcal{X}_p^{\alpha}\right) - \frac{G_p''\left(\mathcal{X}_p^{\alpha}\right)}{G_p'\left(\mathcal{X}_p^{\alpha}\right)} G_{p,\delta^2}'\left(\mathcal{X}_p^{\alpha}\right) = \frac{\delta^2}{2\mathcal{X}_p^{\alpha}} G_{(p+2),\delta^2}'\left(\mathcal{X}_p^{\alpha}\right).
$$

Combining the above steps, we get

$$
\begin{aligned}
P(F_{T,GLS} \leq \kappa\mathcal{F}_{p,K}^{\alpha}|H_1(\delta^2)) = {} & G_{p,\delta^2}\left(\mathcal{X}_p^{\alpha}\right) + (bT)^{-q} G_{p,\delta^2}'\left(\mathcal{X}_p^{\alpha}\right)\mathcal{X}_p^{\alpha}\bar{B} \\
& + \frac{\delta^2}{2} G_{(p+2),\delta^2}'\left(\mathcal{X}_p^{\alpha}\right)\mathcal{X}_p^{\alpha} c_2 b + o(b) + o\left((bT)^{-q}\right).
\end{aligned}
$$

Next, we prove the theorem in the GMM setting. As in the proof of Theorem 3, we can write

$$
pF_T = pF_{T,L}^c + \psi_T^c + \frac{1}{\sqrt{T}}\left[\sqrt{b} + (bT)^{-q}\right]\psi_T^{c*}
$$

where

$$
pF_{T,L}^c = \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t + \Lambda\tilde{c}\right]' \tilde{\Omega}_T^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t + \Lambda\tilde{c}\right],
$$

$\psi_T^c = O_p(1/\sqrt{T})$ does not depend on b, and $\psi_T^{c*} = O_p([\sqrt{b} + (bT)^{-q}]/\sqrt{T} + 1/T)$. $pF_{T,L}^c$ is equal to the Wald statistic of testing whether the mean of the process $u_t := \phi(v_t, \theta_0)$ is zero under the local alternative hypothesis. Using the same argument as the proof of Theorem 3, we have

$$
\begin{aligned}
&P(F_T^* \leq \mathcal{F}_{p,K}^{\alpha}|H_1(\delta^2)) \\
&\quad = G_{p,\delta^2}\left(\mathcal{X}_p^{\alpha}\right) + (bT)^{-q}\, G'_{p,\delta^2}\left(\mathcal{X}_p^{\alpha}\right)\mathcal{X}_p^{\alpha}\bar{B} \\
&\quad + \frac{\delta^2}{2} G'_{(p+2),\delta^2}\left(\mathcal{X}_p^{\alpha}\right)\mathcal{X}_p^{\alpha} c_2 b + O\left(\frac{\log T}{\sqrt{T}}\right) + o(b) + o\left((bT)^{-q}\right),
\end{aligned}
$$

as stated. ■

References

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

[2] Bartlett, M.S., (1937): "Properties of Sufficiency and Statistical Tests." *Proceedings of the Royal Society* A, 160, 268–282.

[3] Bartlett, M.S., (1954): "A Note on the Multiplying Factors for Various χ^2 Approximations." *Journal of the Royal Statistical Society*, Series B, Vol. 16, 296–298.

[4] Bester, C. A., T. G. Conley, and C.B. Hansen (2011): "Inference with Dependent Data Using Cluster Covariance Estimators." *Journal of Econometrics* 165(2), 137–51.

[5] Cribari-Neto, F. and Cordeiro, G.M. (1996): "On Bartlett and Bartlett-type Corrections." *Econometric Reviews*, 15, 339–367.

[6] Doukhan, P. (1995): *Mixing: Properties and Examples.* Springer-Verlag, New York:NY

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

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

[9] Gao, J. and I. Gijbels (2008): "Bandwidth Selection in Nonparametric Kernel Testing." Journal of the American Statistical Association, 484, 1584-1594.

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

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

[12] ——— (2002b): "Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation," *Econometrica,* 70, 2093–2095.

[13] Kiefer, N. M. and T. J. Vogelsang (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests." *Econometric Theory,* 21, 1130–1164.

[14] Lee, W. and C. Kuan (2009): "Testing Over-identifying Restrictions without Consistent Estimation of the Asymptotic Covariance Matrix." Working paper.

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

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

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

[18] Promislow, S. D. (2008): *A First Course in Functional Analysis*, Wiley.

[19] Ravikumar, B., S. Ray, & N. E. Savin (2004): "Robust Wald Tests and the Curse of Dimensionality." Working paper, Department of Economics, University of Iowa.

[20] Ray, S. and N. E. Savin (2008): "The Performance of Heteroskedasticity and Autocorrelation Robust Tests: a Monte Carlo Study with an Application to the Three-factor Fama-French Asset-pricing Model, " *Journal of Applied Econometrics*, 23, 91-109.

[21] Ray, S., N. E. Savin, and A. Tiwari (2009): "Testing the CAPM Revisited." *Journal of Empirical Finance*, 16(5), 721–733.

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

[23] Sun, Y. and P. C. B. Phillips (2009): "Bandwidth Choice for Interval Estimation in GMM Regression," Working paper, Department of Economics, UC San Diego.

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

[25] Tanaka, K. (1996): *Time Series Analysis*, Wiley.

[26] Velasco, C. and P. M. Robinson (2001): "Edgeworth Expansions for Spectral Density Estimates and Studentized Sample Mean." *Econometric Theory*, 17(3), 497–539.

[27] White, H. (2001): *Asymptotic Theory for Econometricians.* Academic Press.

[28] Zhang, X and X. Shao (2012): "Fixed-Smoothing Asymptotics For Time Series." *The Annals of Statistics*, forthcoming.