# Asymptotic F Test in a GMM Framework

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

July 20, 2011

**Abstract**

The paper develops a new and easy-to-use $F$ test in a time series GMM framework that allows for general forms of serial dependence. The test is based on the Wald statistic with a multiplicative correction factor and employs critical values from a standard $F$ distribution. The $F$ critical values are high-order correct under the conventional asymptotics. Monte Carlo simulations show that the $F$ test is more accurate in size than the conventional chi-square test and is as accurate in size as the nonstandard test based on the fixed-$b$ asymptotics.

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

*Keywords:* Asymptotic expansion, F distribution, Heteroscedasticity and Autocorrelation, Hotelling's T-squared distribution, long run variance, robust standard error.

---

*An earlier version of the paper is available as Sun (2010b) with title "Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroscedasticity and Autocorrelation Robust Inference," which contains additional results of independent interest. Email: yisun@ucsd.edu. The author thanks James Stock, three anonymous referees, and many seminar participants for helpful comments. The author gratefully acknowledges partial research support from NSF under Grant No. SES-0752443. Correspondence to: Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508.

# 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 Newey and West (1987) and Andrews (1991). In econometrics, this covariance estimator is often referred to as the heteroscedasticity and autocorrelation robust (HAR) estimator. A major difficulty in using the HAR covariance estimator to perform hypothesis testing lies in how to specify 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 a certain rate with the sample size. In the latter case, $b$ is assumed to be held fixed at a given value. Under these two different asymptotic specifications, the Wald statistic converges in distribution to the standard chi-square distribution and a nonstandard distribution, respectively. See Kiefer, Vogelsang and Bunzel (2000), 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 $\chi^2$ approximation. Jansson (2004) and Sun, Phillips, Jin (2008, hereafter SPJ) provide theoretical analyses for location models.

In this paper, we propose a new $F^*$ test, which is based on the Wald statistic corrected by a multiplicative factor and employs critical values from a standard $F$ distribution. The correction can be regarded as an example of the Bartlett or Bartlett-type correction. See Bartlett (1937, 1954). It 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. The $F^*$ test is as easy to use as the standard Wald test as both the correction factor and the critical values are easy to obtain.

The $F^*$ test can be motivated as an approximation to the nonstandard limiting distribution given in KV (2005). Under the fixed-$b$ asymptotics, the HAR covariance estimator converges in distribution to a weighted sum of independent Wishart distributions. The weighted sum can be approximated by a single Wishart distribution with equivalent degree of freedom $K$. A direct implication is that the nonstandard fixed-$b$ limiting distribution can be approximated by a scaled $F$ distribution. The scaled $F$ distribution can be regarded as a high-order limiting distribution under the sequential asymptotics where the sample size $T \to \infty$ for a fixed $b$ followed by letting $b \to 0$.

We show that critical values from the scaled $F$ distribution are high-order correct under the conventional joint small-$b$ asymptotics where $T \to \infty$ and $b \to 0$ jointly. So under both the sequential asymptotics and the joint asymptotics, the $F$ approximation is a high-order refinement of the conventional $\chi^2$ approximation. This provides a theoretical justification on the accuracy of the $F$ approximation derived in this paper.

On the basis of the $F$ approximation, we provide a theoretical explanation on why the conventional Wald test has severe size distortion when $p$, the number of hypotheses being jointly tested or the dimension of the hypothesis space, is large. We show that the difference between the high-order corrected $F$ critical value and the first-order $\chi^2$ 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 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 establishes an F-approximation to the nonstandard fixed-$b$ asymptotic distribution. Section 4 develops a high-order expansion of the Wald statistic and introduces the $F^*$ test. Section 5 presents simulation evidence and last section concludes. Proofs are given in the Appendix.

## 2 Autocorrelation Robust Testing

The model we consider is the same as KV (2005). 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 $\theta_0$ be the true value and assume that $\theta_0$ is an interior point of the compact parameter space $\Theta$. The moment conditions

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

hold if and only if $\theta = \theta_0$ where $f(\cdot)$ is an $m \times 1$ vector of 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 $\theta_0$ 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

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

As in KV (2005), we make the following high level assumptions.

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

**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 $\theta_0$.*

**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 continuously differentiable functions with first order derivative matrix $R(\theta) = \partial r(\theta)/\partial\theta'$. The Wald statistic is based on the difference $r(\hat{\theta}_T) - r(\theta_0)$. Under Assumptions 1–3, we have, using (1):

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

where

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

This can be regarded as an influence function representation of $\sqrt{T}[r(\hat{\theta}_T) - r(\theta_0)]$.

**Assumption 4** *$T^{-1/2}\sum_{t=1}^{[Tr]}\phi(v_t,\theta_0) \to^d \Lambda W_p(r)$ where $\Lambda\Lambda' = \Omega = \sum_{j=-\infty}^{\infty} Eu_t u_{t-j}$ is the long run variance (LRV) of $u_t := \phi(v_t,\theta_0)$ and $W_p(r)$ is the $p$-dimensional standard Brownian motion.*

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

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

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

$$\hat{\Omega}_T = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}k_b\left(\frac{t-\tau}{T}\right)\hat{u}_t\hat{u}_\tau' \tag{2}$$

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

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

$k(\cdot)$ 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), we can show that under the assumptions given above:

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

for any fixed value of $b$, where

$$F_\infty(p,b) = W_p'(1)\left[\int_0^1\int_0^1 k_b(r-s)dV_p(r)dV_p'(s)\right]^{-1}W_p(1)/p, \tag{3}$$

$W_p(r)$ and $V_p(r)$ are $p$-dimensional Brownian motion and Brownian bridge processes, respectively.

$F_\infty(p,b)$ 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 F-Approximation to the Nonstandard Limits

This section develops an asymptotic expansion of the limit distribution given in (3) as the bandwidth parameter $b \to 0$. On the basis of this expansion, we derive an $F$ approximation to the nonstandard fixed-$b$ limiting distribution.

The asymptotic expansion and later developments in the paper make use of the following kernel conditions:

**Assumption 5** *(i) $k(x)$ is an even function satisfying $k(0) = 1$, $\int_{-\infty}^{\infty} |k(x)x| dx < \infty$, $\int_{-\infty}^{\infty} k^2(x)x^2 dx < \infty$ (ii) The Parzen characteristic exponent defined by*

$$q = \max\{q_0 : q_0 \in \mathbb{Z}^+, \ g_{q_0} = \lim_{x \to 0} \frac{1 - k(x)}{|x|^{q_0}} < \infty\} \tag{4}$$

*is greater than or equal to 1.*

We first establish different representations of the fixed-$b$ limiting distribution. These representations help motivate the $F$ approximation and are used in our proofs. 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, page 199), we can expand $k_b^*(r,s)$ as

$$k_b^*(r,s) = \sum_{n=1}^{\infty} \lambda_n^* f_n^*(r) f_n^*(s), \tag{5}$$

where the right hand side converges in $L^2([0,1] \times [0,1])$. Here $\lambda_n^*$ is an eigenvalue of the centered kernel and $f_n^*(r)$ is the corresponding eigenfunction, i.e. $\lambda_n^* f_n^*(s) = \int_0^1 k_b^*(r,s) f_n^*(r)dr$. It follows from (5) that

$$\int_0^1 \int_0^1 k_b^*(r,s)dW_p(r)dW_p'(s) \overset{d}{=} \sum_{n=1}^{\infty} \lambda_n^* \zeta_n \zeta_n' \tag{6}$$

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

$$pF_\infty(p,b) \overset{d}{=} \eta' \left[ \sum_{n=1}^{\infty} \lambda_n^* \zeta_n \zeta_n' \right]^{-1} \eta, \tag{7}$$

where $\zeta_n \sim iidN(0, \mathbb{I}_p)$, $\eta \sim N(0, \mathbb{I}_p)$ and $\zeta_n$ is independent of $\eta$ for all $n$. 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 $H$ be an orthonormal matrix such that $H = (\eta/\|\eta\|, \Pi)'$ where $\Pi$ is a $p \times (p-1)$ matrix, then

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

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

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

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

$$pF_\infty(p, b) \stackrel{d}{=} \frac{\|\eta\|^2}{\nu_{11\cdot 2}} \text{ for } \nu_{11\cdot 2} = \nu_{11} - \nu_{12}\nu_{22}^{-1}\nu_{21}. \tag{8}$$

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.

Using Lemma A.2 in the appendix and the Markov inequality, we can show that $\nu_{11\cdot 2} \to^p 1$ as $b \to 0$. So as $b$ decreases, the effect of the randomness in $\nu_{11\cdot 2}$ diminishes, and when $b \to 0$, $pF_\infty(p, b)$ approaches the standard $\chi_p^2$ distribution. In other words, under the sequential asymptotics where $T \to \infty$ for a fixed $b$ followed by letting $b \to 0$, we obtain the standard $\chi_p^2$ approximation.

The standard $\chi_p^2$ approximation may be regarded as the first-order sequential asymptotics. To obtain a high-order approximation, we can refine the second stage approximation. A general idea to improve the chi-square approximation is the Bartlett 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 $X \sim pF_\infty(p, b)$ and $EX = pC$ for some constant $C$, then as $b \to 0$, $pF_\infty(p, b)/C$ is closer to the $\chi_p^2$ distribution than the original distribution $pF_\infty(p, b)$. Since $\|\eta\|^2$ and $\nu_{11\cdot 2}$ in (8) are independent, we have $EX = pE\left[\nu_{11\cdot 2}^{-1}\right]$. So $C = E\left[\nu_{11\cdot 2}^{-1}\right]$ or a good approximation to $E\left[\nu_{11\cdot 2}^{-1}\right]$.

We can go one step further. Essentially, the Bartlett correction makes the mean of the $pF_\infty(p, b)$ distribution closer to that of the $\chi_p^2$ distribution. It takes into account only the first moment of $\nu_{11\cdot 2}$ but not its second moment. In view of (8), $F_\infty(p, b)$ is a ratio of a scaled chi-square distribution to another chi-square type distribution. So it is natural to approximate $F_\infty(p, b)$ by an $F$ distribution. We can design an $F$ distribution that matches both the first and second moments of $\nu_{11\cdot 2}$.

It is more convenient to use the matrix representation in (7) to achieve moment matching. Let

$$\Phi = \left( E \sum_{n=1}^{\infty} \lambda_n^* \zeta_n \zeta_n' \right)^{-1} \sum_{n=1}^{\infty} \lambda_n^* \zeta_n \zeta_n'$$

where $E \sum_{n=1}^{\infty} \lambda_n^* \zeta_n \zeta_n' = \mu_1 \mathbb{I}_p$ for $\mu_1 = \sum_{n=1}^{\infty} \lambda_n^* = \int_0^1 k_b^*(r,r)\, dr$. We want to match the second moment of $\Phi$ with a scaled Wishart distribution $\Psi = \mathbb{W}_p(\mathbb{I}_p, K)/K$ for some integer $K > 0$. For any symmetric matrix $D$, we have

$$E\Psi D\Psi = \frac{1}{K} tr(D)\mathbb{I}_p + \left(1 + \frac{1}{K}\right) D. \tag{9}$$

See Example 7.1 in Bilodeau and Brenner (1999). Using this result, we can show that

$$E\Phi D\Phi = \frac{\mu_2}{\mu_1^2} tr(D)\mathbb{I}_p + \left(1 + \frac{\mu_2}{\mu_1^2}\right) D \tag{10}$$

where

$$\mu_2 = \sum_{n=1}^{\infty} (\lambda_n^*)^2 = \int_0^1 \int_0^1 [k_b^*(r,s)]^2\, drds.$$

In view of (9) and (10), we can set

$$K \approx \frac{\mu_1^2}{\mu_2} = \frac{\int_0^1 \int_0^1 k_b^*(r,r)\, k_b^*(s,s)\, drds}{\int_0^1 \int_0^1 k_b^*(r,s) k_b^*(r,s) drds}.$$

That is, $\Phi \approx^d \mathbb{W}_p(\mathbb{I}_p, K)/K$ for the above $K$ value where $\approx^d$ denotes "is approximately equal to in distribution." As a result

$$\mu_1 p F_\infty(p,b) \approx^d \eta' \Psi^{-1} \eta.$$

But $\eta'\Psi^{-1}\eta$ is the Hotelling's $T^2(p,K)$ distribution (Hotelling (1931)). By the well-known relationship between the $F$ distribution and the $T^2$ distribution, we have

$$\frac{\mu_1 (K-p+1)}{K} F_\infty(p,b) \approx^d F_{p,K-p+1}. \tag{11}$$

Using Lemma A.1 in the appendix, we can approximate the correction factor by

$$\frac{\mu_1 (K-p+1)}{K} = \kappa^{-1} + o(b) \tag{12}$$

where

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

and

$$c_1 = \int_{-\infty}^{\infty} k(x)dx \text{ and } c_2 = \int_{-\infty}^{\infty} k^2(x)dx.$$

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, $c_1 = 1.25$, $c_2 = 1$.

Combining (11) and (12) and approximating $F_{p,K-p+1}$ by $F_{p,K}$ if needed, we obtain Theorem 1.

**Theorem 1** *Let Assumption 5(i) hold. Let* $K = K^* - p + 1$ *or* $K^*$ *for*

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

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

*(i)* $P(p\kappa F_{p,K} \le z) = G_p(z) + A(z)b + o(b)$ *where*

$$A(z) = G_p''(z) z^2 c_2 - G_p'(z) z [c_1 + c_2 (p-1)],$$

*and* $G_p(\cdot)$ *is the cdf of the* $\chi_p^2$ *distribution.*

*(ii)* $P(pF_\infty(p,b) \le z) = P(p\kappa F_{p,K} \le z) + o(b)$.

Theorem 1 adjusts the value of $K^*$ to ensure that $K = K^* - p + 1 \ge 1$. The parameter $K$ or $K^*$ can be called the "equivalent degree of freedom (EDF)" of the LRV estimator. The idea of approximating a weighted sum of independent Wishart distributions by a simple Wishart distribution with equivalent degree of freedom can be motivated from the early statistical literature on spectral density estimation. In the scalar case, the distribution of the spectral density estimator is often approximated by a chi-square distribution with equivalent degree of freedom; see Priestley (1981, p. 467).

Note that the EDF $K^*$ is proportional to $1/b$ when $b$ is small while the asymptotic variance of the LRV estimator is proportional to $b$. Hence, as $b$ decreases, i.e. as the degree of smoothing increases, the EDF increases and the variance decreases. In other words, the higher the degree of freedom, the larger the degree of smoothing and the smaller the variance.

A direct implication of Theorem 1 is that

$$P\{pF_\infty(p,b) \le z\} = G_p(z) + A(z)b + o(b). \tag{13}$$

There are two terms in $A(z)b$. The term $G_p''(z) z^2 c_2 b$ arises from the asymptotic mean square error $E(\nu_{11\cdot 2} - 1)^2$ of $\nu_{11\cdot 2}$, while the term $-G_p'(z) z [c_1 + c_2 (p-1)] b$ arises from the asymptotic bias $E(\nu_{11\cdot 2} - 1)$ of $\nu_{11\cdot 2}$. The bias term comes from two sources. The first is the estimating uncertainty of model parameters. This is reflected in the dependence of $\nu_{11\cdot 2}$ 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 $p > 1$, $\nu_{11\cdot 2}$ is not equal to $\nu_{11}$ but its projected version, viz $\nu_{11} - \nu_{12}\nu_{22}^{-1}\nu_{21}$. In contrast, when $p = 1$, $\nu_{11\cdot 2}$ is equal to $\nu_{11}$ 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 $p = 1$, the expansion in (13) 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 hypotheses being jointly tested or the dimension of the hypothesis space. As $p$ increases, the difference between the nonstandard limiting distribution $F_\infty(p,b)$ and the standard distribution $\chi_p^2/p$ becomes larger.

Let $\mathcal{F}_{p,K}^\alpha$ and $\mathcal{F}_\infty^\alpha(p,b)$ be the $1-\alpha$ quantiles of the standard $F_{p,K}$ distribution and the nonstandard $F_\infty(p,b)$ distribution, respectively. Then

$$P\left(F_\infty(p,b) > \kappa \mathcal{F}_{p,K}^\alpha\right) = \alpha + o(b). \tag{14}$$

In other words, $\mathcal{F}_\infty^\alpha(p,b) = \kappa \mathcal{F}_{p,K}^\alpha + o(b)$. So for the original $F$ statistic, we can use $\kappa \mathcal{F}_{p,K}^\alpha$ as the critical value for the test with nominal size $\alpha$.

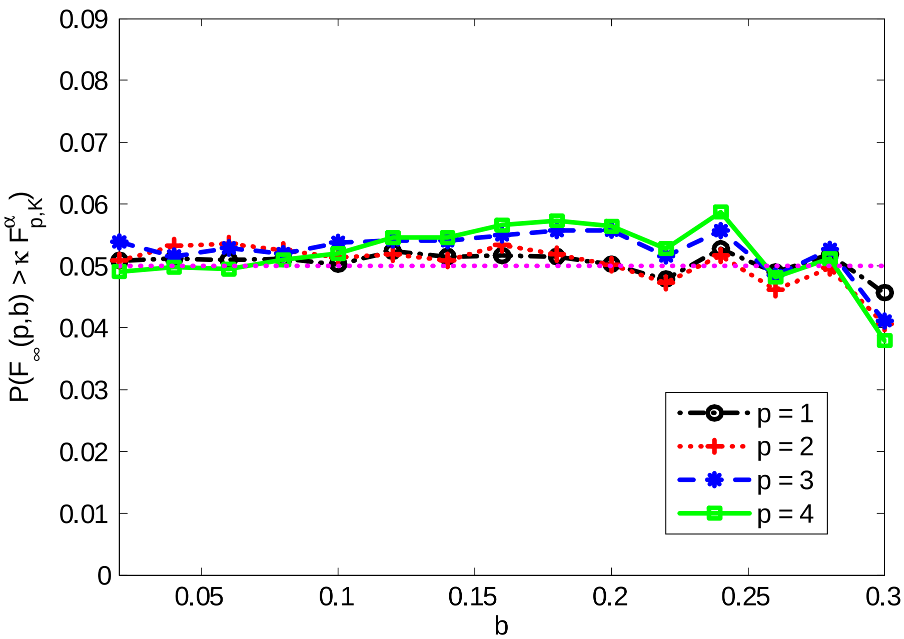


Figure 1: Graph of $P(F_\infty(p,b) > \kappa \mathcal{F}^\alpha_{p,K})$ as a function of $b$ for $\alpha = 5\%$, different values of $p$, and the Bartlett kernel

As an approximation to the nonstandard critical value, the scaled $F$ critical value $\kappa \mathcal{F}^\alpha_{p,K}$ is second-order correct as the approximation error in (14) is of smaller order $o(b)$ rather than $O(b)$ as $b \to 0$. The second-order critical value is larger than the standard critical value from $\chi^2_p/p$ for two reasons. First, $\mathcal{F}^\alpha_{p,K}$ is larger than the corresponding critical values from $\chi^2_p/p$ due to the presence of a random denominator in the $F$ distribution. Second, the correction factor $\kappa$ is larger than 1. As $b$ increases, both the correction factor and $F$ critical value $\mathcal{F}^\alpha_{p,K}$ increase. As a result, the second-order correct critical value $\kappa \mathcal{F}^\alpha_{p,K}$ is an increasing function of $b$.

To evaluate the accuracy of the approximate critical values $\kappa \mathcal{F}^\alpha_{p,K}$, we compute the rejection probability $P(F_\infty(p,b) > \kappa \mathcal{F}^\alpha_{p,K})$ via simulations and compare it with $\alpha$. In our simulations, we approximate the Brownian motion and Brownian bridge processes by normalized partial sums of $T = 1000$ *iid* $N(0,1)$ random variables and the number of replications is 10,000. Figures 1–3 graph the rejection probability $P(F_\infty(p,b) > \kappa \mathcal{F}^\alpha_{p,K})$ as functions of $b$ for the Bartlett, Parzen and QS kernels with $\alpha = 5\%$. The results for $\alpha = 10\%$ are qualitatively similar. For the Bartlett kernel, we use $K = K^*$ and for the Parzen and QS kernels, we use $K = K^* - p + 1$. These two choices are asymptotically equivalent but make a difference when $b$ is not small. For the Bartlett kernel, using $K = K^*$ gives a more accurate approximation. The figures show that the rejection probabilities are very close to the nominal size $\alpha$ when $b \leq 0.3$. This is especially true for the Bartlett kernel and $p = 1$ and 2.

Although the critical values from the nonstandard limiting distribution can be accurately simulated, we develop the $F$ approximation here for two reasons. First, the $F$ approximation is convenient to use for practical testing situations. Second, the $F$ approximation helps shed some new light on the nonstandard asymptotic theory. In the next section, we will show that the critical values from the $F$ distribution are also second-order correct under the conventional asymptotics.

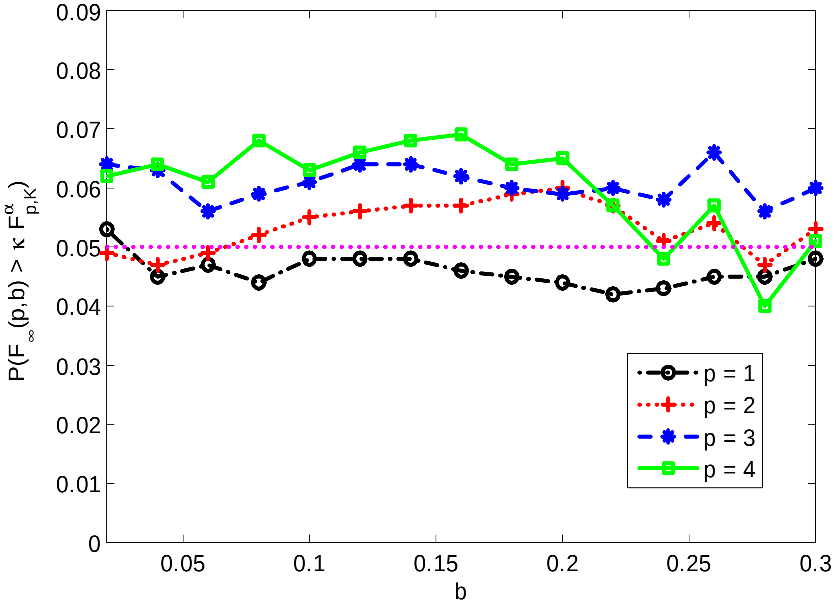


Figure 2: Graph of $P(F_\infty(p,b) > \kappa\mathcal{F}^{\alpha}_{p,K})$ as a function of $b$ for $\alpha = 5\%$, different values of $p$, and the Parzen kernel

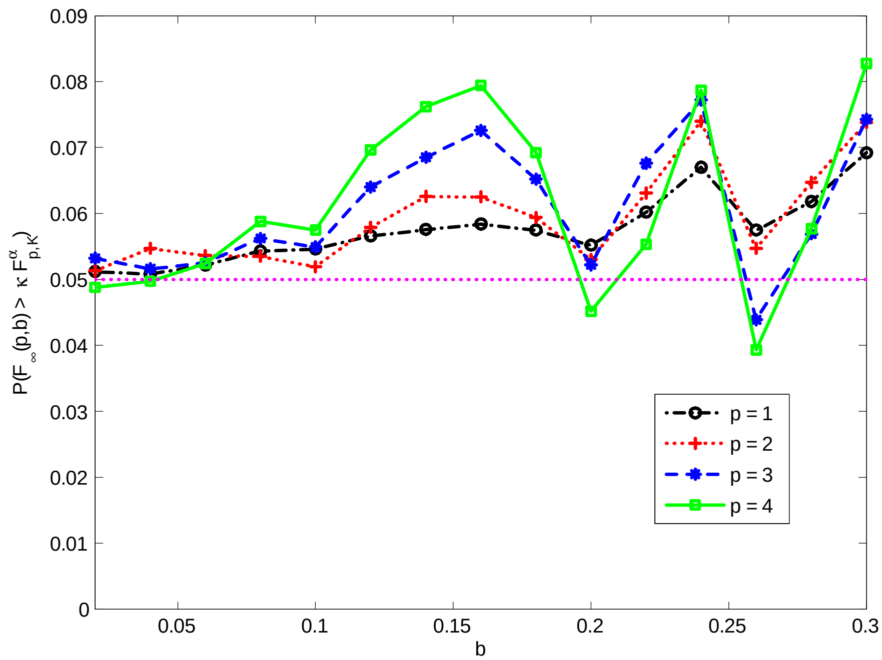


Figure 3: Graph of $P(F_\infty(p,b) > \kappa\mathcal{F}^{\alpha}_{p,K})$ as a function of $b$ for $\alpha = 5\%$, different values of $p$, and the QS kernel

# 4 Second-Order Correctness of the $F$ Approximation

In this section, we show that the $F$ approximation is high-order correct under the small-$b$ asymptotics where $b \to 0$ and $T \to \infty$ jointly. We employ a Gaussian location model to illustrate the basic point. One justification for using the Gaussian location model is that it is the limit of many statistical experiments; see van der Vaart (1995, Corollary 9.5). Another justification is that the high-order terms for the Gaussian location model are expected to appear in more general setting. See the working paper version of Sun (2010b). So it suffices to show that the $F$ approximation is high-order correct in the simple Gaussian location model.

Consider a vector time series $y_t$:

$$y_t = \theta + v_t, t = 1, 2, ..., T, \tag{15}$$

where $y_t = (y_{1t}, ..., y_{dt})'$, $\theta = (\theta_1, ..., \theta_d)'$, $v_t = (v_{1t}, ..., v_{dt})'$ is a stochastic process with zero mean.

The OLS estimator of $\theta$ is the average of $\{y_t\}$, viz $\hat{\theta}_{OLS} = T^{-1}\sum_{t=1}^{T} y_t$. We consider testing $H_0 : R_0\theta = r_0$ against $H_1 : R_0\theta \neq r_0$ for some $p \times d$ matrix $R_0$. This is a special case of tests with nonlinear restrictions. The nonlinear function $r(\theta)$ becomes linear in that $r(\theta) = R_0\theta - r_0$. 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:

$$F_{T,OLS} = \left[\sqrt{T}(R_0\hat{\theta}_{OLS} - r_0)\right]' \hat{\Omega}_T^{-1} \left[\sqrt{T}(R_0\hat{\theta}_{OLS} - r_0)\right]/p \tag{16}$$

where $\hat{\Omega}_T$ is defined as in (2) 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) $v_t$ is a stationary Gaussian process. (ii) For any $c \in \mathbb{R}^d$, the spectral density of $c'v_t$ is bounded above and away from zero in a neighborhood around the origin. (iii) The following FCLT holds:*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} u_t \Rightarrow \Omega^{1/2} W_p(r)$$

*where $\Omega$ is the long run variance matrix of $\{u_t\}$ and $W_p(r)$ is a $p$-dimensional Brownian motion process.*

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

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

where $\Omega_u = var([u_1', u_2', ..., u_T']')$, $y = [y_1', y_2', ..., y_T']'$ and $\ell_T$ is a vector of ones. Define

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

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 $E[(\hat{\theta}_{GLS}-\theta)\Delta']=0$ and $E[(\hat{\theta}_{GLS}-\theta)\hat{u}_t']=0$ for all $t$. So $\hat{\theta}_{GLS}-\theta$ is independent of both $\Delta$ 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 (16).

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

**Lemma 1** *Let Assumption 6 hold. Then*
*(a)* $P\left(pF_{T,GLS}\leq z\right)=EG_p\left(z\Xi_T^{-1}\right)+O\left(T^{-1}\right),$
*(b)* $P\left(pF_{T,OLS}\leq z\right)=P\left(pF_{T,GLS}\leq z\right)+O\left(T^{-1}\right),$
*where*

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

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

Lemma 1 shows that the estimation uncertainty of $\hat{\Omega}_T$ affects the distribution of the Wald statistic only through $\Xi_T$. 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* $\sum_{h=-\infty}^{\infty}|h|^q Eu_tu_{t-h}'<\infty$. *If* $b\to 0$ *such that* $bT\to\infty$, *then*

$$P\left(pF_{T,OLS}\leq z\right)=G_p\left(z\right)+A(z)b+\left(bT\right)^{-q}G_p'\left(z\right)z\bar{B}+o\left(b\right)+o\left(\left(bT\right)^{-q}\right) \tag{17}$$

*where*

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

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

The first term in (17) 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.

Let $\mathcal{X}_p^{\alpha}$ be the critical value from the $\chi_p^2$ distribution, then, up to smaller order terms,

$$P\left(pF_{T,OLS}>\mathcal{X}_p^{\alpha}\right)=\alpha-A(\mathcal{X}_p^{\alpha})b-\left(bT\right)^{-q}G_p'\left(\mathcal{X}_p^{\alpha}\right)\mathcal{X}_p^{\alpha}\bar{B}. \tag{18}$$

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 $-G_p''(\mathcal{X}_p^{\alpha})\left(\mathcal{X}_p^{\alpha}\right)^2c_2$ is positive. This is expected. Using $\chi_p^2$ 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 (e.g. Hannan (1957)). 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 rise to over-rejection. On the other hand, the nonparametric bias term $-(bT)^{-q} G_p'(\mathcal{X}_p^\alpha)\mathcal{X}_p^\alpha \bar{B}$ may be positive or negative, leading to over-rejection or under-rejection. The overall effect of the high-order terms depends on the sign of $\bar{B}$. When $\bar{B}$ is negative, the $F$-test is likely to over-reject. When $\bar{B}$ is positive, the $F$-test may over-reject or under-reject.

Comparing Theorem 1 with Theorem 2, we find that the $F$ approximation captures some terms in the high-order expansion of the small-$b$ asymptotics. By Theorem 1, we have $1 - G_p(p\kappa\mathcal{F}_{p,K}^\alpha) - A(p\kappa\mathcal{F}_{p,K}^\alpha)b = \alpha + o(b)$. Using this result and noting that $p\kappa\mathcal{F}_{p,K}^\alpha = \mathcal{X}_p^\alpha + O(b)$, we obtain

$$
\begin{aligned}
&P\left(F_{T,OLS} > \kappa\mathcal{F}_{p,K}^\alpha\right) = P\left(pF_{T,OLS} > p\kappa\mathcal{F}_{p,K}^\alpha\right) \\
&= 1 - G_p\left(p\kappa\mathcal{F}_{p,K}^\alpha\right) - A(p\kappa\mathcal{F}_{p,K}^\alpha)b - (bT)^{-q} G_p'\left(p\kappa\mathcal{F}_{p,K}^\alpha\right) p\kappa\mathcal{F}_{p,K}^\alpha \bar{B} + o\left(b\right) + o\left((bT)^{-q}\right) \\
&= \alpha - (bT)^{-q} G_p'\left(\mathcal{X}_p^\alpha\right)\mathcal{X}_p^\alpha \bar{B} + o\left(b\right) + o\left((bT)^{-q}\right). \qquad (19)
\end{aligned}
$$

Therefore, use of critical value $\kappa\mathcal{F}_{p,K}^\alpha$ removes the demeaning bias, dimensionality bias and variance term from the high-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 $(bT)^{-q} b^{-1} \to 0$, using critical value $\kappa\mathcal{F}_{p,K}^\alpha$ should lead to size improvements. We have thus shown that critical values from the $F$ distribution are second-order correct under the conventional small-$b$ asymptotics.

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 $F$ approximation and the fixed-$b$ asymptotics 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).

Define the finite sample corrected $F$ statistic as

$$F_{T,OLS}^* = F_{T,OLS}/\kappa. \qquad (20)$$

Then, up to smaller order terms, it follows from (19) that

$$P\left(F_{T,OLS}^* > \mathcal{F}_{p,K}^\alpha\right) = \alpha - (bT)^{-q} G_p'\left(\mathcal{X}_p^\alpha\right)\mathcal{X}_p^\alpha \bar{B}. \qquad (21)$$

So by adjusting both the test statistic and the critical value, we remove three high-order terms that contribute to over-rejection. Only the nonparametric bias term remains. For convenience, we refer to $F_{T,OLS}^*$ as the $F^*$ statistic and the associated test using $F$ critical value $\mathcal{F}_{p,K}^\alpha$ as the $F^*$ test.

In the working paper of Sun (2010b), it is shown that the high-order term of order $b$, namely $A(z)b$, also appears in the high-order expansion of the Wald statistic in a general GMM framework. So the $F$ approximation is high-order correct in more general settings.

# 5 Simulation Study

This section provides some simulation evidence on the finite sample performance of the $F^*$ test. We consider the following data generating process (DGP)

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

where the regressors $x_{t,j}$, $j = 1,2,3,4$ and regressor error $x_{t,0}$ follow mutually independent AR(1) processes, MA(1) processes or MA($m$) processes with the same model parameters. We normalize the regressors and regressor error to have unit variance. For the AR(1) process, let $u_{t,j} = \rho u_{t-1,j} + e_{t,j}$, $e_{t,j} \sim iidN(0,1)$, then $x_{t,j} = u_{t,j}/\sqrt{1-\rho^2}$, $j = 0,...,4$. We consider $\rho = -0.9, -0.6, -0.3, 0, 0.3, 0.6, 0.9$. For the MA(1) process, let $u_{t,j} = e_{t,j} + \psi e_{t,j-1}$, $e_{t,j} \sim iidN(0,1)$, then $x_{t,j} = u_{t,j}/\sqrt{1+\psi^2}$, $j = 0,...,4$. We consider $\psi = -0.99, -0.9, -0.6, -0.3, 0, 0.3, 0.6, 0.9, 0.99$. For the MA($m$) process, let $u_{t,j} = e_{t,j} + \sum_{r=1}^{m} \psi_r e_{t-r,j}$ with $\psi_r = 1 - r/(m+1)$ and $e_{t,j} \sim iidN(0,1)$, then $x_{t,j} = u_{t,j}/\sqrt{1+\sum_{r=1}^{m}\psi_r^2}$. As in Andrews and Monahan (1992), we consider $m = 3, 5, 7, 9, 12, 15$.

Let $\theta = (\theta_0, \theta_1, ..., \theta_4)'$. We estimate $\theta$ by the OLS estimator. Since the model is exactly identified, the weighted matrix $\mathcal{W}_T$ becomes irrelevant. Let $\tilde{x}_t' = [1, x_{1t}, ..., x_{4t}]$ 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:

$$H_{0p} : \theta_1 = ... = \theta_p = 0$$

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 $\mathbb{I}_5$. For each test we consider two significance levels $\alpha = 5\%$ and $\alpha = 10\%$, and two different sample sizes $T = 250, 500$. We consider three commonly used positive semi-definite kernels: Bartlett, Parzen, and QS kernels. The number of simulation replications is 5000.

To construct the Wald statistic, we have to select the bandwidth parameter $b$. Following Sun and Phillips (2009), we select $b$ to minimize the error in rejection probability of the standard Wald test under the null. Equivalently, we select $b$ to minimize the coverage probability error (CPE) of the associated confidence regions. That is

$$b_{CPE}^* = \arg\min_{b \in [0,0.5]} \left| P\left(pF_T \geq \mathcal{X}_p^\alpha\right) - \alpha \right|$$

The upper bound $b = 0.5$ is chosen to speed up the simulations. The simulation results remain more or less the same when $b$ is restricted to be in $[0,1]$. For each parameter configuration, we can discretize $[0, 0.5]$ and find the optimal $b_{CPE}^*$ by simulation. In the Monte Carlo experiments, we discretize $[0, 0.5]$ into a regular grid with step size 0.005. This bandwidth choice is not feasible in practice but it provides an ideal environment to evaluate the finite sample performance of the $F$ approximation.

A feasible version of $b_{CPE}^*$ can be obtained by approximating the coverage probability error. According to (18) or the high-order expansion in the GMM framework given by Sun (2010b), we have, up to smaller order:

$$\begin{aligned} \left| P\left(pF_T \geq \mathcal{X}_p^\alpha\right) - \alpha \right| &= \left| A(\mathcal{X}_p^\alpha) b + (bT)^{-q} G_p'\left(\mathcal{X}_p^\alpha\right) \mathcal{X}_p^\alpha \bar{B} \right| \\ &\leq \left| A(\mathcal{X}_p^\alpha) \right| b + (bT)^{-q} G_p'\left(\mathcal{X}_p^\alpha\right) \mathcal{X}_p^\alpha \left| \bar{B} \right|. \end{aligned} \quad (22)$$

The optimal $b$ that minimizes the upper bound in (22) is given by

$$b_{CPE} = \left( \frac{q \left| \bar{B} \right|}{c_1 + 0.5 c_2 \left( \mathcal{X}_p^{\alpha} + p \right)} \right)^{\frac{1}{q+1}} T^{-\frac{q}{q+1}}.$$

The CPE-optimal bandwidth parameter can be written as $b_{CPE} = b_{CPE}(\bar{B})$ where $\bar{B} = \bar{B}(B, \Omega)$. The parameter $\bar{B}$ is unknown but could be estimated by a standard plug-in procedure as in Andrews (1991). We fit an approximating VAR(1) process to $\{\hat{u}_t\}$ and use the estimated model to compute $\bar{B}(B, \Omega)$.

We examine the finite sample performance of the (modified) Wald tests for different smoothing parameter and reference distribution combinations. We consider both the infeasible bandwidth choice $b^*_{CPE}$ and the feasible one $b_{CPE}$. For each smoothing parameter choice, we consider three different reference distributions: $\chi^2_p/p$, $\kappa F_{p,K}$ and $F_\infty(p, b)$, leading to the conventional $\chi^2$ test, the $F^*$ test introduced in this paper, and the nonstandard test of KV(2005). There are six different testing procedures in total. We label them $b^*$-CPE-$\chi^2$, $b^*$-CPE-$F$, $b^*$-CPE-$N$, $b$-CPE-$\chi^2$, $b$-CPE-$F$, $b$-CPE-$N$ in the figures.

Figures 4-6 report the empirical size of the six different tests for $\alpha = 5\%$ and $T = 250$. The results for the Bartlett kernel and the MA(1) regressors and errors are given in Figure 4; the results for the Parzen kernel and MA($m$) regressors and errors are given in Figure 5; the result for the QS kernel and AR(1) regressors and errors are given in Figure 6. These figures are qualitatively representative of other scenarios. Several patterns emerge. First, it is clear that the conventional $\chi^2$ test can have a large size distortion, especially when the processes are persistent and the number of joint hypotheses is large. Second, the size distortion of the $F^*$ test and the nonstandard test is smaller than the conventional $\chi^2$ test. This is because the $F^*$ test and the nonstandard test employ asymptotic approximations that capture the estimation uncertainty of the LRV estimator. Third, the empirical size of the tests is robust to the bandwidth choice rules. This provides some evidence that the feasible and data-driven bandwidth parameter $b_{CPE}$ is a reasonable proxy to the infeasible bandwidth parameter $b^*_{CPE}$. Figures for the selected bandwidths not reported here show that the feasible and infeasible bandwidths are close to each other. Finally and most importantly, the size difference between the $F^*$ test and the corresponding nonstandard test is small. This is consistent with Figures 1–3.

In our simulations, we have combined different testing procedures with the prewhitening HAR estimators of Andrews and Monahan (1992). To conserve space, we do not report the results here but comment on them briefly. We find that prewhitening helps reduce the size distortion of all tests considered. In addition, the $F^*$ test is as accurate in size as the nonstandard test.

# 6 Conclusion

On the basis of the fixed-$b$ asymptotics and high-order small-$b$ asymptotics, the paper proposes a new $F^*$ test in the GMM framework where the moment conditions may exhibit general forms of serial dependence. The $F^*$ 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.

The $F^*$ test can be combined with any bandwidth selection rule. In the Monte Carlo experiments, we employ a data-driven bandwidth selection that minimizes the error in rejection probability or the coverage probability error. Simulations show that the $F^*$ test is as accurate in size as the nonstandard test of KV(2005). We recommend using the $F^*$ test with a prewhitening HAR variance estimator and testing-oriented bandwidth selection in practical situations. At a

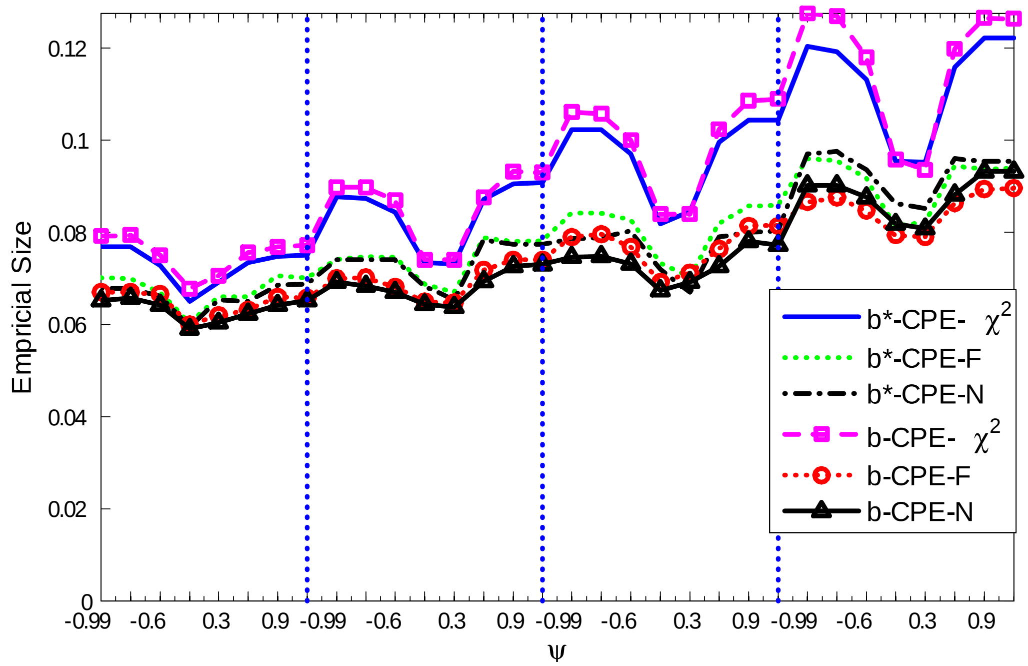


Figure 4: Empirical size of different nominal 5% tests for the Bartlett kernel under MA(1) regressors and errors with the number of joint hypotheses $p = 1, 2, 3, 4$ from left to right panels

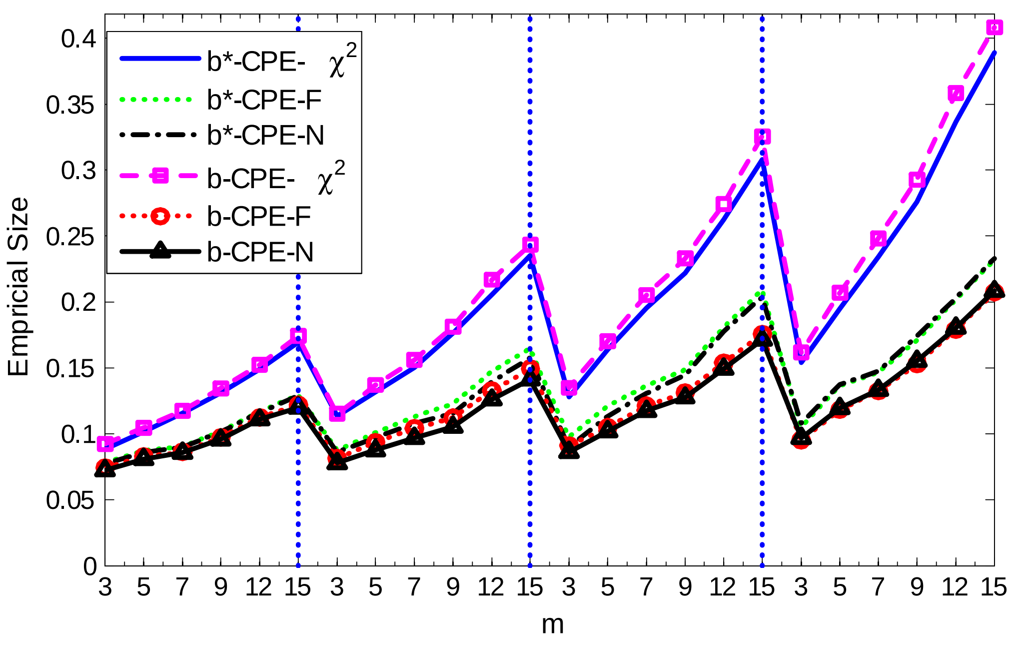


Figure 5: Empirical size of different nominal 5% tests for the Parzen kernel under MA($m$) regressors and errors with the number of joint hypotheses $p = 1, 2, 3, 4$ from left to right panels

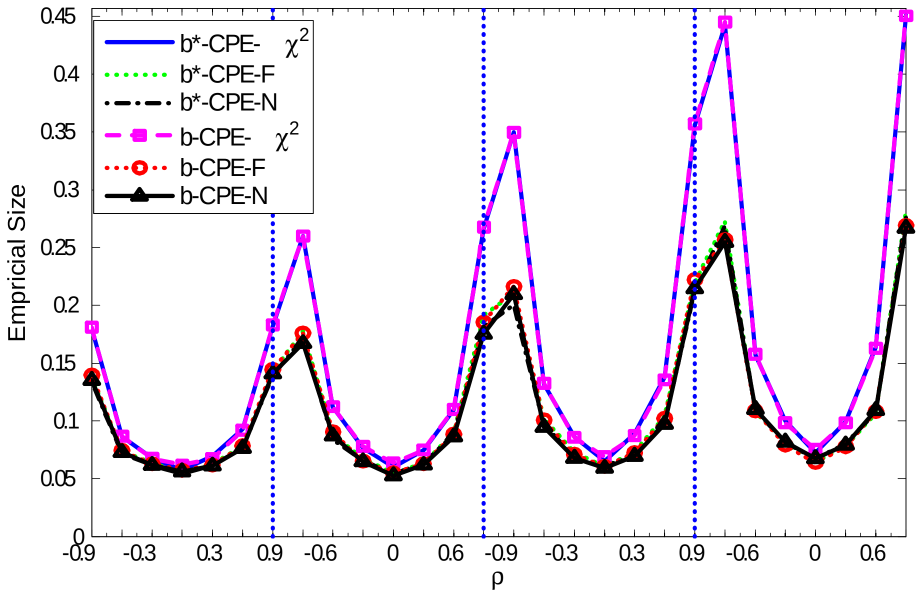


Figure 6: Empirical size of different nominal 5% tests for the QS kernel under AR(1) regressors and errors with the number of joint hypotheses $p = 1, 2, 3, 4$ from left to right panels

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.

# 7 Appendix of Proofs

We first state two lemmas whose proofs are available in Sun (2010b) and an online appendix.

**Lemma A.1** *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)* $\mu_2 = \sum_{n=1}^{\infty} (\lambda_n^*)^2 = bc_2 + O(b^2)$.

**Lemma A.2** *Let Assumption 5(i) hold. As* $b \to 0$, *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 Theorem 1.** We prove the theorem by showing that $P(pF_\infty(p,b) \le z) = G_p(z) + A(z)b + o(b)$ and $P\left(p\kappa F_{p,K} \le z\right) = G_p(z) + A(z)b + o(b)$. First, using (8) and taking a Taylor expansion, we have

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

where $\tilde{z}$ is between $z$ and $z\left(\nu_{11}-\nu_{12}\nu_{22}^{-1}\nu_{21}\right)$. Using Lemma A.2, we have

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

Second, by definition,

$$\begin{aligned} P\left(p\kappa F_{p,K}\leq z\right) &= P\left(\chi_p^2\leq\frac{z}{\kappa}\frac{\chi_K^2}{K}\right)=EG_p\left(\frac{z}{\kappa}\frac{\chi_K^2}{K}\right) \\ &= G_p\left(\frac{z}{\kappa}\right)+\frac{1}{K}G_p''\left(\frac{z}{\kappa}\right)\frac{z^2}{\kappa^2}+o\left(\frac{1}{K}\right) \\ &= G_p\left(z\right)+G_p'\left(z\right)z\left(\frac{1}{\kappa}-1\right)+G_p''\left(z\right)z^2bc_2+o\left(b\right) \\ &= G_p\left(z\right)+\left\{G_p''\left(z\right)z^2c_2-G_p'(z)z\left[c_1+c_2\left(p-1\right)\right]\right\}b+o\left(b\right) \\ &= G_p(z)+A(z)b+o(b), \end{aligned}$$

as desired. ■

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

$$\begin{aligned} pF_{T,GLS} &= \left[R_0T^{1/2}(\hat{\theta}_{GLS}-\theta)\right]'\Omega_{T,GLS}^{-1/2}\left[\Omega_{T,GLS}^{1/2}\hat{\Omega}_T^{-1}\Omega_{T,GLS}^{1/2}\right]\Omega_{T,GLS}^{-1/2}\left[R_0T^{1/2}(\hat{\theta}_{GLS}-\theta)\right] \\ &= \left\|\Omega_{T,GLS}^{-1/2}\left[R_0T^{1/2}(\hat{\theta}_{GLS}-\theta)\right]\right\|^2\times e_T'\left[\Omega_{T,GLS}^{1/2}\hat{\Omega}_T^{-1}\Omega_{T,GLS}^{1/2}\right]e_T \\ &\equiv \Upsilon_T\Xi_T+O_p\left(T^{-1}\right), \end{aligned}$$

where

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

Here we have used

$$\Omega_{T,GLS}=var\left[R_0\sqrt{T}\left(\hat{\theta}_{GLS}-\theta\right)\right]=\Omega\left(1+O\left(T^{-1}\right)\right).$$

Note that $\Upsilon_T$ is independent of $\Xi_T$ because (i) $(\hat{\theta}_{GLS}-\theta)$ is independent of $\hat{\Omega}_T$. (ii) $\Upsilon_T$ 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

$$P\left[pF_{T,GLS}\leq z\right]=P\left[\Upsilon_T\Xi_T\leq z\right]+O\left(T^{-1}\right)=EG_p\left(z\Xi_T^{-1}\right)+O\left(T^{-1}\right)$$

as stated.

**Part (b)**. Let

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

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

$$\begin{aligned}
pF_{T,OLS} &= \left[\sqrt{T}(R_0\hat{\theta}_{OLS} - r_0)\right]'\hat{\Omega}_T^{-1}\left[\sqrt{T}(R_0\hat{\theta}_{OLS} - r_0)\right]/p \\
&= \left[\sqrt{T}(R_0\hat{\theta}_{GLS} - r_0) + R_0\sqrt{T}\Delta\right]'\hat{\Omega}_T^{-1}\left[\sqrt{T}(R_0\hat{\theta}_{GLS} - r_0) + R_0\sqrt{T}\Delta\right]/p \\
&= pF_{T,GLS} + \zeta_{2T} + 2\left[R_0\sqrt{T}\Delta\right]'\hat{\Omega}_T^{-1}\Omega_{T,GLS}^{1/2}\left[\Omega_{T,GLS}^{-1/2}\sqrt{T}R_0(\hat{\theta}_{GLS} - \theta)\right]/p \\
&= pF_{T,GLS} + \zeta_T.
\end{aligned}$$

Note that $\Upsilon_T$ is independent of $\zeta_{1T}, \zeta_{2T}$ and $\Xi_T$, we have

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

where

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

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 $O\left(E\zeta_{2T}\right) = O(1/T)$. Next, let $f_e(x)$ be the pdf of $e_T$. Since $e_T$ is independent of $\hat{\Omega}_T$ and $\Delta$, we have

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

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

$$E\left\{\left[F_1'\left(0, 0, x'\left[\Omega^{1/2}\hat{\Omega}_T^{-1}\Omega^{1/2}\right]x\right)\left(R_0T^{1/2}\Delta\right)'\hat{\Omega}_T^{-1}\Omega_{T,GLS}^{1/2}x\right\} = 0$$

for all $x$. As a result, $EF_1'\left(0, 0, \Xi_T\right)\zeta_{1T} = 0$ and 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

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

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

$$
\left[\Xi_T\left(\hat{\Omega}_T\right)\right]^{-1} = 1 + L + Q + remainder \tag{23}
$$

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

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

$\mathbb{K}_{dd}$ is the $d^2 \times d^2$ commutation matrix, 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

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

So

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

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

$$
EL^2 = 2c_2 b + o\left(b + (bT)^{-q}\right),
$$

and

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

Hence

$$
\left[\Xi_T\left(\hat{\Omega}_T\right)\right]^{-1} = 1 + L + Q + o_p\left(b + (bT)^{-q}\right). \tag{24}
$$

Using the above asymptotic expansion, we have

$$\begin{aligned}
P\left(pF_{T,OLS} \le z\right) &= P\left(\Upsilon_T \le z\Xi_T^{-1}\right) + O\left(T^{-1}\right) \\
&= EG_p\left(z\left(1+L+Q\right)\right) + o\left(b + (bT)^{-q}\right) \\
&= G_p\left(z\right) + G_p'\left(z\right) zE\left(L+Q\right) + \frac{1}{2}EG_p''\left(z\right) z^2\left(EL^2\right) + o\left(b+(bT)^{-q}\right) \\
&= G_p\left(z\right) + (bT)^{-q} G_p'\left(z\right) z\bar{B} - bc_1 G_p'\left(z\right) z \\
&- bc_2 G_p'\left(z\right) z\left(p-1\right) + bc_2 G_p''\left(z\right) z^2 + o\left(b\right) + o\left((bT)^{-q}\right) \\
&= G_p\left(z\right) + A(z)b + (bT)^{-q} G_p'\left(z\right) z\bar{B} + o\left(b\right) + o\left((bT)^{-q}\right)
\end{aligned}$$

as desired. ■

## 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 $\chi^2$ Approximations." *Journal of the Royal Statistical Society*, Series B, Vol. 16, 296–298.

[4] Bilodeau, M. and D. Brenner (1999): *Theory of Multivariate Statistics.* New York: Springer

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

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

[7] Hannan, E. J. (1957): "The Variance of the Mean of a Stationary Process." *Journal of the Royal Statistical Society*, Series B, Vol. 19(2), 282–285

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

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

[10] Kiefer, N. M., T. J. Vogelsang and H. Bunzel (2000): "Simple Robust Testing of Regression Hypotheses." *Econometrica,* 68, 695–714.

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

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

[13] ——— (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests." *Econometric Theory,* 21, 1130–1164.

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

[15] ——— (1994): "Automatic Lag Selection in Covariance Estimation." *Review of Economic Studies*, 61, 631–654.

[16] Priestley, M.B. (1981): *Spectral Analysis and Time Series*, Academic Press.

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

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

[19] 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.

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

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

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

[23] Sun, Y. (2010a): "Autocorrelation Robust Inference Using Nonparametric Series Methods." Working paper, Department of Economics, UC San Diego.

[24] Sun, Y. (2010b): "Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroscedasticity and Autocorrelation Robust Inference." Department of Economics, UC San Diego.

[25] Sun, Y. (2011): "Robust Trend Inference with Series Variance Estimator and Testing-optimal Smoothing Parameter." *Journal of Econometric*s, Forthcoming.

[26] van der Vaart, A.W. (1995): *Asymptotic Statistics*, Cambridge University Press

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

Supplement to
"Asymptotic F Test in a GMM Framework"

by Yixiao Sun

This Supplement provides proofs to the two lemmas in the appendix of the above paper.

**Lemma A.1** *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 A.1.** 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

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

To evaluate $\mu_1$ and $\mu_2$, we let

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

Then

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

For the integral that appears in both $\mu_1$ and $\mu_2$, we have

$$\begin{aligned}&\int_0^1 \int_0^1 k_b(r-s)drds \\ &= \int_{-\infty}^{\infty} \mathcal{K}_1(\lambda)\left[\int_0^1 \exp\left(\frac{i\lambda r}{b}\right)dr\right]\left[\int_0^1 \exp\left(-\frac{i\lambda s}{b}\right)ds\right]d\lambda \\ &= \int_{-\infty}^{\infty} \mathcal{K}_1(\lambda)\frac{b^2}{\lambda^2}\left[\left(1-\cos\left(\frac{\lambda}{b}\right)\right)^2 + \left(\sin\left(\frac{\lambda}{b}\right)\right)^2\right]d\lambda \\ &= b\int_{-\infty}^{\infty} \mathcal{K}_1(\lambda)\, b\left(\frac{\sin\frac{\lambda}{2b}}{\frac{\lambda}{2}}\right)^2 d\lambda \\ &= 2\pi b\mathcal{K}_1(0) + 4b^2\int_{-\infty}^{\infty} \frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\left(\sin\frac{\lambda}{2b}\right)^2 d\lambda,\end{aligned} \tag{A.3}$$

where the last equality holds because

$$\int_{-\infty}^{\infty}\left(\frac{\lambda}{2b}\right)^{-2}\left(\sin\frac{\lambda}{2b}\right)^{2}d\lambda = 2b\int_{-\infty}^{\infty}x^{-2}\sin^2 x dx = 2\pi b. \tag{A.4}$$

Now,

$$\begin{aligned}
&\int_{-\infty}^{\infty}\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\left(\sin\frac{\lambda}{2b}\right)^2 d\lambda \\
&= \int_{-\infty}^{\infty}\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\left(\left(\sin\frac{\lambda}{2b}\right)^2-\frac{1}{2}\right)d\lambda + \frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}d\lambda \\
&= -\frac{1}{2}\int_{-\infty}^{\infty}\left(\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\right)\left(\cos\frac{1}{b}\lambda\right)d\lambda + \frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}d\lambda \\
&= \frac{1}{2}\int_{-\infty}^{\infty}\left(\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\right)d\lambda + o(1)
\end{aligned} \tag{A.5}$$

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 (A.3) and (A.5):

$$\begin{aligned}
&\int_0^1\int_0^1 k_b(r-s)drds \\
&= 2\pi b\mathcal{K}_1(0) + 2b^2\int_{-\infty}^{\infty}\left(\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\right)d\lambda + o(b^2) \\
&= 2\pi b\mathcal{K}_1(0) + b^2\frac{1}{\pi}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}k(x)\frac{\cos\lambda x - 1}{\lambda^2}dx\right)d\lambda + o(b^2) \\
&= 2\pi b\mathcal{K}_1(0) - 2b^2\frac{1}{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}k(x)\frac{\sin^2(\lambda x/2)}{\lambda^2}dxd\lambda + o(b^2) \\
&= 2\pi b\mathcal{K}_1(0) - b^2\int_{-\infty}^{\infty}k(x)|x|\,dx + o\left(b^2\right) \\
&= bc_1 + O(b^2).
\end{aligned} \tag{A.6}$$

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

$$\int_0^1\int_0^1 k_b^2(r-s)drds = bc_2 + O(b^2). \tag{A.7}$$

Next,

$$\begin{aligned}
&\int_0^1 k_b(r-s)ds \\
&= \frac{1}{2}\int_{-\infty}^{\infty}\mathcal{K}_1(\lambda)\int_0^1\left\{\exp\left(\frac{i\lambda(r-s)}{b}\right)+\exp\left(-\frac{i\lambda(r-s)}{b}\right)\right\}dsd\lambda \\
&= \int_{-\infty}^{\infty}\mathcal{K}_1(\lambda)\int_0^1\cos\left(\frac{\lambda(r-s)}{b}\right)dsd\lambda \qquad \text{(A.8)} \\
&= -b\int_{-\infty}^{\infty}\mathcal{K}_1(\lambda)\frac{1}{\lambda}\left[\sin\left(\frac{\lambda(r-1)}{b}\right)-\sin\left(\frac{\lambda r}{b}\right)\right]d\lambda \\
&= -b\int_{-\infty}^{\infty}\mathcal{K}_1(xb)\frac{1}{x}\left[\sin(x(r-1)) - \sin(xr)\right]dx,
\end{aligned}$$

so

$$
\begin{aligned}
&\int_0^1 \int_0^1 \int_0^1 k_b(r-p)k_b(r-q)drdpdq \\
&= b^2 \int_0^1 \left[\int_{-\infty}^{\infty} \mathcal{K}_1\left(xb\right) \frac{1}{x}\left[\sin\left(x(r-1)\right) - \sin\left(xr\right)\right] dx\right]^2 dr \\
&= b^2 \mathcal{K}_1^2\left(0\right) \int_0^1 \left(\int_{-\infty}^{\infty} \frac{1}{x} \sin\left(x(r-1)\right) dx - \int_{-\infty}^{\infty} \frac{1}{x} \sin\left(xr\right) dx\right)^2 dr\left(1+o\left(1\right)\right) \\
&= b^2 \mathcal{K}_1^2\left(0\right) \int_0^1 \left(-\int_{-\infty}^{\infty} \frac{\sin\left(x(r-1)\right)}{x(r-1)} d(x(r-1)) - \int_{-\infty}^{\infty} \frac{1}{xr} \sin\left(xr\right) d\left(xr\right)\right)^2 dr\left(1+o\left(1\right)\right) \\
&= b^2 \mathcal{K}_1^2\left(0\right) \int_0^1 \left(2\int_{-\infty}^{\infty} \frac{1}{y} \sin\left(y\right) dy\right)^2 dr\left(1+o\left(1\right)\right) = c_1^2 b^2 + o(b^2). \qquad \text{(A.9)}
\end{aligned}
$$

Combining (A.6), (A.7), and (A.9) yields the lemma. ■

**Lemma A.2** *Let Assumption 5(i) hold. As $b \to 0$, we have*
*(a)* $E\left(\nu_{11} - \nu_{12}\nu_{22}^{-1}\nu_{21}\right) = 1 - bc_1 - bc_2\left(p-1\right) + o(b)$,
*(b)* $E\left[\left(\nu_{11} - \nu_{12}\nu_{22}^{-1}\nu_{21}\right)^2\right] = 1 - 2b\left(c_1 - c_2\right) - 2\left(p-1\right)bc_2 + o(b)$,
*(c)* $E\left[\left(\nu_{11} - \nu_{12}\nu_{22}^{-1}\nu_{21}\right) - 1\right]^2 = 2bc_2 + o(b)$.

**Proof of Lemma A.2.** (a) Let $W_p(r) = \left[W_1'(r), W_{p-1}'\left(r\right)\right]'$, 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 A.1, and

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

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

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

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 A.1.

(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,

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

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 A.1 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(\lambda_m^*)^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(\lambda_m^*)^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(\lambda_m^*)^2\xi_m\xi_m'\right]\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\} \\
= \; & \left[(p-1)^2\left(\sum_{n=1}^\infty\lambda_n^*\right)^{-2}\left(\sum_{m=1}^\infty(\lambda_m^*)^2\right)^2+2(p-1)\left(\sum_{n=1}^\infty\lambda_n^*\right)^{-2}\left(\sum_{m=1}^\infty(\lambda_m^*)^2\right)^2\right](1+o(1)) \\
= \; & o(b)
\end{aligned}
$$

using Lemma A.1.

Hence

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

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