Fixed-smoothing Asymptotics in a Two-step GMM Framework

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

July 12, 2013, First version: December 2011

Abstract

The paper develops the fixed-smoothing asymptotics in a two-step GMM framework. Under this type of asymptotics, the weighting matrix in the second-step GMM criterion function converges weakly to a random matrix and the two-step GMM estimator is asymptotically mixed normal. Nevertheless, the Wald statistic, the GMM criterion function statistic and the LM statistic remain asymptotically pivotal. It is shown that critical values from the fixed-smoothing asymptotic distribution are high order correct under the conventional increasing-smoothing asymptotics. When an orthonormal series covariance estimator is used, the critical values can be approximated very well by the quantiles of a noncentral F distribution. A simulation study shows that the new statistical tests based on the fixed-smoothing critical values are much more accurate in size than the conventional chi-square test.

JEL Classification: C12, C32

Keywords: F-distribution, Fixed-smoothing Asymptotics, Heteroskedasticity and Autocorrelation Robust, Increasing-smoothing Asymptotics, Noncentral F Test, Two-step GMM Estimation

1 Introduction

Recent research on heteroskedasticity and autocorrelation robust (HAR) inference has been focusing on developing distributional approximations that are more accurate than the conventional chi-square approximation or the normal approximation. To a great extent and from a broad perspective, this development is line with many other areas of research in econometrics where more accurate distributional approximations are the focus of interest. A common theme for coming up with a new approximation is to embed the finite sample situation in a different limiting thought experiment. In the case of HAR inference, the conventional limiting thought experiment assumes that the amount of smoothing increases with the sample size but a slower rate. The new limiting thought experiment assumes that the amount of smoothing is held fixed as the sample

*Email: yisun@ucsd.edu. For helpful comments, the author thanks Jungbin Hwang, David Kaplan, and Min Seong Kim. 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.

size increases. This leads to two types of asymptotics: the conventional increasing-smoothing asymptotics and the more recent fixed-smoothing asymptotics. Sun (2012) coins these two inclusive terms so that they are applicable to different HAR variance estimators, including both kernel HAR variance estimators and orthonormal series (OS) HAR variance estimators.

There is a large and growing literature on the fixed-smoothing asymptotics. For kernel HAR variance estimators, the fixed-smoothing asymptotics is the so-called the fixed-b asymptotics first studied by Kiefer and Vogelsang (2002a, 2002b, 2005, KV hereafter) in the econometrics literature. For other studies, see for example Jansson (2004), Sun, Phillips, Jin (2008), Sun and Phillips (2009), Gonçlaves and Vogelsang (2011) in the time series setting, Bester, Conley, Hansen and Vogelsang (2011, BCHV hereafter) and Sun and Kim (2012b) in the spatial setting, and Gonçalves (2011), Kim and Sun (2012), and Vogelsang (2012) in the panel data setting. For OS HAR variance estimators, the fixed-smoothing asymptotics is the so-called fixed-K asymptotics. For its theoretical development and related simulation evidence, see for example Phillips (2005), Müller (2007), and Sun (2011, 2013). The approximation approaches in some other papers can also be regarded as special cases of the fixed-smoothing asymptotics. This includes, among others, Ibragimov and Müller (2010), Shao (2010) and Bester, Hansen and Conley (2011).

All of the recent developments on the fixed-smoothing asymptotics have been devoted to first-step GMM estimation and inference. In this paper, we establish the fixed-smoothing asymptotics in a general two-step GMM framework. For two-step estimation and inference, the HAR variance estimator not only appears in the covariance estimator but also plays the role of the optimal weighting matrix in the second-step GMM criterion function. Under the fixed-smoothing asymptotics, the weighting matrix converges to a random matrix. As a result, the second-step GMM estimator is not asymptotically normal but rather asymptotically mixed normal. On one hand, the asymptotic mixed normality captures the estimation uncertainty of the GMM weighting matrix and is expected to be closer to the finite sample distribution of the second-step GMM estimator. On the other hand, the lack of asymptotic normality posts a challenge for pivotal inference. It is far from obvious that the Wald statistic is still asymptotically pivotal under the new asymptotics. To confront this challenge, we have to judicially rotate and transform the asymptotic distribution and show that it is equivalent to a distribution that is nuisance parameter free.

The fixed-smoothing asymptotic distribution not only depends on the kernel or basis function and the smoothing parameter, which is the same as in the one-step GMM framework, but also depends on the degree of over-identification, which is different from existing results. In general, the degree of over-identification or the number of moment conditions remains in the asymptotic distribution only under the so-called many-instruments or many-moments asymptotics. Here the number of moment conditions and hence the degree of over-identification are finite and fixed. Intuitively, the degree of over-identification remains in the asymptotic distribution because it is indicative of the dimension of the limiting random weighting matrix.

In the case of OS HAR variance estimation, the fixed-smoothing asymptotic distribution is a mixed noncentral F distribution — a noncentral F distribution with a random noncentrality parameter. This is an intriguing result, as a noncentral F distribution is not expected under the null hypothesis. It is reassuring that the random noncentrality parameter becomes degenerate as the amount of smoothing increases. Replacing the random noncentrality parameter by its mean, we show that the mixed noncentral F distribution can be approximated extremely well by a noncentral F distribution. The noncentral F-distribution is implemented in standard programming languages and packages such as the R language, MATLAB, Mathematica and STATA, so critical values are readily available. No computation intensive simulation is needed. This can be regarded as an advantage of using the OS HAR variance estimator.

In the case of kernel HAR variance estimation, the fixed-smoothing asymptotic distribution is a mixed chi-square distribution — a chi-square distribution scaled by an independent and positive random variable. This resembles an F distribution except that the random denominator has a complicated distribution. Nevertheless, we are able to show that critical values from the fixed-smoothing asymptotic distribution are high order correct under the conventional increasing-smoothing asymptotics. This result is established on the basis of two distributional expansions. The first expansion is the expansion of the fixed-smoothing asymptotic distribution as the amount of smoothing increases, and the second one is the high order Edgeworth expansion established in Sun and Phillips (2009). We arrive at the two expansions via completely different routes, yet at the end some of the terms in the two expansions are exactly the same.

Our framework for establishing the fixed-smoothing asymptotics is general enough to accommodate both the kernel HAR variance estimators and the OS HAR variance estimators. The fixed-smoothing asymptotics is established under weaker conditions than what are typically assumed in the literature. More specifically, instead of maintaining a functional CLT assumption, we make use of a regular CLT, which is weaker than an FCLT. Our method of proof is also novel. It applies directly to both smooth and nonsmooth kernel functions. There is no need to give a separate treatment to non-smooth kernels such as the Bartlett kernel. The unified method of proof leads to a unified representation of the fixed-smoothing asymptotic distribution.

The fixed-smoothing asymptotics is established for three commonly used test statistics in the GMM framework: the Wald statistic, the GMM criterion function statistic, and the score type statistic or the LM statistic. As in the conventional increasing-smoothing asymptotic framework, we show that these three test statistics are asymptotically equivalent and converge to the same limiting distribution.

In the Monte Carlo simulations, we examine the accuracy of the fixed-smoothing approximation. We find that the tests based on the new fixed-smoothing approximation have much more accurate size than the conventional chi-square tests. This is especially true when the degree of over-identification is large. When the model is over-identified, the fixed-smoothing approximation that accounts for the randomness of the GMM weighting matrix is also more accurate than the fixed-smoothing approximation that ignores the randomness. When the OS HAR variance estimator is used, the convenient noncentral F test has almost identical size properties as the nonstandard test whose critical values have to be simulated.

The rest of the paper is organized as follows. Section 2 describes the estimation and testing problems at hand. Section 3 establishes the fixed-smoothing asymptotics for the covariance estimators and the associated test statistics. Section 4 gives different representations of the fixed-smoothing asymptotic distribution. In the case of OS HAR variance estimation, a noncentral F distribution is shown to be a very accurate approximation to the nonstandard fixed-smoothing asymptotic distribution. In the case of kernel HAR variance estimation, the fixed-smoothing approximation is shown to provide a high order refinement over the chi-square approximation. The next section reports simulation evidence on the performance of the new approximation. The last section provides some concluding discussion. Proofs of the main results are given in the Appendix.

A word on notation: we use $\mathcal{F}_{p,K-p-q+1}\left(\delta^2\right)$ to denote a random variable that follows the noncentral F distribution with degrees of freedom $(p, K-p-q+1)$ and noncentrality parameter δ^2. We use $F_{p,K-p-q+1}\left(z,\delta^2\right)$ to denote the CDF of the noncentral F distribution.

2 Two-step GMM Estimation and Testing

We are interested in a $d \times 1$ vector of parameters $\theta \in \Theta \subseteq \mathbb{R}^d$. Let $v_t \in \mathbb{R}^{d_v}$ denote a vector of observations at time t. Let θ_0 denote the true value and assume that θ_0 is an interior point of Θ. The moment conditions

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

hold if and only if $\theta = \theta_0$ where $f\left(\cdot\right)$ is an $m \times 1$ vector of continuously differentiable functions. The process $f\left(v_t, \theta_0\right)$ may exhibit autocorrelation of unknown forms. We assume that $m \geq d$ and rank $E\left[\partial f\left(v_t, \theta_0\right)/\partial \theta'\right] = d$. That is, we consider a model that is possibly over-identified with the degree of over-identification $q = m - d$.

Define

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

then the GMM estimator of θ_0 is given by

$$\hat{\theta}_{GMM} = \arg\min_{\theta \in \Theta} g_T\left(\theta\right)' W_T^{-1} g_T\left(\theta\right)$$

where W_T is a positive definite weighting matrix.

To obtain an initial first step estimator, we often choose a simple weighting matrix W_o that does not depend on model parameters, leading to

$$\tilde{\theta}_T = \arg\min_{\theta \in \Theta} g_T\left(\theta\right)' W_o^{-1} g_T\left(\theta\right).$$

As an example, we may set $W_o = I_m$ in the general GMM setting. In the IV regression, we may set $W_o = Z'Z/T$ where Z is the data matrix for the instruments. We assume that

$$W_o \overset{p}{\to} W_{o,\infty},$$

a matrix that is positive definite almost surely.

According to Hansen (1982), the optimal weighting matrix W_T is the asymptotic variance matrix of $\sqrt{T} g_T\left(\theta_0\right)$. On the basis of the first step estimate $\tilde{\theta}_T$, we can use $\tilde{u}_t := f(v_t, \tilde{\theta}_T)$ to estimate the asymptotic variance matrix. Many nonparametric estimators of the variance matrix are available in the literature. In this paper, we consider a class of quadratic variance estimators, which includes the conventional kernel variance estimators of Andrews (1991), Newey and West (1987), Politis (2011), sharp and steep kernel variance estimators of Phillips, Sun and Jin (2006, 2007), and the orthonormal series (OS) variance estimators of Phillips (2005), Müller (2007), and Sun (2011, 2013) as special cases. Following Phillips, Sun and Jin (2006, 2007), we refer to the conventional kernel estimators as contracted kernel estimators and the sharp and steep kernel estimators as exponentiated kernel estimators.

The quadratic HAR variance estimator is given by

$$\tilde{W}_T\left(\tilde{\theta}_T\right) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q_h\left(\frac{t}{T}, \frac{s}{T}\right)\left(\tilde{u}_t - \frac{1}{T}\sum_{\tau=1}^{T}\tilde{u}_\tau\right)\left(\tilde{u}_s - \frac{1}{T}\sum_{\tau=1}^{T}\tilde{u}_\tau\right)' \tag{2}$$

where $Q_h\left(r, s\right)$ is a weighting function that depends on the smoothing parameter h. For conventional kernel estimators, $Q_h\left(r, s\right) = k\left(\left(r - s\right)/b\right)$ and we take $h = 1/b$. For the sharp kernel estimator, $Q_h(r,s) = (1-|r-s|)^{\rho} 1\{|r-s|<1\}$ and we take $h = \rho$. For steep quadratic kernel estimators, $Q_h(r,s) = k^{\rho}(r-s)$ and we take $h = \sqrt{\rho}$. For the OS estimators $Q_h(r,s) = K^{-1}\sum_{j=1}^{K}\phi_j(r)\phi_j(s)$ and we take $h = K$, where $\{\phi_j(r)\}$ are orthonormal basis functions on $L^2[0,1]$ satisfying $\int_0^1 \phi_j(r)\,dr = 0$. A prewhitened version of the above estimators can also be used. See, for example, Andrews and Monahan (1992) and Xiao and Linton (2002).

For our theoretical development below, we use an asymptotically equivalent estimator $W_T(\tilde{\theta}_T)$ defined by

$$W_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q_h^*\left(\frac{t}{T},\frac{s}{T}\right) f(v_t,\theta)\left[f(v_s,\theta)\right]' \tag{3}$$

where

$$Q_h^*(r,s) = Q_h(r,s) - \int_0^1 Q_h(\tau,s)\,d\tau - \int_0^1 Q_h(r,\tau)\,d\tau + \int_0^1\int_0^1 Q_h(\tau_1,\tau_2)\,d\tau_1 d\tau_2$$

is the centered version of $Q_h(r,s)$ satisfying $\int_0^1 Q_h^*(r,\tau)\,dr = \int_0^1 Q_h^*(\tau,s)\,ds = 0$ for any τ. For OS HAR variance estimators, centering is not necessary as $Q_h^*(r,s) = Q_h(r,s)$. With the variance estimator $W_T(\tilde{\theta}_T)$, the two-step GMM estimator is:

$$\hat{\theta}_T = \arg\min_{\theta\in\Theta} g_T(\theta)' W_T^{-1}(\tilde{\theta}_T) g_T(\theta).$$

Suppose we want to test the linear null hypothesis $H_0 : R\theta_0 = r$ against $H_0 : R\theta_0 \neq r$ where R is a $p\times d$ matrix with full row rank. Nonlinear restrictions can be converted to linear ones by the delta method. We consider three types of test statistics. The first type is the conventional Wald statistic. The normalized Wald statistic is

$$\mathbb{W}_T := \mathbb{W}_T(\hat{\theta}_T) = T(R\hat{\theta}_T - r)'\left\{R\left[G_T(\hat{\theta}_T)'W_T^{-1}(\hat{\theta}_T)G_T(\hat{\theta}_T)\right]^{-1}R'\right\}^{-1}(R\hat{\theta}_T - r)/p,$$

where

$$G_T(\theta) = \frac{\partial g_T(\theta)}{\partial\theta'}.$$

When $p = 1$ and for one-sided alternative hypotheses, we can construct the t statistic $t_T = t_T(\hat{\theta}_T)$ where $t_T(\theta_T)$ is defined to be

$$t_T(\theta_T) = \frac{\sqrt{T}(R\theta_T - r)}{\left\{R\left[G_T(\theta_T)'W_T^{-1}(\theta_T)G_T(\theta_T)\right]^{-1}R'\right\}^{1/2}}.$$

The second type of test statistic is based on the likelihood ratio principle. Let $\hat{\theta}_{T,R}$ be the restricted second-step GMM estimator:

$$\hat{\theta}_{T,R} = \arg\min_{\theta\in\Theta} g_T(\theta)' W_T^{-1}(\tilde{\theta}_T) g_T(\theta) \;\; s.t. \; R\theta = r.$$

The likelihood ratio principle suggests the GMM distance statistic (or GMM criterion function statistic) given by

$$\mathbb{D}_T := \left[T g_T(\hat{\theta}_T)'W_T^{-1}(\breve{\theta}_T)g_T(\hat{\theta}_T) - T g_T(\hat{\theta}_{T,R})'W_T^{-1}(\breve{\theta}_T)g_T(\hat{\theta}_{T,R})\right]/p$$

where $\breve{\theta}_T$ is a $\sqrt{T}$-consistent estimator of θ_0, i.e. $\breve{\theta}_T - \theta_0 = O_p(1/\sqrt{T})$. Typically, we take $\breve{\theta}_T$ to be $\tilde{\theta}_T$, the unrestricted first-step estimator. To ensure that $\mathbb{D}_T \geq 0$, we use the same $W_T^{-1}(\breve{\theta}_T)$ in computing the restricted and unrestricted GMM criterion functions.

The third type of test statistic is the GMM counterpart of the score statistic or Lagrange Multiplier (LM) statistic. It is based on the score or gradient of the GMM criterion function, i.e. $\Delta_T(\theta) = G_T'(\theta) W_T^{-1}(\breve{\theta}_T) g_T(\theta)$. The test statistic is given by

$$\mathbb{S}_T = T\left[\Delta_T(\hat{\theta}_{T,R})\right]'\left[G_T'(\hat{\theta}_{T,R})W_T^{-1}(\breve{\theta}_T)G_T(\hat{\theta}_{T,R})\right]^{-1}\Delta_T(\hat{\theta}_{T,R})/p$$

where as before $\breve{\theta}_T$ is a $\sqrt{T}$-consistent estimator of θ_0.

Under the usual asymptotics where $h \to \infty$ but at a slower rate than the sample size T, all three statistics $\mathbb{W}_T$, $\mathbb{D}_T$ and $\mathbb{S}_T$ are asymptotically distributed as χ_p^2/p, and the t-statistic t_T is asymptotically normal under the null. The question is, what are the limiting distributions of $\mathbb{W}_T$, $\mathbb{D}_T$, $\mathbb{S}_T$ and t_T when h is held fixed as $T \to \infty$? The rationale of considering this type of thought experiment is that it may deliver asymptotic approximations that are more accurate than the chi-square or normal approximation in finite samples.

When h is fixed, that is, b, ρ or K is fixed, the variance matrix estimator involves a fixed amount of smoothing in that it is approximately equal to an average of a fixed number of quantities from a frequency domain perspective. The fixed-b, fixed-ρ or fixed-K asymptotics may be collectively referred to as the fixed-smoothing asymptotics. Correspondingly, the conventional asymptotics under which $h \to \infty$, $T \to \infty$ jointly is referred to as the increasing-smoothing asymptotics. In view of the definition of h for each HAR variance estimator, the magnitude of h indicates the amount or level of smoothing in each case.

3 The Fixed-smoothing Asymptotics

3.1 Fixed-smoothing asymptotics for the variance estimator

To establish the fixed-smoothing asymptotics, we maintain Assumption 1 on the kernel function and basis functions.

Assumption 1 *(i) For kernel HAR variance estimators, the kernel function $k(\cdot)$ satisfies the following condition: for any $b \in (0,1]$ and $\rho \geq 1$, $k_b(x)$ and $k^\rho(x)$ are symmetric, continuous, piecewise monotonic, and piecewise continuously differentiable on $[-1,1]$. (ii) For the OS HAR variance estimator, the basis functions $\phi_j(\cdot)$ are piecewise monotonic, continuously differentiable and orthonormal in $L^2[0,1]$ and $\int_0^1 \phi_j(x)\,dx = 0$.*

Assumption 1 on the kernel function is very mild. It includes many commonly used kernel functions such as the Bartlett kernel, Parzen kernel, QS kernel, Daniel kernel and Tukey-Hanning kernel. Under Assumption 1, we can use the Fourier series expansion to show that $Q_h^*(r,s)$ has the following unified representation for all the HAR variance estimators we consider:

$$Q_h^*(r,s) = \sum_{j=1}^{\infty} \lambda_j \Phi_j(r)\,\Phi_j(s), \tag{4}$$

where $\{\Phi_j(r)\}$ is a sequence of continuously differentiable functions satisfying $\int_0^1 \Phi_j(r)\,dr = 0$. The right hand side of (4) converges absolutely and uniformly over $(r,s) \in [0,1] \times [0,1]$. For positive semi-definite kernels $k(\cdot)$, the above representation also follows from Mercer's theorem, in which case λ_j is the eigenvalue and $\Phi_j(r)$ is the corresponding orthonormal eigen function for the Fredholm operator with kernel $Q_h^*(r,s)$. Alternatively, the representation can be regarded as a spectral decomposition of this Fredholm operator, which can be shown to be compact. This representation enables us to give a unified proof for the contracted kernel HAR estimator, the exponentiated kernel HAR estimator and the OS HAR estimator. For kernel HAR estimation, there is no need to single out non-smooth kernels such as the Bartlett kernel and treat them differently as in KV (2005).

Define

$$G_t(\theta) = \frac{\partial g_t(\theta)}{\partial \theta'} = \frac{1}{T}\sum_{j=1}^{t} \frac{\partial f(v_j, \theta)}{\partial \theta'} \text{ for } t \geq 1 \text{ and } G_0(\theta) = 0.$$

Let $u_t = f(v_t, \theta_0)$ and

$$\Phi_0(t) = 1, \ e_t \sim iidN(0, I_m).$$

We make the following four assumptions on the GMM estimators and the data generating process.

Assumption 2 *$plim_{T\to\infty}\hat{\theta}_T = \theta_0$, $plim_{T\to\infty}\hat{\theta}_{T,R} = \theta_0$, $plim_{T\to\infty}\tilde{\theta}_T = \theta_0$ for an interior point $\theta_0 \in \Theta$.*

Assumption 3 *$\sum_{j=-\infty}^{\infty} \|\Gamma_j\| < \infty$ where $\Gamma_j = Eu_t u_{t-j}'$.*

Assumption 4 *For any $\theta_T = \theta_0 + o_p(1)$, $plim_{T\to\infty} G_{[rT]}(\theta_T) = rG$ uniformly in r where $G = G(\theta_0)$ and $G(\theta) = E\partial f(v_t, \theta)/\partial\theta'$.*

Assumption 5 *(i) $T^{-1/2}\sum_{t=1}^{T}\Phi_j(t/T)u_t$ converges weakly to a continuous distribution, jointly over $j = 0, 1, ..., J$ for every fixed J.*

(ii) The following holds:

$$P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)u_t \leq x \text{ for } j = 0, 1, ..., J\right)$$
$$= P\left(\Lambda\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)e_t \leq x \text{ for } j = 0, 1, ..., J\right) + o(1) \ \text{ as } T \to \infty$$

for every fixed J where $x \in \mathbb{R}^p$ and Λ is the matrix square root of Ω, i.e. $\Lambda\Lambda' = \Omega := \sum_{j=-\infty}^{\infty}\Gamma_j$.

Assumptions 2–4 are standard assumptions in the literature on the fixed-smoothing asymptotics. They are the same as those in Kiefer and Vogelsang (2005), Sun and Kim (2012a), among others. Assumption 5 is a variant of the standard multivariate CLT. If Assumption 1 holds and

$$T^{1/2}g_{[rT]}(\theta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[rT]}u_t \xrightarrow{d} \Lambda B_m(r)$$

where $B_m(r)$ is a standard Brownian motion, then Assumption 5 holds. So Assumption 5 is weaker than the above FCLT, which is typically assumed in the literature on the fixed-smoothing asymptotics.

A great advantage of maintaining only the CLT assumption is that our results can be easily generalized to the case of higher dimensional dependence, such as spatial dependence and spatial-temporal dependence. In fact, Sun and Kim (2012b) have established the fixed-smoothing asymptotics in the spatial setting using only the CLT assumption. They avoid the more restrictive FCLT assumption maintained in BCHV (2011). With some minor notational change and a small change in Assumption 4 as in Sun and Kim (2012b), our results remain valid in the spatial setting.

Assumption 5 only assumes that the approximation error is $o(1)$, which is enough for our first order fixed-smoothing asymptotics. However, it is useful to discuss the composition of the approximation error. Let $\Psi_J(t) = (\Phi_0(t), \Phi_1(t), ..., \Phi_J(t))'$ and $a_J \in \mathbb{R}^{J+1}$ be a vector of the same dimension. It follows from Lemma 1 in Taniguchi and Puri (1996) that under some additional conditions:

$$P\left(a_J' \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Psi_J\left(\frac{t}{T}\right) u_t \leq x\right) = P\left(a_J' \Lambda \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Psi_J\left(\frac{t}{T}\right) e_t \leq x\right) + \frac{c(x)}{\sqrt{T}} + O\left(\frac{1}{T}\right)$$

for some function $c(x)$. In the above Edgeworth expansion, the approximation error of order $c(x)/\sqrt{T}$ captures the skewness of u_t. When u_t is Gaussian or has a symmetric distribution, this term disappears. Part of the approximation error of order $O(1/T)$ comes from the stochastic dependence of u_t. If we replace the iid Gaussian process $\{\Lambda e_t\}$ by a dependent Gaussian process $\{u_t^N\}$ that has the same covariance structure as $\{u_t\}$, then we can remove this part of the approximation error.

To present Assumption 5 more compactly, we introduce the notion of asymptotic equivalence in distribution. Consider two stochastically bounded sequences of random vectors $\xi_T \in \mathbb{R}^p$ and $\eta_T \in \mathbb{R}^p$, we say that they are asymptotically equivalent in distribution and write $\xi_T \overset{a}{\sim} \eta_T$ if and only if $Ef(\xi_T) = Ef(\eta_T) + o(1)$ as $T \to \infty$ for all bounded and continuous functions $f(\cdot)$ on $\mathbb{R}^p$. According to Lemma 3 in the appendix, Assumption 5 is equivalent to

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_j\left(\frac{t}{T}\right) u_t \overset{a}{\sim} \Lambda \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_j\left(\frac{t}{T}\right) e_t$$

jointly over $j = 0, 1, ..., J$.

Let

$$\Omega_j(\theta) = \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_j\left(\frac{t}{T}\right) f(v_t, \theta)\right] \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_j\left(\frac{t}{T}\right) f(v_t, \theta)\right]'.$$

Using the uniform series representation in (4), we can write

$$W_T(\theta) = \sum_{j=1}^{\infty} \lambda_j \Omega_j(\theta),$$

which is an infinite weighted sum of outer-products. To establish the fixed-smoothing asymptotic distribution of $W_T(\theta)$ under Assumption 5, we split $W_T(\theta)$ into a finite sum part and the remainder part: $W_T(\theta) = \sum_{j=1}^{J} \lambda_j \Omega_j(\theta) + \sum_{j=J+1}^{\infty} \lambda_j \Omega_j(\theta)$. For each fixed J, we can use Assumptions 2–5 to obtain the asymptotically equivalent distribution for the first term. However, for the second term to vanish, we require $J \to \infty$. To close the gap in our argument, we use Lemma 1 below, which is similar to Lemma 2 in Kim and Sun (2012b). The lemma puts our proof on a rigorous footing and may be of independent interest.

Lemma 1 *Suppose $\omega_T = \xi_{T,J} + \eta_{T,J}$ and ω_T does not depend on J. Assume that there exist $\xi^*_{T,J}$ and ξ_T such that*

(i) $P\left(\xi_{T,J} < \xi\right) - P\left(\xi^*_{T,J} < \xi\right) = o(1)$ for each fixed J and each $\xi \in \mathbb{R}$ as $T \to \infty$,

(ii) $P\left(\xi^*_{T,J} < \xi\right) - P\left(\xi_T < \xi\right) = o(1)$ uniformly over sufficiently large T for each $\xi \in \mathbb{R}$ as $J \to \infty$,

(iii) the CDF of ξ_T is equicontinuous on $\mathbb{R}$ when T is sufficiently large, and

(iv) $\eta_{T,J} \xrightarrow{p} 0$ uniformly in T as $J \to \infty$. Then

$$P\left(\omega_T < \xi\right) = P\left(\xi_T < \xi\right) + o(1) \text{ for each } \xi \in \mathbb{R} \text{ as } T \to \infty.$$

Lemma 1 allows us to approximate the distribution of ω_T by that of ξ_T. Using this lemma, we can prove the lemma below.

Lemma 2 *Let Assumptions 1-5 hold, then for any $\sqrt{T}$-consistent estimator $\breve{\theta}_T$ and for a fixed h,*

(a) $W_T(\breve{\theta}_T) = W_T\left(\theta_0\right) + o_p\left(1\right);$

(b) $W_T(\breve{\theta}_T) \overset{a}{\sim} W_{eT} := \Lambda T^{-1} \sum_{t=1}^{T} \sum_{\tau=1}^{T} Q^*_h\left(t/T, \tau/T\right) e_t e'_\tau \Lambda';$

(c) $W_T(\breve{\theta}_T) \xrightarrow{d} W_\infty$ with $W_\infty = \Lambda \tilde{W}_\infty \Lambda'$ where

$$\tilde{W}_\infty = \int_0^1 \int_0^1 Q^*_h(r,s) dB_m(r) dB_m(s)'.$$

Lemmas 2(a) and (b) show that the estimation uncertainty in $\breve{\theta}_T$ does not matter asymptotically. It is well known that this type of result holds under the conventional increasing-smoothing asymptotics. The same result holds under the fixed-smoothing asymptotics as long as the weighting function has been centered. Lemma 2(c) gives a unified representation of the limiting distribution. The presentation applies to smooth kernels as well as nonsmooth kernels such as the Bartlett kernel.

Although we do not make the FCLT assumption, the limiting distribution of $W_T(\breve{\theta}_T)$ can still be represented by a functional of Brownian motion as in Lemma 2(c). This representation serves two purposes. First, it gives an explicit representation of the limiting distribution. It is standard practice to obtain the limiting distribution in order to conduct asymptotically valid inference. However, this is not necessary, as we can simply use the asymptotically equivalent distribution in Lemma 2(b). Second, the representation in Lemma 2(c) enables us to compare the results we obtain here with existing results. It is reassuring that the limiting distribution W_∞ is the same as that obtained by Kiefer and Vogelsang (2005), Phillips, Sun and Jin (2006, 2007), and Sun and Kim (2012a), among others.

3.2 Fixed-smoothing asymptotics for the test statistics

We now establish the asymptotic distributions of $\mathbb{W}_T, \mathbb{D}_T, \mathbb{S}_T$ and t_T when h is fixed. For kernel covariance estimators, we focus on the case that $k\left(\cdot\right)$ is positive definite, as otherwise the two-step estimator may not even be consistent. In this case, we can show that, for all the covariance estimators we consider, W_∞ is nonsingular with probability one. Hence, using Lemma 2 and Lemma 4 in the appendix, we have $W_T(\tilde{\theta}_T)^{-1} \overset{a}{\sim} W_{eT}^{-1}$. Using this result and Assumptions 1-5, we have

$$\sqrt{T}\left(\hat{\theta}_T-\theta_0\right) = -\left[G'W_T(\tilde{\theta}_T)^{-1}G\right]^{-1}G'W_T(\tilde{\theta}_T)^{-1}\sqrt{T}g_T\left(\theta_0\right)+o_p\left(1\right)$$
$$\overset{a}{\sim} -\left[G'W_{eT}^{-1}G\right]^{-1}G'W_{eT}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Lambda e_t \tag{5}$$
$$\overset{d}{\to} -\left[G'W_{\infty}^{-1}G\right]^{-1}G'W_{\infty}^{-1}\Lambda B_m\left(1\right).$$

So $\hat{\theta}_T$ is not asymptotically normal but rather asymptotically mixed normal. More specifically, W_{∞}^{-1} is independent of $B_m(1)$, so conditional on W_{∞}^{-1}, the limiting distribution is normal with the conditional variance depending on W_{∞}^{-1}. Due to the asymptotic mixed-normality, it is not straightforward to derive the fixed-smoothing asymptotics of the test statistics $\mathbb{W}_T, \mathbb{D}_T, \mathbb{S}_T$ and t_T.

Using (5) and Lemma 4 in the appendix, we have $\mathbb{W}_T \overset{a}{\sim} F_{eT}$ and $\mathbb{W}_T \overset{d}{\to} F_{\infty}$ where

$$F_{eT} = \left[R\left(G'W_{eT}^{-1}G\right)^{-1}G'W_{eT}^{-1}\Lambda\frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_t\right]'\left[R\left(G'W_{eT}^{-1}G\right)^{-1}R'\right]^{-1}$$
$$\times\left[R\left(G'W_{eT}^{-1}G\right)^{-1}G'W_{eT}^{-1}\Lambda\frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_t\right]/p,$$

and

$$F_{\infty} = \left[R\left(G'W_{\infty}^{-1}G\right)^{-1}G'W_{\infty}^{-1}\Lambda B_m(1)\right]'\left[R\left(G'W_{\infty}^{-1}G\right)^{-1}R'\right]^{-1}$$
$$\times\left[R\left(G'W_{\infty}^{-1}G\right)^{-1}G'W_{\infty}^{-1}\Lambda B_m(1)\right]/p.$$

To show that F_{eT} and F_{∞} are pivotal, we let $e_t := (e_{t,p}', e_{t,d-p}', e_{t,q}')'$. The subscripts p, $d-p$, and q on e indicate not only the dimensions of the random vectors and but also distinguish them so that, for example, $e_{t,p}$ is different and independent from $e_{t,q}$ for all values of p and q. Denote

$$C_{p,T} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_{t,p},\ C_{q,T} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_{t,q}$$

and

$$C_{pp,T} = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}Q_h^*(\frac{t}{T},\frac{\tau}{T})e_{t,p}e_{\tau,p}',\ \ C_{pq,T} = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}Q_h^*(\frac{t}{T},\frac{\tau}{T})e_{t,p}e_{\tau,q}' \tag{6}$$
$$C_{qq,T} = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}Q_h^*(\frac{t}{T},\frac{\tau}{T})e_{t,q}e_{\tau,q}',\ \ D_{pp,T} = C_{pp,T}-C_{pq,T}C_{qq,T}^{-1}C_{pq,T}'.$$

Similarly let

$$B_m(r) := \left(B_p'(r),\ B_{d-p}'(r),\ B_q'(r)\right)'$$

where $B_p(r)$, $B_{d-p}(r)$ and $B_q(r)$ are independent standard Brownian motion processes of dimensions p, $d-p$, and q, respectively. Denote

$$C_{pp} = \int_0^1\int_0^1 Q_h^*(r,s)dB_p(r)dB_p(s)',\ \ C_{pq} = \int_0^1\int_0^1 Q_h^*(r,s)dB_p(r)dB_q(s)' \tag{7}$$
$$C_{qq} = \int_0^1\int_0^1 Q_h^*(r,s)dB_q(r)dB_q(s)',\ \ D_{pp} = C_{pp}-C_{pq}C_{qq}^{-1}C_{pq}'.$$

Then with ingenious use of some theory of multivariate statistics, we obtain the following theorem.

Theorem 1 *Let Assumptions 1-5 hold. Assume that $k(\cdot)$ used in the kernel HAR variance estimation is positive definite. Then, for a fixed h,*

(a) $\mathbb{W}_T(\hat{\theta}_T) \overset{a}{\sim} F_T := \left[C_{p,T} - C_{pq,T}C_{qq,T}^{-1}C_{q,T}\right]' D_{pp,T}^{-1}\left[C_{p,T} - C_{pq,T}C_{qq,T}^{-1}C_{q,T}\right]/p,$

(b) $\mathbb{W}_T(\hat{\theta}_T) \overset{d}{\to} F_\infty := \left[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right]' D_{pp}^{-1}\left[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right]/p,$

(c) $t_T(\hat{\theta}_T) \overset{a}{\sim} \left[C_{p,T} - C_{pq,T}C_{qq,T}^{-1}C_{q,T}\right]/\sqrt{D_{pp,T}},$

(d) $t_T(\hat{\theta}_T) \overset{d}{\to} t_\infty := \left[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right]/\sqrt{D_{pp}},$

where $(C_{p,T}', C_{q,T}')'$ is independent of $C_{pq,T}C_{qq,T}^{-1}$ and $D_{pp,T}$, and $\left(B_p(1)', B_q(1)'\right)'$ is independent of $C_{pq}C_{qq}^{-1}$ and D_{pp}.

Remark 1 *Theorem 1 shows that both the asymptotically equivalent distribution F_T and the limiting distribution F_∞ are pivotal. In particular, they do not depend on the long run variance Ω. Both F_T and F_∞ are nonstandard but can be simulated. It is easier to simulate F_T, as it involves only T iid standard normal vectors.*

Remark 2 *Let $\mathbb{W}_T(\tilde{\theta}_T)$ be the Wald statistic based on the first-step GMM estimator $\tilde{\theta}_T$. $\mathbb{W}_T(\tilde{\theta}_T)$ is constructed in the same way as $\mathbb{W}_T(\hat{\theta}_T)$. It is easy to see that $\mathbb{W}_T(\tilde{\theta}_T) \overset{d}{\to} B_p(1)' C_{pp}^{-1}B_p(1)/p$. See KV (2005) for contracted kernel HAR variance estimators, Phillips, Sun and Jin (2006,2007) for exponentiated kernel HAR variance estimators, and Sun (2013) for OS HAR variance estimators. Comparing the limit $B_p(1)' C_{pp}^{-1}B_p(1)/p$ with F_∞ given in Theorem 1, we can see that F_∞ contains additional adjustment terms.*

Remark 3 *When the model is exactly identified, we have $q = 0$. In this case,*

$$F_\infty = B_p(1)' C_{pp}^{-1}B_p(1)/p.$$

This limit is the same as that in the one-step GMM framework. This is not surprising, as when the model is exactly identified, the GMM weighting matrix becomes irrelevant and the two-step estimator is numerically identical to the one-step estimator.

Remark 4 *It is interesting to see that the limiting distribution F_∞ (and the asymptotically equivalent distribution) depends on the degree of over-identification. Typically such a dependence shows up only when the number of moment conditions is assumed to grow with the sample size. Here the number of moment conditions is fixed. It remains in the limiting distribution because it contains information on the dimension of the random limiting matrix $\tilde{W}_\infty$.*

Theorem 2 *Let Assumptions 1-5 hold. Then, for a fixed h,*

$$\mathbb{D}_T = \mathbb{W}_T + o_p(1) \text{ and } \mathbb{S}_T = \mathbb{W}_T + o_p(1).$$

It follows from Theorem 2 that all three statistics are asymptotically equivalent in distribution to the same distribution F_T defined in Theorem 1(a). They also have the same limiting distribution F_∞. So the asymptotic equivalence of the three statistics under the increasing-smoothing asymptotics remains valid under the fixed-smoothing asymptotics.

Careful inspection of the proofs of Theorems 1 and 2 shows that the two theorems hold regardless of which $\sqrt{T}$-consistent estimator of θ_0 we use in estimating $W_T(\theta_0)$ and $G_T(\theta_0)$ in $\mathbb{D}_T$ and $\mathbb{S}_T$. However, it may make a difference to high orders. We leave this for future research.

4 Representation and Approximation of the Fixed-smoothing Asymptotic Distribution

In this section, we establish different representations of the fixed-smoothing limiting distribution. These representations highlight the connection and difference between the nonstandard approximation and the standard χ^2 or normal approximation.

4.1 The case of series HAR variance estimation

To obtain a new representation of F_∞, we consider the matrix

$$K\begin{bmatrix} C_{pp} & C_{pq} \\ C_{pq}' & C_{qq} \end{bmatrix} \overset{d}{=} \int_0^1\int_0^1 Q_h^*(r,s)\,dB_{p+q}(r)\,dB_{p+q}'(s)$$
$$= \sum_{j=1}^K\left[\int_0^1 \Phi_j(r)\,dB_{p+q}(r)\right]\left[\int_0^1 \Phi_j(r)\,dB_{p+q}(r)\right]'$$

where $B_{p+q}(r)$ is the standard Brownian motion of dimension $p+q$. Since $\{\Phi_j(r)\}$ are orthonormal, $\int_0^1 \Phi_j(r)\,dB_{p+q}(r) \overset{d}{\sim} iidN(0, I_{p+q})$ and the above matrix follows a standard Wishart distribution $\mathcal{W}_{p+q}(K, I_{p+q})$. A well known property of a Wishart random matrix is that $D_{pp} = C_{pp} - C_{pq}C_{qq}^{-1}C_{pq}' \sim \mathcal{W}_p(K-q, I_p)$ and is independent of both C_{pq} and C_{qq}. This implies that D_{pp} is independent of $B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)$, and that F_∞ is equal in distribution to a quadratic form with an independent Wishart matrix as the (inverse) weighting matrix.

This brings F_∞ close to Hotelling's T^2 distribution, which is the same as a standard F distribution after some multiplicative adjustment. The only difference is that $B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)$ is not normal and hence $\left\|B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right\|^2$ does not follow a chi-square distribution. However, conditional on $C_{pq}C_{qq}^{-1}B_q(1)$, $||B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)||^2$ follows a noncentral χ_p^2 distribution with noncentrality parameter

$$\Delta^2 = ||C_{pq}C_{qq}^{-1}B_q(1)||^2.$$

It then follows that F_∞ is conditionally distributed as a noncentral F distribution. Let

$$\kappa = \frac{K}{(K-p-q+1)},\ \delta^2 = E\Delta^2 = \frac{pq}{K-q-1}.$$

The following theorem presents this result formally.

Theorem 3 *For series HAR variance estimation, we have*

(a) $\kappa^{-1}F_\infty \overset{d}{=} \mathcal{F}_{p,K-p-q+1}(\Delta^2)$, *a mixed noncentral F random variable with random noncentrality parameter* Δ^2;

(b) $P(\kappa^{-1}F_\infty < z) = F_{p,K-p-q+1}(z, \delta^2) + o(K^{-1})$ *as* $K \to \infty$;

(c) $P(pF_\infty < z) = \mathcal{G}_p(z) - \mathcal{G}_p'(z)\,z\left(\frac{p+2q-1}{K}\right) + \mathcal{G}_p''(z)\,z^2\frac{1}{K} + o\left(\frac{1}{K}\right)$ *as* $K \to \infty$, *where* $\mathcal{G}_p(z)$ *is the CDF of the chi-square distribution* χ_p^2.

Remark 5 *According to part (a),*

$$F_\infty \overset{d}{=} \frac{\chi_p^2(\Delta^2)/p}{\chi_{K-p-q+1}^2/K}$$

where $\chi_p^2(\Delta^2)$ *is independent of* $\chi^2_{K-p-q+1}$*. It is easy to see that*

$$\chi_p^2\left(\Delta^2\right) \overset{d}{=} ||C_{p\times 1} - C_{p\times K}C'_{q\times K}\left(C_{q\times K}C'_{q\times K}\right)^{-1}C_{q\times 1}||^2$$

where each $C_{m\times n}$ *is an* $m\times n$ *matrix (vector) with iid standard normal elements and* $C_{p\times 1}, C_{p\times K}, C_{q\times K}$ *and* $C_{q\times 1}$ *are mutually independent. This provides a simple way to simulate* F_∞*.*

Remark 6 *Parts (a) and (b) of Theorem 3 are intriguing in that the limiting distribution under the null is approximated by a noncentral F distribution, although the noncentrality parameter goes to zero as* $K\to\infty$*.*

Remark 7 *When* $q=0$*, that is, when the model is just identified, we have* $\Delta^2=\delta^2=0$ *and so*

$$\frac{K-p+1}{K}F_\infty \overset{d}{=} \mathcal{F}_{p,K-p+1}.$$

This result is the same as that obtained by Sun (2013). So an advantage of using orthonormal series HAR variance estimator is that the fixed-smoothing asymptotics is exactly a standard F distribution for just identified models.

Remark 8 *The asymptotics obtained under the specification that* K *is fixed and then letting* $K\to\infty$ *is a sequential asymptotics. As* $K\to\infty$*, we may show that*

$$\kappa^{-1}F_\infty = \frac{\chi_p^2/p}{\chi^2_{K-p-q+1}/\left(K-p-q+1\right)} + o_p(1) = \chi_p^2/p + o_p(1).$$

To the first order, the fixed-smoothing asymptotic distribution reduces to the distribution of χ_p^2/p*. As a result, if first order asymptotics are used in both steps in the sequential asymptotic theory, then the sequential asymptotic distribution is the same as the conventional joint asymptotic distribution. However, Theorem 3 is not based on the first order asymptotics but rather a high order asymptotics. The high order sequential asymptotics can be regarded as a convenient way to obtain an asymptotic approximation that better reflects the finite sample distribution of the test statistic* $\mathbb{D}_T$*,* $\mathbb{S}_T$ *or* $\mathbb{W}_T$*.*

Remark 9 *Instead of approximations via asymptotic expansions, all three types of approximations are distributional approximations. More specifically, the finite sample distributions of* $\mathbb{D}_T$*,* $\mathbb{S}_T$*, and* $\mathbb{W}_T$ *are approximated by the distributions of*

$$\frac{\chi_p^2}{p},\ \frac{\chi_p^2\left(\Delta^2\right)/p}{\chi^2_{K-p-q+1}/K},\ \text{and}\ \frac{\chi_p^2\left(\delta^2\right)/p}{\chi^2_{K-p-q+1}/K} \tag{8}$$

respectively under the conventional joint asymptotics, the fixed-smoothing asymptotics, and the higher order sequential asymptotics. An advantage of using distributional approximations is that they hold uniformly over their supports (by Pólya's Lemma).

Remark 10 *Theorem 3(c) gives a first order distributional expansion of* F_∞*. It is clear that the difference between* pF_∞ *and* χ_p^2 *depends on* K, p *and* q*. Let* $\chi_p^{1-\alpha} = \mathcal{G}_p^{-1}(1-\alpha)$ *be the* $(1-\alpha)$ *quantile of* $\mathcal{G}_p(z)$*, then*

$$P\left(pF_\infty \geq \chi_p^{1-\alpha}\right) = \alpha + \mathcal{G}'_p\left(\chi_p^{1-\alpha}\right)\chi_p^{1-\alpha}\left(\frac{p+2q-1}{K}\right) - \mathcal{G}''_p\left(\chi_p^{1-\alpha}\right)\left(\chi_p^{1-\alpha}\right)^2\frac{1}{K} + o\left(\frac{1}{K}\right).$$

For a typical critical value $\chi_p^{1-\alpha}$, $\mathcal{G}_p'(\chi_p^{1-\alpha}) > 0$ *and* $\mathcal{G}_p''(\chi_p^{1-\alpha}) < 0$, *so we expect* $P(F_\infty \geq \chi_p^{1-\alpha}/p) > \alpha$, *at least when* K *is large. So the critical value from* F_∞ *is expected to be larger than* $\chi_p^{1-\alpha}/p$. *For given* p *and large* K, *the difference between* $P(F_\infty \geq \chi_p^{1-\alpha}/p)$ *and* α *increases with* q, *the degree of over-identification. A practical implication of Theorem 3(c) is that when the degree of over-identification is large, using the chi-square critical value rather than the critical value from* F_∞ *may lead to the finding of a statistically significant relationship that does not actually exist.*

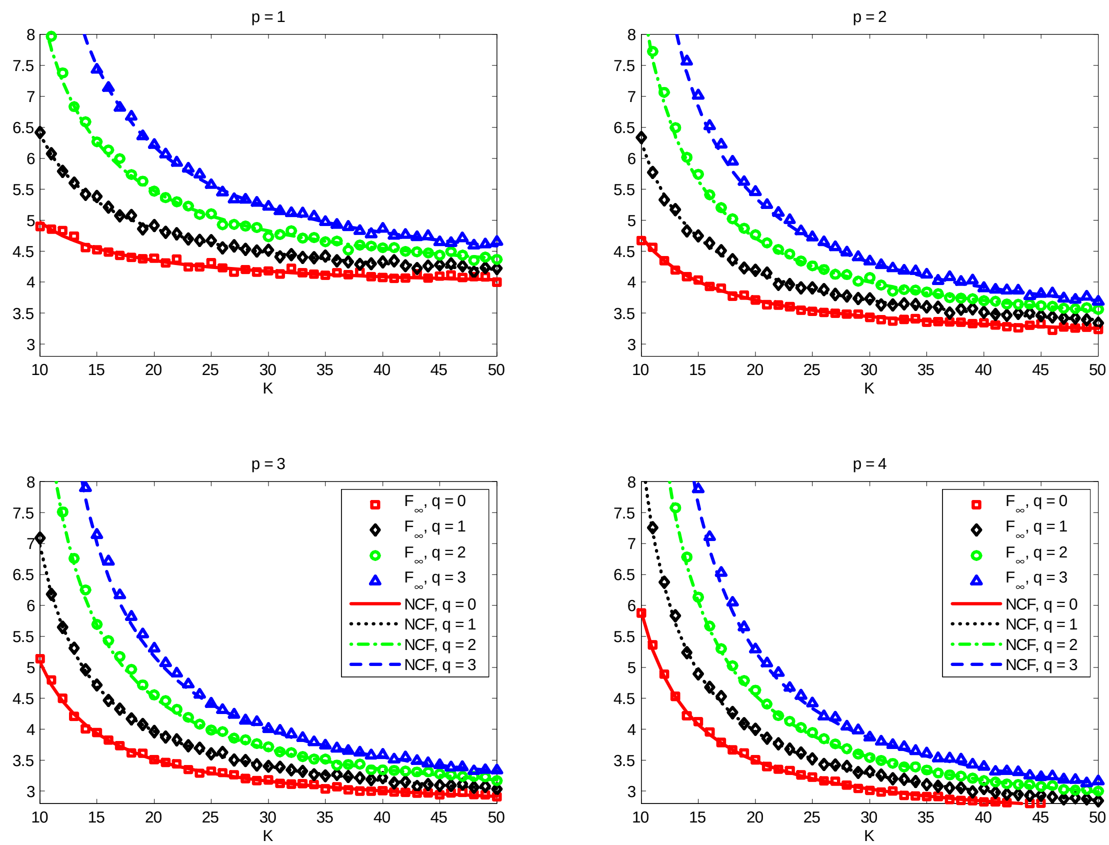

Figure 1: 95% quantitle of the nonstandard distribution and its noncentral F approximation

Figure 1 reports 95% critical values from the two approximations: the noncentral (scaled) $F_{p,K-p-q+1}\left(\cdot,\delta^2\right)$ approximation and the nonstandard F_∞ approximation, for different combinations of p and q. The nonstandard F_∞ distribution is simulated according to the representation in Remark 5 and the nonstandard critical values are based on 10000 simulation replications. Since the noncentral F distribution appears naturally in power analysis, it has been implemented in standard statistical and programming packages. So critical values from the noncentral F distribution can be obtained easily. Let $NCF^{1-\alpha} := \mathcal{F}^{1-\alpha}_{p,K-p-q+1}\left(\delta^2\right)$ be the $(1-\alpha)$ quantile of the noncentral $F_{p,K-p-q+1}\left(\cdot,\delta^2\right)$ distribution, then $\kappa NCF^{1-\alpha}$ corresponds to $\mathcal{F}_\infty^{1-\alpha}$, the $(1-\alpha)$ quantile of the nonstandard F_∞ distribution. Figure 1 graphs $\kappa NCF^{1-\alpha}$ and $\mathcal{F}_\infty^{1-\alpha}$ against different K values. The figure exhibits several patterns, all of which are consistent with Theorem 3(c). First, the noncentral F critical values are remarkably close to the nonstandard critical values. So approximating Δ^2 by its mean δ^2 does not incur much approximation error. Second, the noncentral F and nonstandard critical values increase with the degree of over-identification q. The increase is more significant when K is smaller. This is expected as both the noncentrality parameter δ^2 and the multiplicative factor κ increase with q. In addition, as q increases, the denominators in the noncentral F and nonstandard distributions become more random. All these effects shift the probability mass to the right as q increases. Third, for any given p and q combination, the noncentral F and nonstandard critical values decrease monotonically as K increases and approach the corresponding critical values from the chi-square distribution.

In parallel with Theorem 3, we can represent and approximate t_∞ as follows:

Theorem 4 *Let $\kappa = K/(K-q)$. For series HAR variance estimation, we have*

(a) $t_\infty/\sqrt{\kappa} \overset{d}{=} t_{K-q}(\Delta)$, a mixed noncentral t random variable with random noncentrality parameter $\Delta = C_{1q}C_{qq}^{-1}B_q(1)$;

(b) $P(t_\infty/\sqrt{\kappa} < z) = \frac{1}{2} + sgn(z)\frac{1}{2}F_{1,K-q}\left(z^2, \delta^2\right) + o\left(K^{-1}\right)$ as $K \to \infty$, where $\delta^2 = \frac{q}{K-q-1}$ and $sgn(\cdot)$ is the sign function;

(c) $P(t_\infty < z) = \mathbf{\Phi}(z) - |z|\boldsymbol{\phi}(z)\left[z^2 + (4q+1)\right]/(4K) + o\left(K^{-1}\right)$ as $K \to \infty$, where $\mathbf{\Phi}(z)$ and $\boldsymbol{\phi}(z)$ are the CDF and pdf of the standard normal distribution.

Theorem 4(a) is similar to Theorem 3(a). With a scale adjustment, the fixed-smoothing asymptotic distribution of the t statistic is a mixed noncentral t distribution. The random noncentrality parameter can be simulated according to $\Delta \overset{d}{=} C_{1\times K}C'_{q\times K}(C_{q\times K}C'_{q\times K})^{-1}C_{q\times 1}$ where each matrix $C_{m\times n}$ is an $m \times n$ matrix (vector) with iid standard normal elements and $C_{1\times K}, C_{q\times K}$ and $C_{q\times 1}$ are mutually independent. So the fixed-smoothing asymptotic distribution t_∞ can be simulated easily.

According to Theorem 4(b), we can approximate the quantile of t_∞ by that of a noncentral F distribution. More specifically, let $t_\infty^{1-\alpha}$ be the $(1-\alpha)$ quantile of t_∞, then

$$t_\infty^{1-\alpha} \doteq \begin{cases} \sqrt{\kappa \mathcal{F}_{1,K-q}^{1-2\alpha}(\delta^2)}, & \alpha < 0.5 \\ -\sqrt{\kappa \mathcal{F}_{1,K-q}^{2\alpha-1}(\delta^2)}, & \alpha \geq 0.5 \end{cases}$$

where $\mathcal{F}_{1,K-q}^{1-2\alpha}(\delta^2)$ is the $(1-2\alpha)$ quantile of the noncentral F distribution. As in the case of quantile approximation for F_∞, the above approximation is remarkably accurate. Figure 2, which is similar to Figure 1, illustrates this. The figure graphs $t_\infty^{1-\alpha}$ and its approximation against the values of K for different degrees of over-identification and for $\alpha = 0.05$ and 0.95. As K increases, the quantiles approach the normal quantiles ± 1.645. However when K is small or q is large, there is a significant difference between the normal quantiles and the corresponding quantiles from t_∞. This is consistent with Theorem 4(c). For a given small K, the absolute difference increases with the degree of over-identification q. For a given q, the absolute difference decreases with K.

Figure 2 and Theorem 4(c) suggest that the quantile of t_∞ is larger than the corresponding normal quantile in absolute value. So the test based on the normal approximation rejects the null more often than the test based on the fixed-smoothing approximation. This provides an explanation of the large size distortion of the asymptotic normal test.

4.2 The case of kernel HAR variance estimation

In the case of kernel HAR variance estimation, $D_{pp} = C_{pp} - C_{pq}C_{qq}^{-1}C'_{pq}$ is not independent of C_{pq} or C_{qq}. It is not as easy to simplify the nonstandard distribution as in the case of series HAR variance estimation. Different proofs are needed.

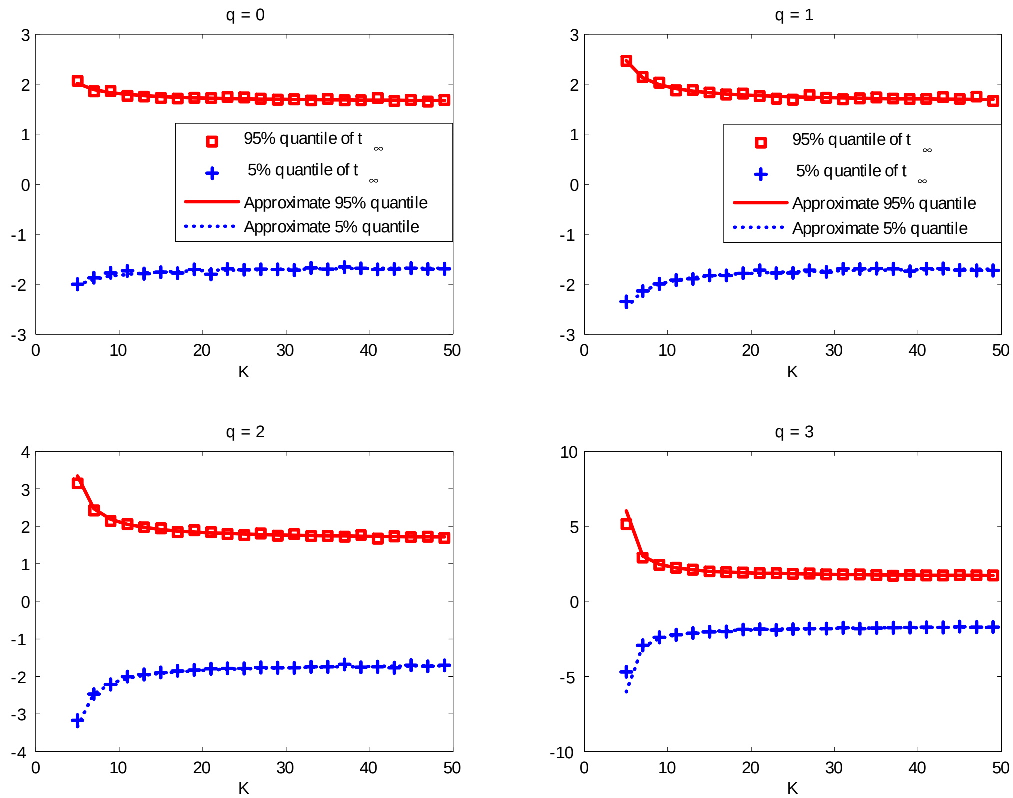

Figure 2: 5% and 95% quantitles of the nonstandard distribution t_∞ and their noncentral F approximation

Let

$$\mu_1 = \int_0^1 Q_h^*(r,r)\, dr \text{ and } \mu_2 = \int_0^1 \int_0^1 \left[Q_h^*(r,s)\right]^2 drds.$$

In Lemma 5 in the appendix, we show that $\mu_1 - 1 \asymp 1/h$ and $\mu_2 \asymp 1/h$ where "$a \asymp b$" indicates that a and b are of the same order of magnitude.

Theorem 5 *For kernel HAR variance estimation, we have*

(a) $pF_\infty \overset{d}{=} \chi_p^2/\eta^2$ *where* χ_p^2 *is independent of* η^2*,*

$$\eta^2 \overset{d}{=} \frac{e_p'\left[I_p + C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'\right]e_p}{e_p'\left[I_p + C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'\right]\left[C_{pp} - C_{pq}C_{qq}^{-1}C_{pq}'\right]^{-1}\left[I_p + C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'\right]e_p}$$

and $e_p = (1, 0, ..., 0)' \in \mathbb{R}^p$.

(b) As $h \to \infty$*, we have* $\eta^2 \overset{p}{\to} 1$ *and*

$$P\left(pF_\infty < z\right) = \mathcal{G}_p(z) + \mathcal{G}_p'(z)\, z\left[(\mu_1 - 1) - \frac{\mu_2}{\mu_1}(p + 2q - 1)\right] + \mathcal{G}_p''(z)\, z^2\mu_2 + o(\mu_2)$$

where as before $\mathcal{G}_p(z)$ *is the CDF of the chi-square distribution* χ_p^2*.*

Remark 11 *Theorem 5(a) shows that pF_∞ follows a scale-mixed chi-square distribution. Since $\eta^2 \to^p 1$ as $h \to \infty$, the sequential limit of $p\mathbb{W}_T$ is the usual chi-square distribution. The result is the same as in the series HAR variance estimation. The virtue of Theorem 5(a) is that it gives an explicit characterization of the random scaling factor η^2.*

Remark 12 *Using Theorem 5(a), we have $P(pF_\infty < z) = E\mathcal{G}_p(z\eta^2)$. Theorem 5(b) follows by taking a Taylor expansion of $\mathcal{G}_p(z\eta^2)$ around $\mathcal{G}_p(z)$ and approximating the moments of η^2. In the case of the contracted kernel HAR variance estimation, Sun (2012) establishes an expansion of the asymptotic distribution for $F_T(\tilde{\theta}_T)$, which is based on the one-step GMM estimator $\tilde{\theta}_T$. Using the notation in this paper, it is shown that*

$$P\left(B_p(1)' C_{pp}^{-1} B_p(1) < z\right) = \mathcal{G}_p(z) + \mathcal{G}_p'(z) z \left[(\mu_1 - 1) - \frac{\mu_2}{\mu_1}(p-1)\right] + \mathcal{G}_p''(z) z^2 \mu_2 + o(\mu_2).$$

So up to the order $O(\mu_2)$, the two expansions agree except that there is an additional term in Theorem 5(b) that reflects the degree of over-identification. If we use the chi-square distribution or the fixed smoothing asymptotic distribution $B_p(1)' C_{pp}^{-1} B_p(1)$ as the reference distribution, then the probability of over-rejection increases with q, at least when μ_2 is small.

We now focus on the case of contracted kernel HAR variance estimation. As $h \to \infty$, i.e. $b \to 0$, we can show that

$$\mu_1 = 1 - bc_1 + o(b) \text{ and } \mu_2 = bc_2 + o(b) \tag{9}$$

where

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

Using Theorem 5(b) and the identity that $-\mathcal{G}'(z^2)(z^2+1) = 2\mathcal{G}''(z^2) z^2$, we have, for $p = 1$:

$$P\left(F_\infty < z^2\right) = \mathcal{G}(z^2) - c_1 \mathcal{G}'(z^2) z^2 b - \frac{c_2}{2}\left[z^4 + (4q+1) z^2\right] \mathcal{G}'(z^2) b + o(b) \tag{10}$$

where $\mathcal{G}(z^2) = \mathcal{G}_1(z^2)$.

The above expansion is related to the high order Edgeworth expansion established in Sun and Phillips (2009) for linear IV regressions. Sun and Phillips (2009) consider the conventional joint limit under which $b \to 0$ and $T \to \infty$ jointly and show that

$$\begin{aligned} P(|t_T(\hat{\theta}_T)| < z) &= \mathbf{\Phi}(z) - \mathbf{\Phi}(-z) \\ &\quad - \left[c_1 z + \frac{1}{2} c_2 \left[z^3 + z(4q+1)\right]\right] b\boldsymbol{\phi}(z) + (bT)^{-g} \rho_{1,\infty} z \boldsymbol{\phi}(z) + s.o. \end{aligned} \tag{11}$$

where $(bT)^{-g} \rho_{1,\infty} z$ captures the nonparametric bias of the kernel HAR variance estimator, and 's.o.' stands for smaller order terms. Here g is the order of the kernel used and the explicit expression for $\rho_{1,\infty}$ is not important here but is given in Sun and Phillips (2009). Observing that $\mathbf{\Phi}(z) - \mathbf{\Phi}(-z) = \mathcal{G}(z^2)$, we have $\boldsymbol{\phi}(z) = \mathcal{G}'(z^2) z$ and $\boldsymbol{\phi}'(z) = \mathcal{G}'(z^2) + 2z^2 \mathcal{G}''(z^2)$. Using these equations, we can represent the high order expansion in (11) as

$$\begin{aligned} P(|t_T(\hat{\theta}_T)|^2 &< z^2) = P(\mathbb{W}_T < z^2) \\ &= \mathcal{G}(z^2) - c_1 \mathcal{G}'(z^2) z^2 b - \frac{c_2}{2}\left[z^4 + (4q+1) z^2\right] \mathcal{G}'(z^2) b + (bT)^{-g} \rho_{1,\infty} \mathcal{G}'(z^2) z^2 + s.o. \end{aligned} \tag{12}$$

Comparing this with (10), the terms of order $O(b)$ are seen to be exactly the same across the two expansions.

Let $z^2 = \mathcal{F}_\infty^{1-\alpha}$ be the $(1-\alpha)$ quantile from the distribution of F_∞, i.e. $P\left(F_\infty < \mathcal{F}_\infty^{1-\alpha}\right) = 1-\alpha$. Then

$$\mathcal{G}\left(\mathcal{F}_\infty^{1-\alpha}\right) - c_1\mathcal{G}'\left(\mathcal{F}_\infty^{1-\alpha}\right)\mathcal{F}_\infty^{1-\alpha} b - \frac{c_2}{2}\left[\left(\mathcal{F}_\infty^{1-\alpha}\right)^2 + (4q+1)\mathcal{F}_\infty^{1-\alpha}\right]\mathcal{G}'\left(\mathcal{F}_\infty^{1-\alpha}\right) b = 1-\alpha+o(b)$$

and so

$$P(\mathbb{W}_T < \mathcal{F}_\infty^{1-\alpha}) = 1-\alpha+(bT)^{-g}\rho_{1,\infty}\mathcal{G}'\left(\mathcal{F}_\infty^{1-\alpha}\right)\mathcal{F}_\infty^{1-\alpha} + s.o.$$

That is, using the critical value $\mathcal{F}_\infty^{1-\alpha}$ eliminates a term in the higher order distributional expansion of $\mathbb{W}_T$ under the conventional asymptotics. The nonstandard critical value $\mathcal{F}_\infty^{1-\alpha}$ is thus high order correct under the conventional asymptotics. In other words, the nonstandard approximation provides a high order refinement over the conventional chi-square approximation.

Although the high order refinement is established for the case of $p=1$ and linear IV regressions, we expect it to be true more generally and for all three types of HAR variance estimators. All we need is a higher order Edgeworth expansion for the Wald statistic in a general GMM setting. In view of Sun and Phillips (2009), this is not difficult conceptually but can be technically very tedious.

5 Simulation Evidence

In this section, we study the finite sample performance of the fixed-smoothing approximations. We consider the following data generating process:

$$y_t = x_{0,t}\alpha + x_{1,t}\beta_1 + x_{2,t}\beta_2 + x_{3,t}\beta_3 + \varepsilon_{y,t}$$

where $x_{0,t}\equiv 1$ and $x_{1,t},\ x_{2,t}$ and $x_{3,t}$ are scalar regressors that are correlated with $\varepsilon_{y,t}$. The dimension of the unknown parameter $\theta = (\alpha,\beta_1,\beta_2,\beta_3)'$ is $d=4$. We have m instruments $z_{0,t}, z_{1,t}, ..., z_{m-1,t}$ with $z_{0,t}\equiv 1$. The reduced-form equations for $x_{1,t},\ x_{2,t}$ and $x_{3,t}$ are given by

$$x_{j,t} = z_{j,t} + \sum_{i=d-1}^{m-1} z_{i,t} + \varepsilon_{x_j,t} \text{ for } j=1,2,3.$$

We assume that $z_{i,t}$ for $i\geq 1$ follows either an AR(1) process

$$z_{i,t} = \rho z_{i,t-1} + \sqrt{1-\rho^2}e_{z_i,t},$$

or an MA(1) process

$$z_{i,t} = \rho e_{z_i,t-1} + \sqrt{1-\rho^2}e_{z_i,t},$$

where

$$e_{z_i,t} = \frac{e_{zt}^i + e_{zt}^0}{\sqrt{2}}$$

and $e_t = [e_{zt}^0, e_{zt}^1, ..., e_{zt}^{m-1}]' \sim iidN(0, I_m)$. By construction, the variance of z_{it} for any $i = 1,2,...,m-1$ is 1. Due to the presence of the common shocks e_{zt}^0, the correlation coefficient between the non-constant $z_{i,t}$ and $z_{j,t}$ for $i\neq j$ is 0.5. The DGP for $\varepsilon_t = (\varepsilon_{yt}, \varepsilon_{x_1t}, \varepsilon_{x_2t}, \varepsilon_{x_3t})'$

is the same as that for $(z_{1,t}, ..., z_{m-1,t})$ except the dimensionality difference. The two vector processes ε_t and $(z_{1,t}, ..., z_{m-1,t})$ are independent from each other.

In the notation of this paper, we have $f(v_t, \theta) = z_t (y_t - x_t'\theta)$ where $v_t = [y_t', x_t', z_t']'$, $x_t = (1, x_{1t}, x_{2t}, x_{3t})'$, $z_t = (1, z_{1t}, ..., z_{m-1,t})'$. We take $\rho = -0.8, -0.5, 0.0, 0.5,$ 0.8 and 0.9. We consider $m = 4, 5, 6$ and the corresponding degrees of over-identification are $q = 0, 1, 2$. The null hypotheses of interest are

$$H_{01} : \beta_1 = 0,$$
$$H_{02} : \beta_1 = \beta_2 = 0,$$
$$H_{03} : \beta_1 = \beta_2 = \beta_3 = 0$$

where $p = 1, 2$ and 3 respectively. The corresponding matrix R is the $2 : p+1$ rows of the identity matrix I_d. We consider two significance levels $\alpha = 5\%$ and $\alpha = 10\%$ and two different sample sizes $T = 100, 200$. The number of simulation replications is 10000.

We first examine the finite sample size accuracy of different tests based on the OS HAR variance estimator. The tests are based on the same Wald test statistic, so they have the same size-adjusted power. The difference lies in the reference distributions or critical values used. We employ the following critical values: $\chi_p^{1-\alpha}/p$, $\frac{K}{K-p+1}\mathcal{F}_{p,K-p+1}^{1-\alpha}$, $\frac{K}{K-p-q+1}\mathcal{F}_{p,K-p-q+1}^{1-\alpha}\left(\delta^2\right)$ with $\delta^2 = pq/(K-q+1)$, and $\mathcal{F}_{\infty}^{1-\alpha}$, leading to the χ^2 test, the CF (central F) test, the NCF (noncentral F) test and the nonstandard F_∞ test. The χ^2 test uses the conventional chi-square approximation. The CF test uses the fixed-smoothing approximation for the Wald statistic based on a first-step GMM estimator. Alternatively, the CF test uses the fixed-smoothing approximation with $q = 0$. The NCF test uses the noncentral F approximation given in Theorem 3. The F_∞ test uses the nonstandard limiting distribution F_∞ with simulated critical values. For each test, the initial first step estimator is the IV estimator with weight matrix $W_o = (Z'Z/T)$ where Z is the matrix of the observed instruments.

To speed up the computation, we assume that K is even and use the basis functions $\Phi_{2j-1}(x) = \sqrt{2}\cos 2j\pi x$, $\Phi_{2j}(x) = \sqrt{2}\sin 2j\pi x$, $j = 1, ..., K/2$. In this case, the OS HAR variance estimator can be computed using discrete Fourier transforms. The OS HAR estimator is a simple average of periodogram. We select K based on the AMSE criterion implemented using the VAR(1) plug-in procedure in Phillips (2005). For completeness, we reproduce the MSE optimal formula for K here:

$$K_{MSE} = 2 \times \left\lceil 0.5 \left(\frac{tr\left[\left(I_{m^2} + \mathbb{K}_{mm} \right) (\Omega \otimes \Omega) \right]}{4vec(B)'vec(B)} \right)^{1/5} T^{4/5} \right\rceil,$$

where $\lceil \cdot \rceil$ is the ceiling function, $\mathbb{K}_{mm}$ is the $m^2 \times m^2$ commutation matrix and

$$B = -\frac{\pi^2}{6} \sum_{j=-\infty}^{\infty} j^2 E f(v_t, \theta_0) f(v_{t-j}, \theta_0)'. \tag{13}$$

Table 1 gives the empirical size of the different tests for the AR(1) case with sample size $T = 100$. The nominal size of the tests is $\alpha = 5\%$. First, as it is clear from the table, the chi-square test can have a large size distortion. The size distortion can be very severe. For example, when $\rho = 0.9$, $p = 3$ and $q = 2$, the empirical size of the chi-square test can be as high as 71.2%, which is far from 5%, the nominal size of the test. Second, the size distortion of the CF test is substantially smaller than the chi-square test when the degree of over-identification is small. This is because the CF test employs the asymptotic approximation that partially captures the estimation uncertainty of the HAR variance estimator. Third, the empirical size of the NCF test is nearly the same as that of the nonstandard F_∞ test. This is consistent with Figure 1. This result provides further evidence that the noncentral F distribution approximates the nonstandard F_∞ distribution very well. Finally, among the four tests, the NCF and F_∞ tests have the most accurate size. For the two-step GMM estimator, the HAR variance estimator appears in two different places and plays two different roles — first as the inverse of the optimal weighting matrix and then as part of the asymptotic variance estimator for the GMM estimator. The nonstandard F_∞ approximation and the noncentral F approximation attempt to capture the estimation uncertainty of the HAR variance estimator in both places. In contrast, a crucial step underlying the central F approximation is that the HAR variance estimator is consistent when it acts as the optimal weighting matrix. As a result, the central F approximation does not adequately capture the estimation uncertainty of the HAR variance estimator. This explains why the NCF and F_∞ tests have more accurate size than the CF test.

Table 2 presents the simulated empirical size for the MA(1) case. The qualitative observations for the AR(1) case remain valid. The chi-square test is most size distorted. The NCF and F_∞ tests are least size distorted. The CF test is in between. As before, the size distortion increases with the serial dependence, the number of joint hypotheses, and the degree of over-identification. This table provides further evidence that the noncentral F distribution and the nonstandard F_∞ distribution provide more accurate approximations to the sampling distribution of the Wald statistic.

Tables 3 and 4 report results for the case when the sample size is 200. Compared to the cases with sample size 100, all tests become more accurate in size. This is well expected. In terms of the size accuracy, the NCF test and the F_∞ test are close to each other. They dominate the CF test, which in turn dominates the chi-square test.

Next, we consider the empirical size of different tests based on kernel HAR variance estimators. Both the contracted kernel approach and the exponentiated kernel approach are considered, but we report only the commonly-used contracted kernel approach as the results are qualitatively similar. We employ three kernels: the Bartlett, Parzen and QS kernels. These three kernels are positive (semi) definite. For each kernel, we use the data-driven AMSE optimal bandwidth and its VAR(1) plug-in implementation from Andrews (1991). For each Wald statistic, three critical values are used: $\chi_p^{1-\alpha}/p$, $\mathcal{F}_\infty^{1-\alpha}(0)$, and $\mathcal{F}_\infty^{1-\alpha}(q)$ where $\mathcal{F}_\infty^{1-\alpha}(q)$ is the $(1-\alpha)$ quantile from the nonstandard F_∞ distribution with degree of over-identification q. $\mathcal{F}_\infty^{1-\alpha}(0)$ coincides with the critical value from the fixed-smoothing asymptotics distribution derived for the first-step IV estimator.

To save space, we only report the results for the Parzen kernel. Tables 5–8 correspond to Tables 1–4. It is clear that the qualitative results exhibited in Tables 1–4 continue to apply. According to the size accuracy, the $F_\infty(q)$ test dominates the $F_\infty(0)$ test, which in turn dominates the conventional χ^2 test.

6 Conclusion

The paper has developed the fixed-smoothing asymptotics for heteroskedasticity and autocorrelation robust inference in a two-step GMM framework. We have shown that the conventional chi-square test that ignores the estimation uncertainty of the GMM weighting matrix and the covariance estimator can have very large size distortion. This is especially true when the number of joint hypotheses being tested and the degree of over-identification are high or the underlying processes are persistent. The test based on our new fixed-smoothing approximation reduces the size distortion substantially and is thus recommended for practical use.

There are a number of interesting extensions. First, given that our proof uses only a CLT, the results of the paper can be extended easily to the spatial setting, spatial-temporal setting or panel data setting. See Kim and Sun (2012) for an extension along this line. Second, in the Monte Carlo experiments, we use the conventional MSE criterion to select the smoothing parameters. We could have used the methods proposed in Sun and Phillips (2009), Sun, Phillips and Jin (2008) or Sun (2012) that are tailored towards confidence interval construction or hypothesis testing. But in their present forms these methods are either available only for the t-test or work only in the first-step GMM framework. It will be interesting to extend them to more general tests in the two-step GMM framework. Third, we can use the asymptotically equivalent distribution to conduct asymptotically valid inferences. Simulation results not reported here show that the size difference between using the asymptotically equivalent distribution and the limiting distribution is negligible. However, it is easy to replace the iid process in the asymptotically equivalent distribution F_{eT} by a dependent process that mimics the degree of autocorrelation in the moment process. Such a replacement may lead to even more accurate tests and is similar in spirit to Zhang and Shao (2013) who have shown the high order refinement of a Gaussian bootstrap procedure under the fixed-smoothing asymptotics. Finally, the general results of the paper can be extended to a nonparametric sieve GMM framework. See Chen, Liao and Sun (2012) for a recent development on autocorrelation robust sieve inference for time series models.

Table 1: Empirical size of the χ^2 test, F test, noncentral F test and nonstandard F_∞ test based on the series LRV estimator for the AR(1) case with $T = 100$

$\boldsymbol{\rho}$	χ^2	CF	NCF	F_∞	χ^2	CF	NCF	F_∞	χ^2	CF	NCF	F_∞
	$p=1,q=0$				$p=2,q=0$				$p=3,q=0$			
-0.8	0.114	0.072	0.072	0.073	0.197	0.087	0.087	0.084	0.310	0.109	0.109	0.111
-0.5	0.081	0.060	0.060	0.059	0.117	0.066	0.066	0.066	0.174	0.077	0.077	0.078
0.0	0.063	0.051	0.051	0.050	0.083	0.052	0.052	0.053	0.112	0.060	0.060	0.062
0.5	0.094	0.063	0.063	0.063	0.142	0.065	0.065	0.065	0.222	0.077	0.077	0.078
0.8	0.134	0.086	0.086	0.088	0.229	0.100	0.100	0.097	0.355	0.119	0.119	0.122
0.9	0.166	0.117	0.117	0.120	0.290	0.150	0.150	0.146	0.437	0.181	0.181	0.184
	$p=1,q=1$				$p=2,q=1$				$p=3,q=1$			
-0.8	0.186	0.129	0.081	0.077	0.307	0.153	0.088	0.087	0.457	0.193	0.107	0.113
-0.5	0.113	0.086	0.065	0.065	0.175	0.099	0.069	0.068	0.247	0.116	0.079	0.080
0.0	0.081	0.065	0.053	0.052	0.113	0.075	0.057	0.056	0.155	0.085	0.060	0.060
0.5	0.128	0.091	0.064	0.063	0.204	0.110	0.073	0.072	0.308	0.130	0.079	0.080
0.8	0.196	0.140	0.089	0.086	0.331	0.172	0.101	0.099	0.489	0.209	0.112	0.118
0.9	0.252	0.184	0.126	0.123	0.420	0.242	0.155	0.153	0.589	0.301	0.183	0.192
	$p=1,q=2$				$p=2,q=2$				$p=3,q=2$			
-0.8	0.260	0.190	0.080	0.079	0.425	0.256	0.090	0.083	0.602	0.318	0.100	0.097
-0.5	0.154	0.117	0.061	0.062	0.244	0.146	0.065	0.061	0.351	0.175	0.074	0.073
0.0	0.104	0.085	0.055	0.055	0.148	0.101	0.062	0.058	0.211	0.119	0.065	0.063
0.5	0.171	0.129	0.065	0.066	0.279	0.165	0.065	0.061	0.415	0.200	0.073	0.072
0.8	0.268	0.199	0.085	0.082	0.449	0.269	0.090	0.082	0.623	0.330	0.100	0.097
0.9	0.332	0.257	0.124	0.121	0.529	0.358	0.148	0.137	0.712	0.433	0.160	0.157

Table 2: Empirical size of the χ^2 test, F test, noncentral F test and nonstandard F_∞ test based on the series LRV estimator for the MA(1) case with $T = 100$

ρ	χ^2	CF	NCF	F_∞	χ^2	CF	NCF	F_∞	χ^2	CF	NCF	F_∞
		$p=1,q=0$				$p=2,q=0$				$p=3,q=0$		
-0.8	0.072	0.055	0.055	0.055	0.106	0.058	0.058	0.058	0.153	0.069	0.069	0.070
-0.5	0.074	0.055	0.055	0.055	0.101	0.058	0.058	0.059	0.144	0.069	0.069	0.070
0.0	0.063	0.051	0.051	0.050	0.083	0.052	0.052	0.053	0.112	0.060	0.060	0.062
0.5	0.078	0.054	0.054	0.054	0.119	0.060	0.060	0.061	0.182	0.071	0.071	0.072
0.8	0.081	0.056	0.056	0.056	0.125	0.060	0.060	0.062	0.195	0.071	0.071	0.071
0.9	0.077	0.055	0.055	0.055	0.114	0.059	0.059	0.060	0.170	0.070	0.070	0.070
		$p=1,q=1$				$p=2,q=1$				$p=3,q=1$		
-0.8	0.099	0.074	0.057	0.056	0.160	0.090	0.061	0.060	0.223	0.099	0.067	0.068
-0.5	0.097	0.075	0.057	0.057	0.150	0.088	0.061	0.060	0.207	0.096	0.068	0.068
0.0	0.081	0.065	0.053	0.052	0.113	0.075	0.057	0.056	0.155	0.085	0.060	0.060
0.5	0.108	0.078	0.057	0.056	0.171	0.093	0.062	0.062	0.253	0.107	0.068	0.068
0.8	0.116	0.082	0.057	0.057	0.185	0.097	0.062	0.062	0.274	0.115	0.069	0.070
0.9	0.107	0.077	0.058	0.057	0.163	0.093	0.063	0.062	0.239	0.107	0.070	0.071
		$p=1,q=2$				$p=2,q=2$				$p=3,q=2$		
-0.8	0.137	0.105	0.054	0.054	0.216	0.133	0.058	0.056	0.305	0.153	0.066	0.066
-0.5	0.131	0.100	0.055	0.055	0.204	0.128	0.063	0.060	0.284	0.146	0.066	0.065
0.0	0.104	0.085	0.055	0.055	0.148	0.101	0.062	0.058	0.211	0.119	0.065	0.063
0.5	0.143	0.109	0.060	0.061	0.231	0.138	0.063	0.060	0.339	0.168	0.067	0.067
0.8	0.153	0.115	0.060	0.061	0.247	0.146	0.062	0.059	0.366	0.177	0.067	0.066
0.9	0.137	0.105	0.058	0.059	0.218	0.134	0.061	0.059	0.316	0.159	0.070	0.070

Table 3: Empirical size of the χ^2 test, F test, noncentral F test and nonstandard F_∞ test based on the series LRV estimator for the AR(1) case with $T = 200$

ρ	χ^2	CF	NCF	F_∞	χ^2	CF	NCF	F_∞	χ^2	CF	NCF	F_∞
		$p=1,q=0$				$p=2,q=0$				$p=3,q=0$		
-0.8	0.102	0.071	0.071	0.071	0.150	0.073	0.073	0.073	0.234	0.091	0.091	0.091
-0.5	0.070	0.058	0.058	0.058	0.088	0.062	0.062	0.063	0.117	0.071	0.071	0.072
0.0	0.060	0.054	0.054	0.054	0.065	0.052	0.052	0.052	0.081	0.059	0.059	0.057
0.50	0.079	0.063	0.063	0.063	0.104	0.063	0.063	0.064	0.144	0.066	0.066	0.070
0.8	0.114	0.072	0.072	0.073	0.187	0.079	0.079	0.078	0.283	0.092	0.092	0.093
0.9	0.140	0.091	0.091	0.094	0.243	0.107	0.107	0.104	0.366	0.124	0.124	0.128
		$p=1,q=1$				$p=2,q=1$				$p=3,q=1$		
-0.8	0.147	0.107	0.072	0.072	0.234	0.126	0.080	0.081	0.348	0.147	0.092	0.094
-0.5	0.083	0.070	0.058	0.057	0.117	0.081	0.066	0.065	0.154	0.094	0.073	0.072
0.0	0.069	0.060	0.053	0.053	0.081	0.064	0.058	0.058	0.100	0.071	0.062	0.061
0.5	0.099	0.080	0.062	0.061	0.139	0.088	0.065	0.062	0.182	0.092	0.070	0.072
0.8	0.164	0.112	0.076	0.073	0.270	0.138	0.083	0.083	0.394	0.159	0.087	0.090
0.9	0.205	0.145	0.095	0.090	0.347	0.178	0.110	0.108	0.498	0.219	0.121	0.127
		$p=1,q=2$				$p=2,q=2$				$p=3,q=2$		
-0.8	0.195	0.148	0.071	0.073	0.319	0.192	0.075	0.070	0.460	0.233	0.083	0.082
-0.5	0.101	0.082	0.060	0.060	0.144	0.098	0.066	0.063	0.190	0.116	0.069	0.070
0.0	0.074	0.066	0.053	0.054	0.096	0.074	0.056	0.057	0.117	0.083	0.059	0.060
0.5	0.113	0.091	0.056	0.056	0.166	0.107	0.064	0.059	0.234	0.129	0.070	0.069
0.8	0.225	0.170	0.072	0.073	0.371	0.216	0.075	0.070	0.539	0.264	0.081	0.078
0.9	0.276	0.210	0.090	0.088	0.460	0.280	0.099	0.090	0.639	0.349	0.106	0.103

Table 4: Empirical size of the χ^2 test, F test, noncentral F test and nonstandard F_∞ test based on the series LRV estimator for the MA(1) case with $T = 200$

ρ	χ^2	CF	NCF	F_∞	χ^2	CF	NCF	F_∞	χ^2	CF	NCF	F_∞
	$p=1, q=0$				$p=2, q=0$				$p=3, q=0$			
-0.8	0.066	0.054	0.054	0.053	0.082	0.056	0.056	0.056	0.106	0.062	0.062	0.064
-0.5	0.066	0.056	0.056	0.055	0.079	0.054	0.054	0.056	0.103	0.061	0.061	0.064
0.0	0.060	0.054	0.054	0.054	0.065	0.052	0.052	0.052	0.081	0.059	0.059	0.057
0.5	0.070	0.058	0.058	0.057	0.089	0.057	0.057	0.057	0.120	0.060	0.060	0.063
0.8	0.072	0.057	0.057	0.057	0.091	0.058	0.058	0.058	0.130	0.060	0.060	0.063
0.9	0.069	0.056	0.056	0.056	0.084	0.057	0.057	0.057	0.114	0.061	0.061	0.063
	$p=1, q=1$				$p=2, q=1$				$p=3, q=1$			
-0.8	0.080	0.066	0.055	0.055	0.105	0.075	0.063	0.061	0.141	0.084	0.069	0.068
-0.5	0.077	0.065	0.056	0.056	0.102	0.076	0.064	0.062	0.134	0.082	0.066	0.066
0.0	0.069	0.060	0.053	0.053	0.081	0.064	0.058	0.058	0.100	0.071	0.062	0.061
0.5	0.085	0.071	0.058	0.057	0.115	0.077	0.060	0.058	0.152	0.083	0.064	0.063
0.8	0.089	0.072	0.057	0.056	0.123	0.077	0.058	0.057	0.164	0.085	0.064	0.065
0.9	0.080	0.067	0.057	0.056	0.110	0.074	0.059	0.058	0.142	0.084	0.066	0.065
	$p=1, q=2$				$p=2, q=2$				$p=3, q=2$			
-0.8	0.096	0.080	0.056	0.056	0.130	0.090	0.059	0.058	0.177	0.109	0.067	0.066
-0.5	0.095	0.079	0.055	0.057	0.123	0.089	0.060	0.060	0.165	0.106	0.065	0.064
0.0	0.074	0.066	0.053	0.054	0.096	0.074	0.056	0.057	0.117	0.083	0.059	0.060
0.5	0.100	0.080	0.054	0.055	0.136	0.094	0.059	0.058	0.184	0.110	0.065	0.066
0.8	0.103	0.083	0.054	0.055	0.146	0.096	0.059	0.056	0.202	0.113	0.063	0.063
0.9	0.093	0.079	0.055	0.056	0.131	0.093	0.059	0.058	0.172	0.105	0.065	0.065

Table 5: Empirical size of the χ^2 test, the $F_\infty(0)$ test and the $F_\infty(q)$ test based on the Parzen kernel LRV estimator for the AR(1) case with $T = 100$

ρ	χ^2	$F_\infty(0)$	$F_\infty(q)$	χ^2	$F_\infty(0)$	$F_\infty(q)$	χ^2	$F_\infty(0)$	$F_\infty(q)$
	$p=1, q=0$			$p=2, q=0$			$p=3, q=0$		
-0.8	0.132	0.091	0.091	0.215	0.112	0.112	0.326	0.153	0.153
-0.5	0.085	0.073	0.073	0.116	0.082	0.082	0.160	0.100	0.100
0.0	0.063	0.059	0.059	0.071	0.058	0.058	0.093	0.070	0.070
0.5	0.093	0.074	0.074	0.131	0.076	0.076	0.182	0.096	0.096
0.8	0.164	0.102	0.102	0.273	0.126	0.126	0.408	0.154	0.154
0.9	0.216	0.124	0.124	0.382	0.164	0.164	0.559	0.212	0.212
	$p=1, q=1$			$p=2, q=1$			$p=3, q=1$		
-0.8	0.197	0.145	0.110	0.313	0.174	0.127	0.443	0.224	0.164
-0.5	0.105	0.091	0.076	0.150	0.104	0.088	0.201	0.126	0.104
0.0	0.073	0.067	0.059	0.096	0.077	0.067	0.121	0.084	0.074
0.5	0.117	0.095	0.079	0.169	0.111	0.090	0.236	0.132	0.104
0.8	0.223	0.151	0.110	0.362	0.189	0.132	0.521	0.230	0.160
0.9	0.315	0.204	0.145	0.521	0.271	0.182	0.704	0.329	0.226
	$p=1, q=2$			$p=2, q=2$			$p=3, q=2$		
-0.8	0.257	0.195	0.115	0.406	0.242	0.139	0.555	0.305	0.166
-0.5	0.128	0.110	0.078	0.188	0.131	0.091	0.248	0.155	0.103
0.0	0.086	0.079	0.060	0.114	0.091	0.070	0.150	0.108	0.080
0.5	0.139	0.115	0.081	0.210	0.140	0.088	0.291	0.167	0.105
0.8	0.282	0.201	0.113	0.456	0.260	0.135	0.613	0.319	0.157
0.9	0.384	0.269	0.144	0.605	0.365	0.190	0.774	0.438	0.221

Table 6: Empirical size of the χ^2 test, the $F_\infty(0)$ test and the $F_\infty(q)$ test based on the Parzen kernel LRV estimator for the MA(1) case with $T = 100$

ρ	χ^2	$F_\infty(0)$	$F_\infty(q)$	χ^2	$F_\infty(0)$	$F_\infty(q)$	χ^2	$F_\infty(0)$	$F_\infty(q)$
	$p=1,q=0$			$p=2,q=0$			$p=3,q=0$		
-0.8	0.075	0.065	0.065	0.101	0.070	0.070	0.133	0.086	0.086
-0.5	0.073	0.065	0.065	0.096	0.070	0.070	0.126	0.085	0.085
0.0	0.063	0.059	0.059	0.071	0.058	0.058	0.093	0.070	0.070
0.5	0.079	0.064	0.064	0.107	0.069	0.069	0.150	0.084	0.084
0.8	0.081	0.066	0.066	0.113	0.070	0.070	0.160	0.086	0.086
0.9	0.076	0.063	0.063	0.103	0.070	0.070	0.140	0.082	0.082
	$p=1,q=1$			$p=2,q=1$			$p=3,q=1$		
-0.8	0.091	0.079	0.067	0.133	0.094	0.078	0.173	0.108	0.088
-0.5	0.086	0.076	0.067	0.126	0.094	0.078	0.162	0.106	0.087
0.0	0.073	0.067	0.059	0.096	0.077	0.067	0.121	0.084	0.074
0.5	0.097	0.081	0.070	0.140	0.094	0.077	0.191	0.111	0.088
0.8	0.104	0.083	0.069	0.149	0.097	0.077	0.203	0.116	0.090
0.9	0.095	0.081	0.069	0.134	0.092	0.075	0.181	0.108	0.091
	$p=1,q=2$			$p=2,q=2$			$p=3,q=2$		
-0.8	0.114	0.099	0.072	0.169	0.118	0.083	0.215	0.139	0.093
-0.5	0.110	0.096	0.069	0.157	0.114	0.081	0.201	0.133	0.091
0.0	0.086	0.079	0.060	0.114	0.091	0.070	0.150	0.108	0.080
0.5	0.120	0.100	0.071	0.168	0.117	0.083	0.229	0.138	0.092
0.8	0.125	0.105	0.072	0.181	0.122	0.081	0.249	0.144	0.091
0.9	0.113	0.098	0.071	0.160	0.114	0.081	0.216	0.133	0.091

Table 7: Empirical size of the χ^2 test, the $F_\infty(0)$ test and the $F_\infty(q)$ test based on the Parzen kernel LRV estimator for the AR(1) case with $T = 200$

ρ	χ^2	$F_\infty(0)$	$F_\infty(q)$	χ^2	$F_\infty(0)$	$F_\infty(q)$	χ^2	$F_\infty(0)$	$F_\infty(q)$
	$p=1,q=0$			$p=2,q=0$			$p=3,q=0$		
-0.8	0.110	0.090	0.090	0.150	0.098	0.098	0.218	0.123	0.123
-0.5	0.075	0.073	0.073	0.090	0.078	0.078	0.114	0.093	0.093
0.0	0.059	0.061	0.061	0.061	0.058	0.058	0.074	0.070	0.070
0.5	0.080	0.073	0.073	0.096	0.075	0.075	0.126	0.086	0.086
0.8	0.124	0.091	0.091	0.186	0.102	0.102	0.261	0.122	0.122
0.9	0.170	0.105	0.105	0.286	0.129	0.129	0.421	0.161	0.161
	$p=1,q=1$			$p=2,q=1$			$p=3,q=1$		
-0.8	0.143	0.117	0.096	0.208	0.136	0.110	0.293	0.166	0.134
-0.5	0.083	0.079	0.071	0.108	0.091	0.083	0.137	0.109	0.100
0.0	0.063	0.064	0.061	0.074	0.070	0.066	0.086	0.079	0.075
0.5	0.091	0.085	0.074	0.117	0.091	0.079	0.147	0.102	0.088
0.8	0.155	0.118	0.094	0.239	0.142	0.110	0.328	0.167	0.127
0.9	0.235	0.160	0.116	0.391	0.199	0.140	0.542	0.243	0.170
	$p=1,q=2$			$p=2,q=2$			$p=3,q=2$		
-0.8	0.168	0.138	0.094	0.257	0.174	0.114	0.349	0.213	0.137
-0.5	0.092	0.087	0.073	0.124	0.103	0.083	0.149	0.116	0.095
0.0	0.066	0.067	0.061	0.078	0.074	0.065	0.092	0.084	0.074
0.5	0.098	0.090	0.071	0.131	0.103	0.079	0.172	0.121	0.092
0.8	0.193	0.154	0.096	0.295	0.185	0.113	0.415	0.230	0.129
0.9	0.296	0.215	0.121	0.469	0.275	0.147	0.637	0.338	0.173

Table 8: Empirical size of the χ^2 test, the $F_\infty(0)$ test and the $F_\infty(q)$ test based on the Parzen kernel LRV estimator for the MA(1) case with $T = 200$

ρ	χ^2	$F_\infty(0)$	$F_\infty(q)$	χ^2	$F_\infty(0)$	$F_\infty(q)$	χ^2	$F_\infty(0)$	$F_\infty(q)$
	$p=1,q=0$			$p=2,q=0$			$p=3,q=0$		
-0.8	0.065	0.064	0.064	0.077	0.068	0.068	0.098	0.077	0.077
-0.5	0.065	0.065	0.065	0.075	0.067	0.067	0.093	0.077	0.077
0.0	0.059	0.061	0.061	0.061	0.058	0.058	0.074	0.070	0.070
0.5	0.070	0.066	0.066	0.082	0.066	0.066	0.104	0.074	0.074
0.8	0.071	0.066	0.066	0.085	0.066	0.066	0.111	0.075	0.075
0.9	0.069	0.066	0.066	0.079	0.065	0.065	0.101	0.076	0.076
	$p=1,q=1$			$p=2,q=1$			$p=3,q=1$		
-0.8	0.076	0.074	0.065	0.095	0.082	0.074	0.118	0.093	0.085
-0.5	0.074	0.073	0.066	0.091	0.082	0.074	0.113	0.093	0.086
0.0	0.063	0.064	0.061	0.074	0.070	0.066	0.086	0.079	0.075
0.5	0.077	0.074	0.066	0.097	0.080	0.070	0.118	0.089	0.080
0.8	0.081	0.076	0.066	0.102	0.081	0.071	0.126	0.090	0.078
0.9	0.075	0.072	0.064	0.095	0.078	0.069	0.114	0.089	0.080
	$p=1,q=2$			$p=2,q=2$			$p=3,q=2$		
-0.8	0.087	0.083	0.069	0.107	0.092	0.073	0.135	0.108	0.089
-0.5	0.085	0.083	0.069	0.104	0.088	0.072	0.131	0.106	0.086
0.0	0.066	0.067	0.061	0.078	0.074	0.065	0.092	0.084	0.074
0.5	0.085	0.082	0.066	0.111	0.092	0.072	0.137	0.104	0.083
0.8	0.089	0.083	0.065	0.118	0.094	0.072	0.145	0.107	0.082
0.9	0.083	0.081	0.066	0.108	0.090	0.071	0.132	0.102	0.085

7 Appendix of Proofs

Consider two stochastically bounded sequences of random variables $\{\xi_{1,T} \in \mathbb{R}^p\}$ and $\{\xi_{2,T} \in \mathbb{R}^p\}$. Since $\{\xi_{1,T}\}$ is stochastically bounded, there exists a subsequence $\{\xi_{1,s_1(T)} : T = 1, 2, ...\}$ of $\{\xi_{1,T} : T = 1, 2, ...\}$ such that $\xi_{1,s_1(T)} \overset{d}{\to} \xi_{1,\infty}$ for some random variable $\xi_{1,\infty}$. Note that the subsequence $\{\xi_{2,s_1(T)}\}$ is also stochastically bounded, so there exists a subsequence $\{\xi_{2,s_2(s_1(T))}\}$ of $\{\xi_{2,s_1(T)}\}$ such that $\xi_{2,s_2(s_1(T))} \overset{d}{\to} \xi_{2,\infty}$ for some random variable $\xi_{2,\infty}$. Define $s(T) = s_2(s_1(T))$, then $\xi_{1,s(T)} \overset{d}{\to} \xi_{1,\infty}$ and $\xi_{2,s(T)} \overset{d}{\to} \xi_{2,\infty}$. Let $\mathcal{C}(\xi_{1,\infty}, \xi_{2,\infty})$ be the set of points at which the CDFs of $\xi_{1,\infty}$ and $\xi_{2,\infty}$ are continuous. The following lemma uses the definitions of $s(T)$ and $\mathcal{C}(\xi_{1,\infty}, \xi_{2,\infty})$. The proof is similar to that for the Lévy-Cramér continuity theorem and is omitted here.

Lemma 3 *The following two statements are equivalent:*

(i) $P(\xi_{1,s(T)} \leq x) = P(\xi_{2,s(T)} \leq x) + o(1)$ *for all* $x \in \mathcal{C}(\xi_{1\infty}, \xi_{2\infty})$.

(ii) For any $f \in \mathcal{BC}$, $Ef(\xi_{1,s(T)}) - Ef(\xi_{2,s(T)}) \to 0$ *as* $T \to \infty$ *where* $\mathcal{BC}$ *is the class of bounded, continuous and real valued functions on* $\mathbb{R}^p$.

Lemma 4 *If* (ξ_{1T}, η_{1T}) *converges weakly to* (ξ, η) *and*

$$P(\xi_{1T} \leq x, \eta_{1T} \leq y) = P(\xi_{2T} \leq x, \eta_{2T} \leq y) + o(1) \text{ as } T \to \infty \tag{14}$$

for all continuity point (x, y) *of the CDF of* (ξ, η), *then*

$$g(\xi_{1T}, \eta_{1T}) \overset{a}{\sim} g(\xi_{2T}, \eta_{2T}),$$

where $g(\cdot, \cdot)$ *is continuous on a set* $\mathcal{C}$ *such that* $P((\xi, \eta) \in \mathcal{C}) = 1$.

Proof of Lemma 4. Since (ξ_{1T}, η_{1T}) converges weakly, we can apply Lemma 3 with $s(T) = T$. It follows from the condition in (14) that $Ef(\xi_{1T}, \eta_{1T}) - Ef(\xi_{2T}, \eta_{2T}) \to 0$ for any $f \in \mathcal{BC}$. But $Ef(\xi_{1T}, \eta_{1T}) \to Ef(\xi, \eta)$ and so $Ef(\xi_{2T}, \eta_{2T}) \to Ef(\xi, \eta)$ for any $f \in \mathcal{BC}$. That is, (ξ_{2T}, η_{2T}) also converges weakly to (ξ, η). Using the same proof for proving the continuous mapping theorem, we have $Ef(g(\xi_{1T}, \eta_{1T})) - Ef(g(\xi_{2T}, \eta_{2T})) \to 0$ for any $f \in \mathcal{BC}$. Therefore $g(\xi_{1T}, \eta_{1T}) \overset{a}{\sim} g(\xi_{2T}, \eta_{2T})$.
■

Proof of Lemma 1. Let $\varepsilon > 0$. Under condition (iii), we can find $\delta > 0$ such that for some integer $T_{\min} > 0$

$$P(\xi - \delta \leq \xi_T < \xi + \delta) \leq \varepsilon$$

for all $T \geq T_{\min}$. Here $T_{\min}$ does not depend on δ or ε. Under condition (iv), we can find a $J_{\min}$ that does not depend on δ or ε such that

$$P(|\eta_{T,J}| > \delta) \leq \varepsilon$$

for all $J \geq J_{\min}$ and all T. From condition (ii), we can find $J'_{\min} \geq J_{\min}$ and $T'_{\min} \geq T_{\min}$ such that

$$\left|P\left(\xi^*_{T,J} < \xi\right) - P(\xi_T < \xi)\right| \leq \varepsilon$$

for all $J \geq J'_{\min}$ and all $T \geq T'_{\min}$. It follows from condition (i) that for any fixed $J_0 \geq J'_{\min}$, there exists a $T''_{\min}(J_0) \geq T'_{\min} \geq T_{\min}$ such that

$$\left|P\left(\xi_{T,J_0} < \xi + \delta\right) - P\left(\xi^*_{T,J_0} < \xi + \delta\right)\right| \leq \varepsilon$$
$$\left|P\left(\xi_{T,J_0} < \xi - \delta\right) - P\left(\xi^*_{T,J_0} < \xi - \delta\right)\right| \leq \varepsilon$$

for $T \geq T''_{\min}(J_0)$.

When $T \geq T''_{\min}(J_0)$, we have

$$\begin{aligned} P\left(\omega_T \leq \xi\right) &= P\left(\xi_{T,J_0} + \eta_{T,J_0} \leq \xi\right) \leq P\left(\xi_{T,J_0} \leq \xi + \delta\right) + P\left(\left|\eta_{T,J_0}\right| > \delta\right) \\ &\leq P\left(\xi^*_{T,J_0} \leq \xi + \delta\right) + 2\varepsilon \leq P\left(\xi_T < \xi + \delta\right) + 3\varepsilon \\ &\leq P\left(\xi_T < \xi\right) + 4\varepsilon. \end{aligned}$$

Similarly,

$$\begin{aligned} P\left(\omega_T \leq \xi\right) &= P\left(\xi_{T,J_0} + \eta_{T,J_0} \leq \xi\right) \geq P(\xi_{T,J_0} \leq \xi - \delta) - P\left(\left|\eta_{T,J_0}\right| \geq \delta\right) \\ &\geq P(\xi^*_{T,J_0} \leq \xi - \delta) - 2\varepsilon \geq P(\xi_T \leq \xi - \delta) - 3\varepsilon \\ &\geq P(\xi_T \leq \xi) - 4\varepsilon. \end{aligned}$$

Since the above two inequalities hold for all $\varepsilon > 0$, we must have $P\left(\omega_T < \xi\right) = P\left(\xi_T < \xi\right) + o(1)$ as $T \to \infty$. ■

Proof of Lemma 2. Part (a). For some $\bar{\theta}^*_T$ between $\breve{\theta}_T$ and θ_0, we have

$$\begin{aligned} W_T(\breve{\theta}_T) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} Q^*_h\left(\frac{t}{T},\frac{\tau}{T}\right)\left\{u_t + \frac{\partial f(v_t,\bar{\theta}^*_T)}{\partial\theta'}\left(\breve{\theta}_T - \theta_0\right)\right\}\left\{u_\tau + \frac{\partial f(v_\tau,\bar{\theta}^*_T)}{\partial\theta'}\left(\breve{\theta}_T - \theta_0\right)\right\}' \\ &= W_T\left(\theta_0\right) + I_1 + I_1' + I_2, \end{aligned} \tag{15}$$

where

$$\begin{aligned} I_1 &= \frac{1}{T}\sum_{\tau=1}^{T}\sum_{t=1}^{T} Q^*_h\left(\frac{t}{T},\frac{\tau}{T}\right)\frac{\partial f(v_t,\bar{\theta}^*_T)}{\partial\theta'}\left(\breve{\theta}_T - \theta_0\right)u_\tau', \\ I_2 &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T} Q^*_h\left(\frac{t}{T},\frac{\tau}{T}\right)\left[\frac{\partial f(v_t,\bar{\theta}^*_T)}{\partial\theta'}\left(\breve{\theta}_T - \theta_0\right)\right]\left[\frac{\partial f(v_\tau,\bar{\theta}^*_T)}{\partial\theta'}\left(\breve{\theta}_T - \theta_0\right)\right]'. \end{aligned}$$

Here $\bar{\theta}^*_T$ in the matrix $\frac{\partial f(v_t,\bar{\theta}^*_T)}{\partial\theta'}$ can be different for different rows. For notational economy, we do not make this explicit.

For a sequence of vectors or matrices $\{C_t\}$, define $\mathcal{S}_0\left(C\right) = 0$ and $\mathcal{S}_t\left(C\right) = T^{-1}\sum_{\tau=1}^{t} C_\tau$. Using summation by parts, we obtain, for two sequence of matrices $\{A_t\}$ and $\{B_t\}$:

$$\begin{aligned} &\frac{1}{T^2}\sum_{\tau=1}^{T}\sum_{t=1}^{T} Q^*_h\left(\frac{t}{T},\frac{\tau}{T}\right)A_t B_\tau' \\ &= \sum_{\tau=1}^{T-1}\sum_{t=1}^{T-1}\nabla Q^*_h\left(\frac{t}{T},\frac{\tau}{T}\right)\mathcal{S}_t\left(A\right)\mathcal{S}_\tau'\left(B\right) \\ &+ \sum_{t=1}^{T-1}\left[Q^*_h\left(\frac{t}{T},1\right) - Q^*_h\left(\frac{t+1}{T},1\right)\right]\mathcal{S}_t\left(A\right)\mathcal{S}_T'\left(B\right) \\ &+ \sum_{\tau=1}^{T-1}\left[Q^*_h\left(1,\frac{\tau}{T}\right) - Q^*_h\left(1,\frac{\tau+1}{T}\right)\right]\mathcal{S}_T\left(A\right)\mathcal{S}_\tau'\left(B\right) + Q^*_h\left(1,1\right)\mathcal{S}_T\left(A\right)\mathcal{S}_T'\left(B\right). \end{aligned}$$

where

$$\nabla Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) := Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) - Q_h^*\left(\frac{t+1}{T},\frac{\tau}{T}\right) - Q_h^*\left(\frac{t}{T},\frac{\tau+1}{T}\right) + Q_h^*\left(\frac{t+1}{T},\frac{\tau+1}{T}\right).$$

Plugging $A_t = \frac{\partial f(v_t,\bar{\theta}_T^*)}{\partial\theta'}\left(\breve{\theta}_T-\theta_0\right) = G_t(\bar{\theta}_T^*)\left(\breve{\theta}_T-\theta_0\right)$ and $B_t = u_t$ into the above expression and letting $\epsilon_t = G_t(\bar{\theta}_T^*) - \frac{t}{T}G$, we obtain a new representation of I_1 :

$$\begin{aligned}
I_1 &= T\sum_{\tau=1}^{T-1}\sum_{t=1}^{T-1}\nabla Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right)\left(\frac{t}{T}G+\epsilon_t\right)\left(\breve{\theta}_T-\theta_0\right)\mathcal{S}_\tau'(u) \\
&+ T\sum_{t=1}^{T-1}\left[Q_h^*\left(\frac{t}{T},1\right)-Q_h^*\left(\frac{t+1}{T},1\right)\right]\left(\frac{t}{T}G+\epsilon_t\right)\left(\breve{\theta}_T-\theta_0\right)\mathcal{S}_T'(u) \\
&+ T\sum_{\tau=1}^{T-1}\left[Q_h^*\left(1,\frac{\tau}{T}\right)-Q_h^*\left(1,\frac{\tau+1}{T}\right)\right](G+\epsilon_T)\left(\breve{\theta}_T-\theta_0\right)\mathcal{S}_\tau'(u) \\
&+ TQ_h^*(1,1)(G+\epsilon_T)\left(\breve{\theta}_T-\theta_0\right)\mathcal{S}_T'(u) \\
&= T\sum_{\tau=1}^{T-1}\sum_{t=1}^{T-1}\nabla Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right)\epsilon_t\left(\breve{\theta}_T-\theta_0\right)\mathcal{S}_\tau'(u) \\
&+ T\sum_{t=1}^{T-1}\left[Q_h^*\left(\frac{t}{T},1\right)-Q_h^*\left(\frac{t+1}{T},1\right)\right]\epsilon_t\left(\breve{\theta}_T-\theta_0\right)\mathcal{S}_T'(u) \\
&+ T\sum_{\tau=1}^{T-1}\left[Q_h^*\left(1,\frac{\tau}{T}\right)-Q_h^*\left(1,\frac{\tau+1}{T}\right)\right]\epsilon_T\left(\breve{\theta}_T-\theta_0\right)\mathcal{S}_\tau'(u) \\
&+ TQ_h^*(1,1)\epsilon_T\left(\breve{\theta}_T-\theta_0\right)\mathcal{S}_T'(u) + G\left(\breve{\theta}_T-\theta_0\right)\frac{1}{T}\sum_{\tau=1}^{T}\sum_{t=1}^{T}Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right)u_\tau' \\
&:= I_{11}+I_{12}+I_{13}+I_{14}+I_{15}
\end{aligned}$$

where ϵ_t satisfies $\sup_t\|\epsilon_t\| = o_p(1)$ by Assumption 4.

We now show that each of the five terms in the above expression is $o_p(1)$. It is easy to see that

$$\begin{aligned}
I_{15} &= G\left(\breve{\theta}_T-\theta_0\right)\frac{1}{T}\sum_{\tau=1}^{T}\sum_{t=1}^{T}Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right)u_\tau' \\
&= G\sqrt{T}\left(\breve{\theta}_T-\theta_0\right)\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}\left[\frac{1}{T}\sum_{t=1}^{T}Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right)\right]u_\tau' \\
&= G\sqrt{T}\left(\breve{\theta}_T-\theta_0\right)\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}\left[\int_0^1 Q_h^*\left(t,\frac{\tau}{T}\right)dt+O\left(\frac{1}{T}\right)\right]u_\tau' \\
&= G\sqrt{T}\left(\breve{\theta}_T-\theta_0\right)\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}\left[O\left(\frac{1}{T}\right)\right]u_\tau' = o_p(1),
\end{aligned}$$

$$
\begin{aligned}
I_{14} &= TQ_h^*(1,1)\,\epsilon_T\left(\breve{\theta}_T-\theta_0\right)\mathcal{S}_T'(u)\\
&= Q_h^*(1,1)\,\epsilon_T\sqrt{T}\left(\breve{\theta}_T-\theta_0\right)\sqrt{T}\mathcal{S}_T'(u)=o_p(1)\,,
\end{aligned}
$$

and

$$
\begin{aligned}
I_{13} &= T\sum_{\tau=1}^{T-1}\left[Q_h^*\left(1,\frac{\tau}{T}\right)-Q_h^*\left(1,\frac{\tau+1}{T}\right)\right]\epsilon_T\left(\breve{\theta}_T-\theta_0\right)\mathcal{S}_\tau'(u)\\
&= \epsilon_T\left[\sqrt{T}\left(\breve{\theta}_T-\theta_0\right)\right]\left[\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}Q_h^*\left(1,\frac{\tau}{T}\right)u_\tau\right]\\
&\quad+\epsilon_T\left[\sqrt{T}\left(\breve{\theta}_T-\theta_0\right)\right]Q_h^*(1,1)\left[\sqrt{T}\mathcal{S}_T(u)\right]\\
&= o_p(1)\,.
\end{aligned}
$$

To show that $I_{12}=o_p(1)$, we note that under the piecewise monotonicity condition in Assumption 1, $Q_h^*(t/T,1)$ is piecewise monotonic in t/T. So for some finite J we can partition the set $\{1,2,...,T\}$ into J maximal non-overlapping subsets $\cup_{j=1}^J\mathcal{I}_j$ such that $Q_h^*(t/T,1)$ is monotonic on each $\mathcal{I}_j:=\{\mathcal{I}_{jL},...,\mathcal{I}_{jU}\}$. Now

$$
\begin{aligned}
&\left\|\sum_{t=1}^{T-1}\left[Q_h^*\left(\frac{t}{T},1\right)-Q_h^*\left(\frac{t+1}{T},1\right)\right]\epsilon_t\right\|\\
\leq&\ \left\|\sum_{j=1}^{J}\sum_{t\in\mathcal{I}_j}\left[Q_h^*\left(\frac{t}{T},1\right)-Q_h^*\left(\frac{t+1}{T},1\right)\right]\epsilon_t\right\|+o_p(1)\\
\leq&\ \sum_{j=1}^{J}\sum_{t\in\mathcal{I}_j}\left|Q_h^*\left(\frac{t}{T},1\right)-Q_h^*\left(\frac{t+1}{T},1\right)\right|\sup_s\|\epsilon_s\|+o_p(1)\\
=&\ \sum_{j=1}^{J}(\pm)_j\sum_{t\in\mathcal{I}_j}\left[Q_h^*\left(\frac{t+1}{T},1\right)-Q_h^*\left(\frac{t}{T},1\right)\right]\sup_s\|\epsilon_s\|+o_p(1)\\
=&\ \sum_{j=1}^{J}\left|\left[Q_h^*\left(\frac{\mathcal{I}_{jU}}{T},1\right)-Q_h^*\left(\frac{\mathcal{I}_{jL}}{T},1\right)\right]\right|\sup_s\|\epsilon_s\|+o_p(1)\\
=&\ O(1)\sup_s\|\epsilon_s\|+o_p(1)=o_p(1)
\end{aligned}
$$

where the $o_p(1)$ term in the first inequality reflects the case when t and $t+1$ belong to different partitions and "$(\pm)_j$" takes "+" or "−" depending on whether $Q_h^*(t/T,1)$ is increasing or decreasing on the interval $[\mathcal{I}_{jL},\mathcal{I}_{jU}]$. As a result,

$$
\begin{aligned}
&\|I_{12}\|=\left\|T\sum_{t=1}^{T-1}\left[Q_h^*\left(\frac{t}{T},1\right)-Q_h^*\left(\frac{t+1}{T},1\right)\right]\epsilon_t\left(\breve{\theta}_T-\theta_0\right)\mathcal{S}_T'(u)\right\|\\
\leq&\ \left\|\sum_{t=1}^{T-1}\left[Q_h^*\left(\frac{t}{T},1\right)-Q_h^*\left(\frac{t+1}{T},1\right)\right]\epsilon_t\right\|\left\|\sqrt{T}\left(\breve{\theta}_T-\theta_0\right)\right\|\left\|\sqrt{T}\mathcal{S}_T'(u)\right\|=o_p(1)\,. \quad (16)
\end{aligned}
$$

To show that $I_{11}=o_p(1)$, we use a similar argument. Under the piecewise monotonicity condition, we can partition the set of lattice points $\{(i,j):i,j=1,2,...,T\}$ into a finite number of maximal non-overlapping subsets

$$\mathcal{I}_\ell = \{(i,j) : \mathcal{I}^*_{\ell 1L} \le i \le \mathcal{I}^*_{\ell 1U}, \mathcal{I}^*_{\ell 2L} \le j \le \mathcal{I}^*_{\ell 2U}\}$$

for $\ell = 1, ..., \ell_{\max}$ such that $\nabla Q^*_h(i/T, j/T)$ have the same sign for all $(i,j) \in \mathcal{I}_\ell$. Now

$$\begin{aligned}
\|I_{11}\| &= \left\| \sum_{\tau=1}^{T-1} \sum_{t=1}^{T-1} \nabla Q^*_h\left(\frac{t}{T}, \frac{\tau}{T}\right) \epsilon_t \left[\sqrt{T}\left(\breve{\theta}_T - \theta_0\right)\right] \left[\sqrt{T}\mathcal{S}'_\tau(u)\right] \right\| \\
&= \left\| \sum_{\ell=1}^{\ell_{\max}} \sum_{(i,j)\in\mathcal{I}_\ell} \nabla Q^*_h\left(\frac{i}{T}, \frac{j}{T}\right) \epsilon_i \left[\sqrt{T}\left(\breve{\theta}_T - \theta_0\right)\right] \left[\sqrt{T}\mathcal{S}'_j(u)\right] \right\| \\
&= \left\| \sum_{\ell=1}^{\ell_{\max}} (\pm)_\ell \sum_{(i,j)\in\mathcal{I}_\ell} \nabla Q^*_h\left(\frac{i}{T}, \frac{j}{T}\right) \right\| \sup_t \|\epsilon_t\| \times \left\| \sqrt{T}\left(\breve{\theta}_T - \theta_0\right) \right\| \times \left[\sup_\tau \left\| \sqrt{T}\mathcal{S}'_\tau(u) \right\| \right] \\
&= \sum_{\ell=1}^{\ell_{\max}} \left| Q^*_h\left(\frac{\mathcal{I}^*_{\ell 1L}}{T}, \frac{\mathcal{I}^*_{\ell 2L}}{T}\right) - Q^*_h\left(\frac{\mathcal{I}^*_{\ell 1U}}{T}, \frac{\mathcal{I}^*_{\ell 2L}}{T}\right) - Q^*_h\left(\frac{\mathcal{I}^*_{\ell 1L}}{T}, \frac{\mathcal{I}^*_{\ell 2U}}{T}\right) + Q^*_h\left(\frac{\mathcal{I}^*_{\ell 1U}}{T}, \frac{\mathcal{I}^*_{\ell 2U}}{T}\right) \right| \times o_p(1) \\
&= o_p(1).
\end{aligned}$$

We have therefore proved that $I_1 = o_p(1)$.

We can use similar arguments to show that $I_2 = o_p(1)$. Details are omitted. This completes the proof of Part (a).

Parts (b) and (c). We start by establishing the series representation of the weighting function as given in (4). The representation trivially holds with $\Phi_j(\cdot) = \phi_j(\cdot)$ for the OS-HAR variance estimator. It remains to show that it also holds for the contracted kernel $k_b(\cdot)$ and the exponentiated kernel $k^\rho(\cdot)$. We focus on the former as the proof for the latter is similar. Under Assumption 1(a), $k_b(x)$ has a Fourier cosine series expansion

$$k_b(x) = c + \sum_{j=1}^{\infty} \tilde{\lambda}_j \cos j\pi x \tag{17}$$

where c is the constant term in the expansion, $\sum_{j=1}^{\infty} \left|\tilde{\lambda}_j\right| < \infty$, and the right hand side converges uniformly over $x \in [-1, 1]$. This implies that

$$k_b(r-s) = c + \sum_{j=1}^{\infty} \tilde{\lambda}_j \cos j\pi r \cos j\pi s + \sum_{j=1}^{\infty} \tilde{\lambda}_j \sin j\pi r \sin j\pi s$$

and the right hand side converges uniformly over $(r,s) \in [0,1] \times [0,1]$. Now

$$\begin{aligned}
k^*_h(r,s) &= k_b(r,s) - \int_0^1 k_b(\tau - s)\, d\tau - \int_0^1 k_b(r-\tau)\, d\tau + \int_0^1 k_b(\tau_1 - \tau_2)\, d\tau_1 d\tau_2 \qquad (18) \\
&= \sum_{j=1}^{\infty} \tilde{\lambda}_j \cos j\pi r \cos j\pi s + \sum_{j=1}^{\infty} \tilde{\lambda}_j \left(\sin j\pi r - \int_0^1 \sin j\pi\tau d\tau\right)\left(\sin j\pi s - \int_0^1 \sin j\pi\tau d\tau\right) \\
&:= \sum_{j=1}^{\infty} \lambda_j \Phi_j(r) \Phi_j(s)
\end{aligned}$$

by taking

$$\Phi_j(r) = \begin{cases} \cos\left(\frac{1}{2}\pi j r\right), & j \text{ is even} \\ \sin\left(\frac{1}{2}\pi (j+1) r\right) - \int_0^1 \left(\sin \frac{1}{2}\pi (j+1)\tau\right) d\tau, & j \text{ is odd} \end{cases}$$

and

$$\lambda_j = \begin{cases} \tilde{\lambda}_{j/2}, & j \text{ is even} \\ \tilde{\lambda}_{(j+1)/2}, & j \text{ is odd} \end{cases}.$$

Obviously $\Phi_j(r)$ is continuously differentiable and satisfies $\int_0^1 \Phi_j(r)\,dr = 0$ for all j. The uniform convergence of $\sum_{j=1}^{\infty} \lambda_j \Phi_j(r) \Phi_j(s)$ to $k_h^*(r,s)$ inherits from the uniform convergence of the Fourier series in (17).

In view of part (a), it suffices to show that $W_T(\theta_0) \overset{a}{\sim} W_{eT}$. Using the Cramér-Wold device, it suffices to show that

$$tr\left[W_T(\theta_0)A\right] \overset{a}{\sim} tr\left[W_{eT}A\right]$$

for any conformable symmetry matrix A. Note that

$$tr\left[W_T(\theta_0)A\right] = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) u_t' A u_\tau$$

$$tr\left(W_{eT}A\right) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) (\Lambda e_t)' A (\Lambda e_t).$$

We write $tr\left[W_T(\theta_0)A\right]$ as

$$\begin{aligned} tr\left[W_T(\theta_0)A\right] &= \sum_{j=1}^{\infty} \lambda_j \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)u_t\right)' A \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)u_t\right) \\ &:= \xi_{T,J} + \eta_{T,J} \end{aligned}$$

where

$$\xi_{T,J} = \sum_{j=1}^{J} \lambda_j \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)u_t\right]' A \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)u_t\right],$$

$$\eta_{T,J} = \sum_{j=J+1}^{\infty} \lambda_j \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)u_t\right]' A \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)u_t\right].$$

We proceed to apply Lemma 1 by verifying the four conditions. First, it follows from Assumption 5 that for every J fixed,

$$\xi_{T,J} \overset{a}{\sim} \xi_{T,J}^* := \sum_{j=1}^{J} \lambda_j \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)\Lambda e_t\right]' A \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)\Lambda e_t\right] \tag{19}$$

as $T \to \infty$. Second, let

$$\begin{aligned} \xi_T &= \sum_{j=1}^{\infty} \lambda_j \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)\Lambda e_t\right]' A \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)\Lambda e_t\right] \\ &= tr\left(W_T^* A\right) \end{aligned}$$

and $A = \sum_{\ell=1}^{m} \mu_\ell a_\ell a_\ell'$ be the spectral decomposition of A. Then

$$\xi_T - \xi_{T,J}^* = \sum_{j=J+1}^{\infty} \lambda_j \sum_{\ell=1}^{m} \mu_\ell \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_j \left(\frac{t}{T} \right) a_\ell' \Lambda e_t \right]^2 .$$

So

$$E \left| \xi_T - \xi_{T,J}^* \right| \le C \sum_{j=J+1}^{\infty} |\lambda_j| \frac{1}{T} \sum_{t=1}^{T} \Phi_j^2 \left(\frac{t}{T} \right) \le C \sum_{j=J+1}^{\infty} |\lambda_j|$$

for some constant C that does not depend on T. That is, $E \left| \xi_T - \xi_{T,J}^* \right| \to 0$ uniformly in T as $J \to \infty$. So for any $\delta > 0$ and $\varepsilon > 0$, there exists a J_0 that does not depend on T such that

$$P \left(\left| \xi_T - \xi_{T,J}^* \right| \ge \delta \right) \le E \left| \xi_T - \xi_{T,J}^* \right| / \delta \le \varepsilon$$

for all $J \ge J_0$ and all T. It then follows that there exists a T_0 which does not depend on J_0 such that

$$\begin{aligned} P \left(\xi_{T,J}^* < \xi \right) &= P \left(\xi_T + \left(\xi_{T,J}^* - \xi_T \right) < \xi \right) \\ &\le P \left(\xi_T < \xi + \delta \right) + P \left(\left| \xi_T - \xi_{T,J}^* \right| \ge \delta \right) \\ &\le P \left(\xi_T < \xi + \delta \right) + \varepsilon \le P \left(\xi_T < \xi \right) + 2\varepsilon \end{aligned}$$

for $J \ge J_0$ and $T \ge T_0$. Here we have used the equicontinuity of the CDF of ξ_T when T is sufficiently large. This is verified below. Similarly, we can show that for $J \ge J_0$ and $T \ge T_0$ we have

$$P \left(\xi_{T,J}^* < \xi \right) \ge P \left(\xi_T < \xi \right) - 2\varepsilon.$$

Since the above two inequalities hold for all $\varepsilon > 0$, it must be true that $P \left(\xi_{T,J}^* < \xi \right) - P \left(\xi_T < \xi \right) = o(1)$ uniformly over sufficiently large T as $J \to \infty$.

Third, we can use Lemma 1 to show that ξ_T converges in distribution to ξ_∞ where

$$\begin{aligned} \xi_\infty &= \sum_{j=1}^{\infty} \lambda_j \left[\int_0^1 \Phi_j (r) \Lambda dB_m (r) \right]' A \left[\int_0^1 \Phi_j (s) \Lambda dB_m (s) \right] \\ &= \int_0^1 \int_0^1 Q_h^* (r, s) \left[\Lambda dB_m (r) \right]' A \Lambda dB_m (s) . \end{aligned}$$

Details are omitted here. So for any ξ and δ, we have

$$P \left(\xi - \delta \le \xi_T < \xi + \delta \right) = P \left(\xi - \delta \le \xi_\infty < \xi + \delta \right) + o(1)$$

as $T \to \infty$. Given the continuity of the CDF of ξ_∞, for any $\varepsilon > 0$, we can find a $\delta > 0$ such that $P \left(\xi - \delta \le \xi_T < \xi + \delta \right) \le \varepsilon$ for all T when T is sufficiently large. We have thus verified Condition (iii) in Lemma 1. Finally, for any $x \in \mathbb{R}$, we have

$$\begin{aligned} P \left(\left| \eta_{T,J} \right| > x \right) &\le P \left(\sum_{j=J+1}^{\infty} |\lambda_j| \sum_{\ell=1}^{m} |\mu_\ell| \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_j \left(\frac{t}{T} \right) a_\ell' u_t \right]^2 > x \right) \\ &\le x^{-1} \sum_{\ell=1}^{m} |\mu_\ell| \, E \sum_{j=J+1}^{\infty} |\lambda_j| \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_j \left(\frac{t}{T} \right) a_\ell' u_t \right]^2 . \end{aligned}$$

But by Assumption 3,

$$
\begin{aligned}
&E\sum_{j=J+1}^{\infty}|\lambda_j|\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)a_\ell' u_t\right]^2 \\
&=\sum_{j=J+1}^{\infty}|\lambda_j|\frac{1}{T}\sum_{t=1}^{T}\Phi_j^2\left(\frac{t}{T}\right)E\left(a_\ell' u_t\right)^2+\sum_{j=J+1}^{\infty}|\lambda_j|\frac{1}{T}\sum_{t\neq s}\Phi_j\left(\frac{t}{T}\right)\Phi_j\left(\frac{s}{T}\right)Ea_\ell' u_t u_s' a_\ell \\
&\leq C\sum_{j=J+1}^{\infty}|\lambda_j|+\sum_{j=J+1}^{\infty}|\lambda_j|\frac{1}{T}\sum_{t\neq s}\left|a_\ell'\Gamma_{t-s}a_\ell\right|=O\left(\sum_{j=J+1}^{\infty}|\lambda_j|\right)=o(1)
\end{aligned}
$$

uniformly over T, as $J\to\infty$. So we have proved that $\eta_{T,J}=o_p(1)$ uniformly over T, as $J\to\infty$.

Combining the above steps, we have

$$
tr\left[W_T(\theta_0)A\right]\overset{a}{\sim}\xi_T=tr\left(W_{eT}A\right),
$$

and

$$
tr\left(W_{eT}A\right)\overset{d}{\to}\int_0^1\int_0^1 Q_h^*(r,s)\left[\Lambda dB_m(r)\right]' A\Lambda dB_m(s)=tr(W_\infty A).
$$

As a consequence,

$$
W_T\left(\breve{\theta}_T\right)\overset{a}{\sim}W_{eT}\overset{d}{\to}W_\infty
$$

as desired. ■

Proof of Theorem 1. We prove part (b) only as the proofs for other parts are similar. Let $G_\Lambda=\Lambda^{-1}G$ be an $m\times d$ matrix, then

$$
\begin{aligned}
F_\infty=&\left\{R\left[G_\Lambda'\tilde{W}_\infty^{-1}G_\Lambda\right]^{-1}G_\Lambda'\tilde{W}_\infty^{-1}B_m(1)\right\}'\left\{R\left[G_\Lambda'\tilde{W}_\infty^{-1}G_\Lambda\right]^{-1}R'\right\}^{-1} \\
&\times\left\{R\left[G_\Lambda'\tilde{W}_\infty^{-1}G_\Lambda\right]^{-1}G_\Lambda'\tilde{W}_\infty^{-1}B_m(1)\right\}/p.
\end{aligned}
$$

Let $G_\Lambda=U_{m\times m}\Sigma_{m\times d}V_{d\times d}'$ be the singular value decomposition of G_Λ. By definition, $U'U=UU'=I_m$, $VV'=V'V=I_d$ and

$$
\Sigma=\left[\begin{array}{c}A_{d\times d}\\ O_{q\times d}\end{array}\right],
$$

where A is a diagonal matrix with singular values on the main diagonal and O is a matrix of zeros. Then

$$
\begin{aligned}
&R\left[G_\Lambda'\tilde{W}_\infty^{-1}G_\Lambda\right]^{-1}G_\Lambda'\tilde{W}_\infty^{-1}B_m(1)=R\left[V\Sigma' U'\tilde{W}_\infty^{-1}U\Sigma V'\right]^{-1}V\Sigma' U'\tilde{W}_\infty^{-1}B_m(1) \\
&=RV\left[\Sigma' U'\tilde{W}_\infty^{-1}U\Sigma\right]^{-1}\Sigma' U'\tilde{W}_\infty^{-1}B_m(1)=RV\left[\Sigma' U'\tilde{W}_\infty^{-1}U\Sigma\right]^{-1}\Sigma'\left[U'\tilde{W}_\infty^{-1}U\right]\left[U'B_m(1)\right] \\
&\overset{d}{=}RV\left[\Sigma'\tilde{W}_\infty^{-1}\Sigma\right]^{-1}\Sigma'\tilde{W}_\infty^{-1}B_m(1),
\end{aligned}
$$

and

$$
R\left[G_\Lambda'\tilde{W}_\infty^{-1}G_\Lambda\right]^{-1}R'=R\left[V\Sigma' U'\tilde{W}_\infty^{-1}U\Sigma V'\right]^{-1}R'\overset{d}{=}RV\left[\Sigma'\tilde{W}_\infty^{-1}\Sigma\right]^{-1}V'R',
$$

where we have used the distributional equivalence of $[\tilde{W}_\infty^{-1}, B_m(1)]$ and $[U'\tilde{W}_\infty^{-1}U, U'B_m(1)]$. Let

$$\tilde{W}_\infty = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \text{ and } \tilde{W}_\infty^{-1} = \begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix}$$

where C_{11} and C^{11} are $d\times d$ matrices, C_{22} and C^{22} are $q\times q$ matrices, and $C_{12} = C_{21}'$, $C^{12} = (C^{21})'$. By definition

$$C_{11} = \int_0^1 \int_0^1 Q_h^*(r,s)dB_d(r)dB_d(s)' = \begin{pmatrix} C_{pp} & C_{p,d-p} \\ C_{p,d-p}' & C_{d-p,d-p} \end{pmatrix} \tag{20}$$

$$C_{12} = \int_0^1 \int_0^1 Q_h^*(r,s)dB_d(r)dB_q(s)' = \begin{pmatrix} C_{pq} \\ C_{d-p,q} \end{pmatrix} \tag{21}$$

$$C_{22} = \int_0^1 \int_0^1 Q_h^*(r,s)dB_q(r)dB_q(s)' = C_{qq} \tag{22}$$

where C_{pp}, C_{pq}, and C_{qq} are defined in (7), and $C_{d-p,d-p}$, $C_{p,d-p}$ and $C_{d-p,q}$ are similarly defined. It follows from the partitioned inverse formula that

$$C^{11} = \left[C_{11} - C_{12}C_{22}^{-1}C_{21}\right]^{-1}, \; C^{12} = -C^{11}C_{12}C_{22}^{-1}.$$

With the above partition of $\tilde{W}_\infty^{-1}$, we have

$$\begin{aligned} \left[\Sigma'\tilde{W}_\infty^{-1}\Sigma\right]^{-1} &= \left\{ \begin{pmatrix} A' & O' \end{pmatrix} \begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix} \begin{pmatrix} A \\ O \end{pmatrix} \right\}^{-1} \\ &= \left[A'C^{11}A\right]^{-1} = A^{-1}\left(C^{11}\right)^{-1}\left(A'\right)^{-1} \end{aligned}$$

and so

$$\begin{aligned} & RV\left[\Sigma'\tilde{W}_\infty^{-1}\Sigma\right]^{-1}\Sigma'\tilde{W}_\infty^{-1} \\ &= RVA^{-1}\left(C^{11}\right)^{-1}\left(A'\right)^{-1}\begin{pmatrix} A' & O' \end{pmatrix}\begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix} \\ &= RVA^{-1}\left(C^{11}\right)^{-1}\left(A'\right)^{-1}A'\begin{pmatrix} C^{11} & C^{12} \end{pmatrix} \\ &= RVA^{-1}\begin{pmatrix} I_d, & \left(C^{11}\right)^{-1}C^{12} \end{pmatrix}, \end{aligned} \tag{23}$$

and

$$RV\left[\Sigma'\tilde{W}_\infty^{-1}\Sigma\right]^{-1}V'R' = RVA^{-1}\left(C^{11}\right)^{-1}\left(A'\right)^{-1}V'R'.$$

As a result,

$$\begin{aligned} F_\infty &\overset{d}{=} \left[RVA^{-1}\begin{pmatrix} I_d, & \left(C^{11}\right)^{-1}C^{12} \end{pmatrix}B_m(1)\right]'\left(RVA^{-1}\left(C^{11}\right)^{-1}\left(A'\right)^{-1}V'R'\right)^{-1} \\ &\times \left[RVA^{-1}\begin{pmatrix} I_d, & \left(C^{11}\right)^{-1}C^{12} \end{pmatrix}B_m(1)\right]/p \\ &= \left[RVA^{-1}\begin{pmatrix} I_d, & -C_{12}C_{22}^{-1} \end{pmatrix}B_m(1)\right]'\left(RVA^{-1}\left(C^{11}\right)^{-1}\left(A'\right)^{-1}V'R'\right)^{-1} \\ &\times \left[RVA^{-1}\begin{pmatrix} I_d, & -C_{12}C_{22}^{-1} \end{pmatrix}B_m(1)\right]/p. \end{aligned}$$

Let $B_m(1) = [B_d'(1), B_q'(1)]'$ and

$$RVA^{-1} = \tilde{U}_{p\times p}\tilde{\Sigma}_{p\times d}\tilde{V}_{d\times d}'$$

be the SVD of RVA^{-1}, where

$$\tilde{\Sigma}_{p\times d} = \left(\begin{array}{cc} \tilde{A}_{p\times p}, & \tilde{O}_{p\times(d-p)} \end{array} \right).$$

Then

$$\begin{aligned}
F_\infty &\overset{d}{=} \left\{\tilde{U}\tilde{\Sigma}\tilde{V}'\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]\right\}'\left(\tilde{U}\tilde{\Sigma}\tilde{V}'\left(C^{11}\right)^{-1}\tilde{V}\tilde{\Sigma}'\tilde{U}'\right)^{-1} \\
&\times \tilde{U}\tilde{\Sigma}\tilde{V}'\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]/p \\
&= \left\{\tilde{\Sigma}\tilde{V}'\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]\right\}'\left(\tilde{\Sigma}\tilde{V}'C_{11}^{-1}\tilde{V}\tilde{\Sigma}'\right)^{-1} \\
&\times \tilde{\Sigma}\tilde{V}'\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]/p \\
&= \left\{\tilde{\Sigma}\left[\tilde{V}'B_d(1) - \tilde{V}'C_{12}C_{22}^{-1}B_q(1)\right]\right\}'\left(\tilde{\Sigma}\tilde{V}'\left(C^{11}\right)^{-1}\tilde{V}\tilde{\Sigma}'\right)^{-1} \\
&\times \tilde{\Sigma}\left[\tilde{V}'B_d(1) - \tilde{V}'C_{12}C_{22}^{-1}B_q(1)\right].
\end{aligned}$$

Noting that the joint distribution of $\left[\tilde{V}'B_d(1), \tilde{V}'C_{12}, C_{22}, \ \tilde{V}'\left(C^{11}\right)^{-1}\tilde{V}\right]$ is invariant to the orthonormal matrix $\tilde{V}'$, we have

$$\begin{aligned}
F_\infty &\overset{d}{=} \left\{\tilde{\Sigma}\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]\right\}'\left(\tilde{\Sigma}\left(C^{11}\right)^{-1}\tilde{\Sigma}'\right)^{-1} \\
&\times \tilde{\Sigma}\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]/p \\
&= \left\{\left(\begin{array}{cc} \tilde{A}, & \tilde{O} \end{array} \right)\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]\right\}' \\
&\times \left(\left(\begin{array}{cc} \tilde{A}, & \tilde{O} \end{array} \right)\left(C^{11}\right)^{-1}\left(\begin{array}{cc} \tilde{A}, & \tilde{O} \end{array} \right)'\right)^{-1} \\
&\times \left\{\left(\begin{array}{cc} \tilde{A}, & \tilde{O} \end{array} \right)\left[B_d(1) - C_{12}C_{22}^{-1}B_q(1)\right]\right\}/p.
\end{aligned}$$

Writing

$$\left(C^{11}\right)^{-1} = C_{11} - C_{12}C_{22}^{-1}C_{21} = \left(\begin{array}{cc} D_{pp} & D^{12} \\ D^{21} & D^{22} \end{array} \right)$$

where $D_{pp} = C_{pp} - C_{pq}C_{qq}^{-1}C_{pq}'$ and D^{22} is a $(d-p)\times(d-p)$ matrix and using equations (20)–(22), we have

$$F_\infty \overset{d}{=} \left[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right]' D_{pp}^{-1}\left[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right]/p.$$

■

Proof of Theorem 2. Let $\boldsymbol{\lambda}_T$ be the $p\times 1$ vector of Lagrange multipliers for the constrained GMM estimation. The first order conditions for $\hat{\theta}_{T,R}$ are

$$\frac{\partial g_T'\left(\hat{\theta}_{T,R}\right)}{\partial\theta}W_T^{-1}(\tilde{\theta}_T)g_T\left(\hat{\theta}_{T,R}\right) + R_T'\boldsymbol{\lambda}_T = 0 \text{ and } R\hat{\theta}_{T,R} = r. \tag{24}$$

Linearizing the first set of conditions and using Assumption 4, we have the following system of equations:

$$\begin{pmatrix} \tilde{\Psi} & R' \\ R_{p\times d} & \mathbf{0}_{p\times p} \end{pmatrix} \begin{pmatrix} \sqrt{T}\left(\hat{\theta}_{T,R}-\theta_0\right) \\ \sqrt{T}\boldsymbol{\lambda}_T \end{pmatrix} = \begin{pmatrix} -G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T\left(\theta_0\right) \\ \mathbf{0}_{p\times 1} \end{pmatrix} + o_p\left(1\right)$$

where $\tilde{\Psi} := \tilde{\Psi}\left(\tilde{\theta}_T\right) = G'W_T^{-1}(\tilde{\theta}_T)G$. From this, we get

$$\begin{aligned} \sqrt{T}\left(\hat{\theta}_{T,R}-\theta_0\right) &= -\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T\left(\theta_0\right) \\ &\quad - \tilde{\Psi}^{-1}R'\left\{R\tilde{\Psi}^{-1}R'\right\}^{-1}R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T\left(\theta_0\right) + o_p\left(1\right) \end{aligned} \tag{25}$$

and

$$\sqrt{T}\boldsymbol{\lambda}_T = -\left\{R\tilde{\Psi}^{-1}R'\right\}^{-1}R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T\left(\theta_0\right) + o_p\left(1\right). \tag{26}$$

Combining (5) with (25), we have

$$\sqrt{T}\left(\hat{\theta}_T-\hat{\theta}_{T,R}\right) = \tilde{\Psi}^{-1}R'\left\{R\tilde{\Psi}^{-1}R'\right\}^{-1}R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T\left(\theta_0\right) + o_p\left(1\right) \tag{27}$$

which implies that $\sqrt{T}\left(\hat{\theta}_T-\hat{\theta}_{T,R}\right) = O_p\left(1\right)$. So

$$g_T\left(\hat{\theta}_T\right) = g_T\left(\hat{\theta}_{T,R}\right) + G_T(\hat{\theta}_T)\left(\hat{\theta}_T-\hat{\theta}_{T,R}\right) + o_p\left(1/\sqrt{T}\right).$$

Using this and the first order conditions for $\hat{\theta}_T : g_T'\left(\hat{\theta}_T\right)W_T^{-1}(\tilde{\theta}_T)G_T(\hat{\theta}_T) = 0$, we have

$$\begin{aligned} \mathbb{D}_T &= T\left(\hat{\theta}_T-\hat{\theta}_{T,R}\right)'G_T'(\hat{\theta}_T)W_T^{-1}(\breve{\theta}_T)G_T(\hat{\theta}_T)\left(\hat{\theta}_T-\hat{\theta}_{T,R}\right)/p \\ &\quad + 2Tg_T'\left(\hat{\theta}_T\right)W_T^{-1}(\tilde{\theta}_T)G(\hat{\theta}_T)\left(\hat{\theta}_T-\hat{\theta}_{T,R}\right)/p + o_p\left(1\right) \\ &= T\left(\hat{\theta}_T-\hat{\theta}_{T,R}\right)'G_T'(\hat{\theta}_T)W_T^{-1}(\breve{\theta}_T)G_T(\hat{\theta}_T)\left(\hat{\theta}_T-\hat{\theta}_{T,R}\right)/p + o_p\left(1\right). \end{aligned} \tag{28}$$

Invoking Lemma 2(a), we obtain

$$\mathbb{D}_T = T\left(\hat{\theta}_T-\hat{\theta}_{T,R}\right)'\tilde{\Psi}\left(\hat{\theta}_T-\hat{\theta}_{T,R}\right)/p + o_p\left(1\right). \tag{29}$$

Plugging (27) into (29) and simplifying the resulting expression, we have

$$\begin{aligned} \mathbb{D}_T &= \left[R'\left\{R\tilde{\Psi}^{-1}R'\right\}^{-1}R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T\left(\theta_0\right)\right]' \\ &\quad \times \tilde{\Psi}^{-1}\left[R'\left\{R\tilde{\Psi}^{-1}R'\right\}^{-1}R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T\left(\theta_0\right)\right]/p + o_p\left(1\right) \\ &= \left[R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T\left(\theta_0\right)\right]'\left[R\tilde{\Psi}^{-1}R'\right]^{-1}R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T\left(\theta_0\right)/p + o_p\left(1\right) \\ &= \sqrt{T}\left[R\left(\hat{\theta}_T-\theta_0\right)\right]'\left[R\tilde{\Psi}^{-1}R'\right]^{-1}\sqrt{T}R\left(\hat{\theta}_T-\theta_0\right)/p + o_p\left(1\right) \\ &= \mathbb{W}_T + o_p\left(1\right). \end{aligned} \tag{30}$$

Next, we prove the second result in the theorem. In view of the first order conditions in (24) and equation (26), we have

$$\begin{aligned}\sqrt{T}\Delta_T\left(\hat{\theta}_{T,R}\right) &= \frac{\partial g_T'\left(\hat{\theta}_{T,R}\right)}{\partial\theta}W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T\left(\hat{\theta}_{T,R}\right)+o_p(1) = -\sqrt{T}R'\boldsymbol{\lambda}_T+o_p(1)\\ &= R'\left\{R\tilde{\Psi}^{-1}R'\right\}^{-1}R\tilde{\Psi}^{-1}G'W_T^{-1}(\tilde{\theta}_T)\sqrt{T}g_T(\theta_0)+o_p(1)\\ &= \tilde{\Psi}\sqrt{T}\left(\hat{\theta}_T-\hat{\theta}_{T,R}\right)+o_p(1),\end{aligned}$$

and so

$$\begin{aligned}\mathbb{S}_T &= T\left(\hat{\theta}_T-\hat{\theta}_{T,R}\right)'\tilde{\Psi}\left(\hat{\theta}_T-\hat{\theta}_{T,R}\right)/p+o_p(1)\\ &= \mathbb{D}_T+o_p(1)=\mathbb{W}_T+o_p(1).\end{aligned}$$

■

Proof of Theorem 3. Part (a). Let

$$H=\left(\frac{B_p(1)-C_{pq}C_{qq}^{-1}B_q(1)}{\left\|B_p(1)-C_{pq}C_{qq}^{-1}B_q(1)\right\|},\bar{H}\right)$$

be an orthonormal matrix, then

$$\begin{aligned}F_\infty &= \left\|B_p(1)-C_{pq}C_{qq}^{-1}B_q(1)\right\|^2\left[\frac{B_p(1)-C_{pq}C_{qq}^{-1}B_q(1)}{\left\|B_p(1)-C_{pq}C_{qq}^{-1}B_q(1)\right\|}\right]'H\\ &\times H'D_{pp}^{-1}HH'\left[\frac{B_p(1)-C_{pq}C_{qq}^{-1}B_q(1)}{\left\|B_p(1)-C_{pq}C_{qq}^{-1}B_q(1)\right\|}\right]/p\\ &= \left\|B_p(1)-C_{pq}C_{qq}^{-1}B_q(1)\right\|^2e_p'\left[H'D_{pp}^{-1}H\right]e_p/p,\end{aligned}$$

where $e_p=(1,0,0,...,0)'\in\mathbb{R}^p$. But $H'D_{pp}^{-1}H$ has the same distribution as D_{pp}^{-1} and D_{pp} is independent of $B_p(1)-C_{pq}C_{qq}^{-1}B_q(1)$. So

$$F_\infty\overset{d}{=}\frac{\left\|B_p(1)-C_{pq}C_{qq}^{-1}B_q(1)\right\|^2/p}{\left[e_p'D_{pp}^{-1}e_p\right]^{-1}}. \tag{31}$$

That is, F_∞ is equal in distribution to a ratio of two independent random variables.

It is easy to see that

$$\left[e_p'D_{pp}^{-1}e_p\right]^{-1}\overset{d}{=}\left[e_{p+q}'\left[\int_0^1\int_0^1Q_h^*(r,s)\,dB_{p+q}(r)\,dB_{p+q}'(s)\right]^{-1}e_{p+q}\right]^{-1}\overset{d}{=}\frac{1}{K}\chi^2_{K-p-q+1}$$

where $e_{p+q}=(1,0,...,0)'\in\mathbb{R}^{p+q}$. With this, we can now represent F_∞ as

$$F_\infty\overset{d}{=}\frac{\chi_p^2\left(\Delta^2\right)/p}{\chi^2_{K-p-q+1}/K} \tag{32}$$

and so

$$\kappa^{-1}F_{\infty} = \frac{\chi_p^2\left(\Delta^2\right)/p}{\chi_{K-p-q+1}^2/\left(K-p-q+1\right)} \overset{d}{=} \mathcal{F}_{p,K-p-q+1}\left(\Delta^2\right). \tag{33}$$

Part (b). Since the numerator and the denominator in (33) are independent, $\kappa^{-1}F_{\infty}$ is distributed as a noncentral F distribution, conditional on Δ^2. More specifically, we have

$$P\left(\kappa^{-1}F_{\infty} < z\right) = P\left(\mathcal{F}_{p,K-p-q+1}\left(\Delta^2\right) < z\right) = EF_{p,K-p-q+1}\left(z,\Delta^2\right)$$

where $F_{p,K-p-q+1}(z,\lambda)$ is the CDF of the noncentral F distribution with degrees of freedom $(p, K-p-q+1)$ and noncentrality parameter λ, and $\mathcal{F}_{p,K-p-q+1}(\lambda)$ is a random variable with CDF $F_{p,K-p-q+1}(z,\lambda)$.

We proceed to compute the mean of Δ^2. Let

$$\xi_j = \int_0^1 \Phi_j(r)\,dB_p(r) \sim iidN(0,I_p) \text{ and } \eta_j = \int_0^1 \Phi_j(r)\,dB_q(r) \sim iidN(0,I_q).$$

Then we can represent C_{pq} and C_{qq} as

$$C_{pq} = K^{-1}\sum_{j=1}^{K}\xi_j\eta_j' \text{ and } C_{qq} = K^{-1}\sum_{j=1}^{K}\eta_j\eta_j'. \tag{34}$$

So

$$\begin{aligned}
E\Delta^2 &= EB_q(1)'\,C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}B_q(1) = Etr\left(C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}\right) \\
&= Etr\left[\left(\frac{1}{K}\sum_{j=1}^{K}\eta_j\eta_j'\right)^{-1}\left(\frac{1}{K}\sum_{j=1}^{K}\eta_j\xi_j'\right)\left(\frac{1}{K}\sum_{j=1}^{K}\xi_j\eta_j'\right)\left(\frac{1}{K}\sum_{j=1}^{K}\eta_j\eta_j'\right)^{-1}\right] \\
&= ptrE(\Pi),
\end{aligned}$$

where $\Pi := \left(\sum_{j=1}^{K}\eta_j\eta_j'\right)^{-1}$ follows an inverse-Wishart distribution and

$$E(\Pi) = \frac{I_q}{K-q-1}$$

for K large enough. Therefore $E\Delta^2 = \frac{pq}{K-q-1} = \delta^2$.

Next we compute the variance of Δ^2. It follows from the law of total variance that

$$var\left(\Delta^2\right) = E\left[var\left(\Delta^2|C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}\right)\right] + var\left(\left[tr\left(C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}\right)\right]\right).$$

Note that $B_q(1)$ is independent of $C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}$. So conditional on $C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}$, Δ^2 is a quadratic form in standard normals. Hence the conditional variance of Δ^2 is

$$var\left(\Delta^2|C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}\right) = 2tr\left(C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}\right).$$

Using the representation in (34), we have

$$
\begin{aligned}
&Etr\left(C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}\right)\\
&= Etr\left(\frac{1}{K}\sum_{j=1}^{K}\eta_j\xi_j'\right)\left(\frac{1}{K}\sum_{j=1}^{K}\xi_j\eta_j'\right)C_{qq}^{-2}\left(\frac{1}{K}\sum_{j=1}^{K}\eta_j\xi_j'\right)\left(\frac{1}{K}\sum_{j=1}^{K}\xi_j\eta_j'\right)C_{qq}^{-2}\\
&= Etr\left[\frac{1}{K^4}\sum_{j_1=1}^{K}\sum_{i_1=1}^{K}\eta_{j_1}\left(\xi_{j_1}'\xi_{i_1}\right)\eta_{i_1}'C_{qq}^{-2}\sum_{j_2=1}^{K}\sum_{i_2=1}^{K}\eta_{j_2}\left(\xi_{j_2}'\xi_{i_2}\right)\eta_{i_2}'C_{qq}^{-2}\right]\\
&= Etr\left[\frac{1}{K^4}\sum_{j_1=1}^{K}\sum_{i_1=1}^{K}\sum_{j_2=1}^{K}\sum_{i_2=1}^{K}\eta_{j_1}\eta_{i_1}'C_{qq}^{-2}\eta_{j_2}\eta_{i_2}'C_{qq}^{-2}E\left(\xi_{j_1}'\xi_{i_1}\right)\left(\xi_{j_2}'\xi_{i_2}\right)\right].
\end{aligned}
$$

Since

$$
E\left(\xi_{j_1}'\xi_{i_1}\right)\left(\xi_{j_2}'\xi_{i_2}\right)=\begin{cases} p^2, & j_1=i_1, j_2=i_2, \text{ and } j_1\neq j_2,\\ p, & j_1=j_2, i_1=i_2, \text{ and } j_1\neq i_1,\\ p & j_1=i_2, i_1=j_2, \text{ and } j_1\neq i_1,\\ p^2+2p, & j_1=j_2=i_1=i_2,\\ 0, & \text{otherwise,}\end{cases}
$$

we have

$$
\begin{aligned}
&Etr\left(C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}\right)\\
&= Etr\left[\frac{1}{K^4}\sum_{j_1=1}^{K}\sum_{j_2=1}^{K}\eta_{j_1}\eta_{j_1}'C_{qq}^{-2}\eta_{j_2}\eta_{j_2}'C_{qq}^{-2}\right]p^2+Etr\left[\frac{1}{K^4}\sum_{j_1=1}^{K}\sum_{i_1=1}^{K}\eta_{j_1}\left(\eta_{i_1}'C_{qq}^{-2}\eta_{j_1}\right)\eta_{i_1}'C_{qq}^{-2}\right]p\\
&+Etr\left[\frac{1}{K^4}\sum_{j_1=1}^{K}\sum_{i_1=1}^{K}\eta_{j_1}\left(\eta_{i_1}'C_{qq}^{-2}\eta_{i_1}\right)\eta_{j_1}'C_{qq}^{-2}\right]p\\
&= Etr\left(\sum_{\ell_1=1}^{K}\eta_{\ell_1}\eta_{\ell_1}'\right)^{-2}\left(p^2+p\right)+Etr\left[\frac{1}{K^2}\sum_{j_1=1}^{K}\eta_{j_1}\eta_{j_1}'C_{qq}^{-2}\right]tr\left[\frac{1}{K^2}\sum_{i_1=1}^{K}\eta_{i_1}'C_{qq}^{-2}\eta_{i_1}\right]p\\
&= E\left[tr\left(\Pi^2\right)\right]\left(p^2+p\right)+E\left[tr\left(\Pi\right)\right]^2p\\
&= \left(\sum_{i=1}^{q}\sum_{j=1}^{q}E\Pi_{ij}^2\right)\left(p^2+p\right)+E\left(\sum_{i=1}^{q}\Pi_{ii}\right)^2p.
\end{aligned}
$$

It follows from Theorem 5.2.2 of Press (2005, pp. 119, using the notation here) that

$$
E\Pi_{ij}^2=\frac{\left(K-q+1\right)\delta_{ij}+\left(K-q-1\right)}{\left(K-q\right)\left(K-q-1\right)^2\left(K-q-3\right)}+\frac{\delta_{ij}}{\left[K-q-1\right]^2}=O(\frac{1}{K^2})
$$

and for $i\neq j$,

$$
E\Pi_{ii}\Pi_{jj}=\frac{2}{\left(K-q\right)\left(K-q-1\right)^2\left(K-q-3\right)}+\frac{1}{\left[K-q-1\right]^2}=O(\frac{1}{K^2}).
$$

Hence

$$
Etr\left(C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}\right)=O(\frac{1}{K^2}). \tag{35}
$$

Next

$$
\begin{aligned}
& var\left(\left[tr\left(C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}\right)\right]\right) \\
& \leq E\left[tr\left(C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}\right) tr\left(C_{qq}^{-1}C_{pq}'C_{pq}C_{qq}^{-1}\right)\right] \\
& = Etr\left[\left(\frac{1}{K}\sum_{j=1}^{K}\eta_j\xi_j'\right)\left(\frac{1}{K}\sum_{j=1}^{K}\xi_j\eta_j'\right)C_{qq}^{-2}\right] tr\left[\left(\frac{1}{K}\sum_{j=1}^{K}\eta_j\xi_j'\right)\left(\frac{1}{K}\sum_{j=1}^{K}\xi_j\eta_j'\right)C_{qq}^{-2}\right] \\
& = Etr\left[\frac{1}{K}\sum_{i_1=1}^{K}\sum_{j_1=1}^{K}\eta_{i_1}\left(\xi_{i_1}'\xi_{j_1}\right)\eta_{j_1}'C_{qq}^{-2}\right] tr\left[\frac{1}{K}\sum_{i_2=1}^{K}\sum_{j_2=1}^{K}\eta_{i_2}\left(\xi_{i_2}'\xi_{j_2}\right)\eta_{j_2}'C_{qq}^{-2}\right] \\
& = E\frac{1}{K^2}\sum_{i_1=1}^{K}\sum_{j_1=1}^{K}\sum_{i_2=1}^{K}\sum_{j_2=1}^{K} tr\left(\eta_{i_1}\eta_{j_1}'C_{qq}^{-2}\right) tr\left(\eta_{i_2}\eta_{j_2}'C_{qq}^{-2}\right) E\left[\left(\xi_{i_1}'\xi_{j_1}\right)\left(\xi_{i_2}'\xi_{j_2}\right)\right] \\
& = E\left[tr\left(\Pi\right)\right]^2 p^2 + E\frac{1}{K^2}\sum_{i_1=1}^{K}\sum_{j_1=1}^{K} tr\left[C_{qq}^{-2}\eta_{j_1}\eta_{j_1}'C_{qq}^{-2}\eta_{i_1}\eta_{i_1}'\right]2p \\
& = E\left[tr\left(\Pi\right)\right]^2 p^2 + Etr\left[\Pi^2\right]2p.
\end{aligned}
$$

Using the same formulae from Press (2005), we can show that the last term is of order $O\left(K^{-2}\right).$ This, combined with (35), leads to $var\left(\Delta^2\right) = O(K^{-2})$.

Taking a Taylor expansion and using the mean and variance of Δ^2, we have

$$
\begin{aligned}
& P\left(\kappa^{-1}F_\infty < z\right) \\
& = EF_{p,K-p-q+1}\left(z,\Delta^2\right) \\
& = EF_{p,K-p-q+1}\left(z,\delta^2\right) + E\frac{\partial F_{p,K-p-q+1}\left(z,\delta^2\right)}{\partial\lambda}\left(\Delta^2-\delta^2\right) + E\frac{\partial^2 F_{p,K-p-q+1}\left(z,\tilde{\delta}^2\right)}{\partial\lambda^2}\left(\Delta^2-\delta^2\right)^2 \qquad (36) \\
& = EF_{p,K-p-q+1}\left(z,\delta^2\right) + E\frac{\partial^2 F_{p,K-p-q+1}\left(z,\tilde{\delta}^2\right)}{\partial\lambda^2}\left(\Delta^2-\delta^2\right)^2
\end{aligned}
$$

for some $\tilde{\Delta}^2$ between Δ^2 and δ^2. By definition,

$$
\begin{aligned}
F_{p,K-p-q+1}\left(z,\lambda\right) &= P\left(\frac{\chi_p^2\left(\lambda\right)}{p}\left[\frac{\chi_{K-p-q+1}^2}{K-p-q+1}\right]^{-1} < z\right) \\
&= E\mathcal{G}_p\left(pz\left[\frac{\chi_{K-p-q+1}^2}{K-p-q+1}\right],\lambda\right)
\end{aligned}
$$

where $\mathcal{G}_p(z,\lambda)$ is the CDF of the noncentral chi-square distribution $\chi_p^2(\lambda)$ with noncentrality parameter λ. In view of the relationship $\mathcal{G}_p(z,\lambda) = \exp\left(-\frac{\lambda}{2}\right)\sum_{j=0}^{\infty}\frac{(\lambda/2)^j}{j!}\mathcal{G}_{p+2j}(z)$, we have

$$\begin{aligned}
&\frac{\partial F_{p,K-p-q+1}(z,\lambda)}{\partial\lambda} \\
&= \sum_{j=0}^{\infty}\frac{\partial}{\partial\lambda}\left[\exp\left(-\frac{\lambda}{2}\right)\frac{(\lambda/2)^j}{j!}\right]E\mathcal{G}_{p+2j}\left(pz\left[\frac{\chi^2_{K-p-q+1}}{K-p-q+1}\right]\right) \\
&= -\frac{1}{2}\exp\left(-\frac{\lambda}{2}\right)\sum_{j=0}^{\infty}\frac{(\lambda/2)^j}{j!}E\mathcal{G}_{p+2j}\left(pz\left[\frac{\chi^2_{K-p-q+1}}{K-p-q+1}\right]\right) \\
&+ \frac{1}{2}\exp\left(-\frac{\lambda}{2}\right)\sum_{j=0}^{\infty}\frac{(\lambda/2)^j}{j!}E\mathcal{G}_{p+2+2j}\left(pz\left[\frac{\chi^2_{K-p-q+1}}{K-p-q+1}\right]\right) \\
&= \frac{1}{2}\sum_{j=0}^{\infty}\exp\left(-\frac{\lambda}{2}\right)\frac{(\lambda/2)^j}{j!}E\left[\mathcal{G}_{p+2+2j}\left(pz\left[\frac{\chi^2_{K-p-q+1}}{K-p-q+1}\right]\right) - \mathcal{G}_{p+2j}\left(pz\left[\frac{\chi^2_{K-p-q+1}}{K-p-q+1}\right]\right)\right]
\end{aligned}$$

and so

$$\begin{aligned}
\left|\frac{\partial^2 F_{p,K-p-q+1}(z,\lambda)}{\partial\lambda^2}\right| &\leq \frac{1}{2}\sum_{j=0}^{\infty}\left|\frac{\partial}{\partial\lambda}\left[\exp\left(-\frac{\lambda}{2}\right)\frac{(\lambda/2)^j}{j!}\right]\right| \\
&\leq \frac{1}{4}\sum_{j=0}^{\infty}\exp\left(-\frac{\lambda}{2}\right)\frac{(\lambda/2)^j}{j!} + \frac{1}{4}\sum_{j=1}^{\infty}\exp\left(-\frac{\lambda}{2}\right)\frac{(\lambda/2)^{(j-1)}}{(j-1)!} \\
&\leq \frac{1}{4}+\frac{1}{4} = \frac{1}{2}
\end{aligned}$$

for all z and λ. Combining the boundedness of $\partial^2 F_{p,K-p-q+1}(z,\lambda)/\partial\lambda^2$ with (36) yields

$$\begin{aligned}
P\left(\kappa^{-1}F_\infty < z\right) &= F_{p,K-p-q+1}\left(z,\delta^2\right) + O\left[var(\Delta^2)\right] \\
&= F_{p,K-p-q+1}\left(z,\delta^2\right) + O\left(\frac{1}{K^2}\right) = F_{p,K-p-q+1}\left(z,\delta^2\right) + o\left(\frac{1}{K}\right).
\end{aligned}$$

Part (c). It follows from Part (b) that

$$\begin{aligned}
P(pF_\infty < z) &= E\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K},\delta^2\right) + o\left(\frac{1}{K}\right) \\
&= E\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K},0\right) + E\frac{\partial}{\partial\lambda}\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K},0\right)\delta^2 + o\left(\frac{1}{K}\right) \\
&+ E\left[\frac{\partial^2}{\partial\lambda^2}\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K},\tilde{\delta}^2\right)\right]\delta^4,
\end{aligned} \tag{37}$$

where $\tilde{\delta}^2$ is between 0 and δ^2. As in the proof of Part (b), we can show that $\left|\frac{\partial^2}{\partial\lambda^2}\mathcal{G}_p(z,\lambda)\right| \leq 1$. As a result,

$$E\left[\frac{\partial^2}{\partial\lambda^2}\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K},\tilde{\delta}^2\right)\right]\delta^4 = O(\frac{1}{K^2}) = o(\frac{1}{K}). \tag{38}$$

Consequently,

$$P(pF_\infty < z) = E\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K}, 0\right) + E\frac{\partial}{\partial\lambda}\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K}, 0\right)\delta^2 + o(\frac{1}{K}).$$

By direct calculations, it is easy to show that

$$\frac{\partial}{\partial\lambda}\mathcal{G}_p(z,0) = -\frac{1}{2}\left[\mathcal{G}_p(z) - \mathcal{G}_{p+2}(z)\right] = -\frac{1}{p}\frac{z^{p/2}e^{-z/2}}{2^{p/2}\Gamma\left(\frac{p}{2}\right)} = -\frac{1}{p}\mathcal{G}'_p(z)\,z. \tag{39}$$

Therefore,

$$\begin{aligned}
&P(pF_\infty < z)\\
&= E\mathcal{G}_p\left(\frac{z\chi^2_{K-p-q+1}}{K}\right) - \frac{1}{p}E\mathcal{G}'_p\left(\frac{z\chi^2_{K-p-q+1}}{K}\right)\frac{z\chi^2_{K-p-q+1}}{K}\delta^2 + o\left(\frac{1}{K}\right)\\
&= \mathcal{G}_p(z) + \mathcal{G}'_p(z)\,zE\left(\frac{\chi^2_{K-p-q+1}}{K} - 1\right) + \frac{1}{2}\mathcal{G}''_p(z)\,z^2 var\left(\frac{\chi^2_{K-p-q+1}}{K}\right) - \frac{\delta^2}{p}\mathcal{G}'_p(z)\,z + o\left(\frac{1}{K}\right)\\
&= \mathcal{G}_p(z) + \mathcal{G}'_p(z)\,z\frac{-p-q+1}{K} + \mathcal{G}''_p(z)\,z^2\frac{K-p-q+1}{K^2} - \frac{q}{K-q-1}\mathcal{G}'_p(z)\,z\\
&= \mathcal{G}_p(z) - \mathcal{G}'_p(z)\,z\left(\frac{p+2q-1}{K}\right) + \mathcal{G}''_p(z)\,z^2\frac{1}{K} + o\left(\frac{1}{K}\right).
\end{aligned}$$

■

Proof of Theorem 4. Part (a). Using the same argument for proving Theorem 3(a), we have

$$t_\infty \overset{d}{=} \frac{B_1(1) - C_{1q}C_{qq}^{-1}B_q(1)}{\sqrt{\chi^2_{K-q}/K}} \tag{40}$$

and so

$$\frac{t_\infty}{\sqrt{\kappa}} \overset{d}{=} \frac{B_1(1) - C_{1q}C_{qq}^{-1}B_q(1)}{\sqrt{\chi^2_{K-q}/(K-q)}} \overset{d}{=} t_{K-q}(\Delta).$$

Part (b). Since the distribution of t_∞ is symmetric about 0, we have for any $z \in \mathbb{R}^+$:

$$\begin{aligned}
&P\left(\frac{t_\infty}{\sqrt{\kappa}} < z\right)\\
&= \frac{1}{2} + \frac{1}{2}P(|t_\infty/\sqrt{\kappa}| < |z|) = \frac{1}{2} + \frac{1}{2}P(t^2_\infty/\kappa < z^2)\\
&= \frac{1}{2} + \frac{1}{2}F_{1,K-q}\left(z^2, \delta^2\right) + o(\frac{1}{K})
\end{aligned}$$

where the second last equality follows from Theorem 3(b). When $z \in \mathbb{R}^-$, we have

$$\begin{aligned}
&P\left(\frac{t_\infty}{\sqrt{\kappa}} < z\right)\\
&= \frac{1}{2} - \frac{1}{2}P(|t_\infty/\sqrt{\kappa}| < |z|) = \frac{1}{2} - \frac{1}{2}P(t^2_\infty/\kappa < z^2)\\
&= \frac{1}{2} - \frac{1}{2}F_{1,K-q}\left(z^2, \delta^2\right) + o(\frac{1}{K}).
\end{aligned}$$

Therefore

$$P\left(\frac{t_\infty}{\sqrt{\kappa}}<z\right)=\frac{1}{2}+\frac{1}{2}sgn(z)F_{1,K-q}\left(z^2,\delta^2\right)+o(\frac{1}{K}).$$

Part (c). Using Theorem 3(c) and the symmetry of the distribution of t_∞ about 0, we have for any $z\in\mathbb{R}^+$:

$$\begin{aligned}P(t_\infty<z)&=\frac{1}{2}+\frac{1}{2}P(|t_\infty|<|z|)=\frac{1}{2}+\frac{1}{2}P(t_\infty^2<z^2)\\&=\frac{1}{2}+\frac{1}{2}\mathcal{G}\left(z^2\right)-\mathcal{G}'\left(z^2\right)z^2\left(\frac{q}{K}\right)+\frac{1}{2}\mathcal{G}''\left(z^2\right)z^4\frac{1}{K}+o\left(\frac{1}{K}\right).\end{aligned}$$

Using the relationships that

$$\frac{1}{2}+\frac{1}{2}\mathcal{G}\left(z^2\right)=\mathbf{\Phi}(z),$$

and

$$-\mathcal{G}'\left(z^2\right)z^2\left(\frac{q}{K}\right)+\frac{1}{2}\mathcal{G}''\left(z^2\right)z^4\frac{1}{K}=-\frac{1}{4K}\boldsymbol{\phi}(z)\left[z^3+z(4q+1)\right],$$

we have

$$P(t_\infty<z)=\mathbf{\Phi}(z)-\frac{1}{4K}z\boldsymbol{\phi}(z)\left[z^2+(4q+1)\right]+o(\frac{1}{K}).$$

Similarly when $z\in\mathbb{R}^-$, we have

$$P(t_\infty<z)=\mathbf{\Phi}(z)+\frac{1}{4K}z\boldsymbol{\phi}(z)\left[z^2+(4q+1)\right]+o(\frac{1}{K}).$$

Therefore

$$P(t_\infty<z)=\mathbf{\Phi}(z)-\frac{1}{4K}|z|\boldsymbol{\phi}(z)\left[z^2+(4q+1)\right]+o(\frac{1}{K}).$$

■

Before proving Theorem 5, we present a technical lemma. Part (i) of the lemma is proved in Sun (2012). Parts (ii) and (iii) of the lemma are proved in Sun, Phillips and Jin (2011)

Define $g=\lim_{x\to 0}\left[1-k\left(x\right)\right]/x^{q_0}$, q_0 is the Parzen exponent of the kernel function, $c_1=\int_{-\infty}^{\infty}k\left(x\right)dx$, $c_2=\int_{-\infty}^{\infty}k^2\left(x\right)dx$.

Lemma 5 *(i) For conventional kernel HAR variance estimators, we have, as* $h\to\infty$,

(a) $\mu_1=1-bc_1+O(b^2)$,

(b) $\mu_2=bc_2+O(b^2)$.

(ii) For sharp kernel HAR variance estimators, we have, as $h\to\infty$,

(a) $\mu_1=1-\frac{2}{\rho+2}$,

(b) $\mu_2=\frac{1}{\rho+1}+O\left(\frac{1}{\rho^2}\right)$.

(iii) For steep kernel HAR variance estimator, we have, as $h\to\infty$,

(a) $\mu_1=1-\left(\frac{\pi}{\rho g}\right)^{1/2}+O\left(\frac{1}{\rho}\right)$,

(b) $\mu_2=\left(\frac{\pi}{2\rho g}\right)^{1/2}+O\left(\frac{1}{\rho}\right)$.

Proof of Theorem 5. Part (a). Recall

$$pF_\infty = \left[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right]' D_{pp}^{-1}\left[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right].$$

Conditional on (C_{pq}, C_{qq}, C_{pp}), $B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)$ is normal with mean zero and variance $I_p + C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'$. Let L be the lower triangular matrix such that LL' is the Choleski decomposition of $I_p + C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'$. Then the conditional distribution of $\zeta := L^{-1}\left[B_p(1) - C_{pq}C_{qq}^{-1}B_q(1)\right]$ is $N(0, I_p)$. Since the conditional distribution does not depend on (C_{pq}, C_{qq}, C_{pp}), we can conclude that ζ is independent of (C_{pq}, C_{qq}, C_{pp}). So we can write

$$pF_\infty \overset{d}{=} \zeta' A \zeta$$

where $A = L'D_{pp}^{-1}L$. Given that A is a function of (C_{pq}, C_{qq}, C_{pp}), we know that ζ and A are independent. As a result, $\zeta' A\zeta \overset{d}{=} \zeta'(OAO')\zeta$ for any orthonormal matrix O that is independent of ζ.

Let $H = \left(\zeta/\|\zeta\|, \tilde{H}\right)$ be an orthonormal matrix with first column $\zeta/\|\zeta\|$. We choose H to be independent of A. This is possible as ζ and A are independent. Then

$$\begin{aligned} pF_\infty \overset{d}{=} \|\zeta\|^2 \frac{\zeta'}{\|\zeta\|}(OAO')\frac{\zeta}{\|\zeta\|} &= \|\zeta\|^2\left[\frac{\zeta'}{\|\zeta\|}H\right]H'(OAO')H\left[H'\frac{\zeta}{\|\zeta\|}\right] \\ &= \|\zeta\|^2 e_p'\left(H'OAO'H\right)e_p \end{aligned}$$

where $e_p = (1, 0, ..., 0)'$ is the first basis vector in $\mathbb{R}^p$. Since $\|\zeta\|^2$, H and A are mutually independent from each other, we can write

$$pF_\infty \overset{d}{=} \|\zeta\|^2 e_p'\left(\mathcal{H}'O'AO\mathcal{H}\right)e_p$$

for any orthonormal matrix $\mathcal{H}$ that is independent of both ζ and A. Letting $O = \mathcal{H}'$, we obtain:

$$pF_\infty \overset{d}{=} \|\zeta\|^2\left(e_p'Ae_p\right) = \frac{\|\zeta\|^2}{\left[e_p'Ae_p\right]^{-1}} \overset{d}{=} \frac{\|\zeta\|^2}{\left[e_p'L'D_{pp}^{-1}Le_p\right]^{-1}}.$$

Since Le_p is the first column of L, we have, using the definition of the Choleski decomposition:

$$Le_p = \frac{\left[I_p + C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'\right]e_p}{\sqrt{e_p'\left[I_p + C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'\right]e_p}}.$$

As a result,

$$pF_\infty \overset{d}{=} \frac{\|\zeta\|^2}{\eta^2}$$

where

$$\eta^2 = \left[e_p'L'D_{pp}^{-1}Le_p\right]^{-1} = \frac{e_p'\left[I_p + C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'\right]e_p}{e_p'\left[I_p + C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'\right]D_{pp}^{-1}\left[I_p + C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'\right]e_p}.$$

Part (b). It is easy to show that

$$EC_{qq} = \mu_1 I_q \text{ and } var\left[vec\left(C_{qq}\right)\right] = \mu_2\left(I_{qq} + \mathbb{K}_{qq}\right)$$

where

$$\mu_1 = \int_0^1 Q_h^*(r,r)\,dr, \mu_2 = \int_0^1 \int_0^1 [Q_h^*(r,s)]^2\,drds,$$

I_{qq} is the $q^2 \times q^2$ identity matrix, and $\mathbb{K}_{qq}$ is the $q^2 \times q^2$ commutation matrix. So $C_{qq} = \mu_1 I_q + o_p(1)$ and $C_{qq}^{-1} = \mu_1^{-1} I_q + o_p(1)$. Similarly, $C_{pp} = \mu_1 I_p + o_p(1)$ and $C_{pp}^{-1} = \mu_1^{-1} I_p + o_p(1)$. In addition, using the same argument, we can show that $C_{pq} = o_p(1)$. Therefore

$$\begin{aligned}
\eta^2 &= \frac{e_p' \left[I_p + C_{pq}C_{pq}'/\mu_1^2\right] e_p}{e_p' \left[I_p + C_{pq}C_{pq}'/\mu_1^2\right] \left[I_p - C_{pq}C_{pq}'/\mu_1^2\right]^{-1} \left[I_p + C_{pq}C_{pq}'/\mu_1^2\right] e_p} (1 + o_p(1)) \\
&= 1 + o_p(1)
\end{aligned}$$

That is, $\eta^2 \to^p 1$.

We proceed to prove the distributional expansion. The (i,j)-th elements $C_{pp}(i,j)$, $C_{pq}(i,j)$ and $C_{qq}(i,j)$ of C_{pp}, C_{pq} and C_{qq} are equal to either $\int_0^1 \int_0^1 Q_h^*(r,s)\,dB(r)dB(s)$ or $\int_0^1 \int_0^1 Q_h^*(r,s)\,dB(r)d\tilde{B}(s)$ where $B(\cdot)$ and $\tilde{B}(\cdot)$ are independent standard Brownian motion processes. By direct calculations, we have, for any $\varsigma \in (0, 3/8)$,

$$P\left(|C_{ef}(i,j) - EC_{ef}(i,j)| > \mu_2^{\varsigma}\right) \leq \frac{E\left|C_{ef}(i,j) - EC_{ef}(i,j)\right|^8}{\mu_2^{8\varsigma}} = O\left(\frac{\mu_2^4}{\mu_2^{8\varsigma}}\right) = o(\mu_2)$$

where $e, f = p$ or q. Define the event $\mathcal{E}$ as

$$\mathcal{E} = \left\{\omega : |C_{ef}(i,j) - EC_{ef}(i,j)| \leq \mu_2^{\varsigma} \text{ for all } i,j \text{ and all } e \text{ and } f\right\},$$

then the complement $\mathcal{E}^c$ of $\mathcal{E}$ satisfies $P(\mathcal{E}^c) = o(\mu_2)$ as $h \to \infty$. Let $\tilde{\mathcal{E}}$ be another event, then

$$\begin{aligned}
P\left(\tilde{\mathcal{E}}\right) &= P\left(\tilde{\mathcal{E}} \cap \mathcal{E}\right) + P\left(\tilde{\mathcal{E}} \cap \mathcal{E}^c\right) = P\left(\tilde{\mathcal{E}} \cap \mathcal{E}\right) + o(\mu_2) \\
&= P(\tilde{\mathcal{E}}|\mathcal{E})P(\mathcal{E}) + o(\mu_2) = P(\tilde{\mathcal{E}}|\mathcal{E})(1 - o(\mu_2)) + o(\mu_2) \\
&= P(\tilde{\mathcal{E}}|\mathcal{E}) + o(\mu_2).
\end{aligned}$$

That is, up to an error of order $o(\mu_2)$, $P\left(\tilde{\mathcal{E}}\right)$, $P\left(\tilde{\mathcal{E}} \cap \mathcal{E}\right)$ and $P(\tilde{\mathcal{E}}|\mathcal{E})$ are equivalent. So for the purpose of proving the theorem, it is innocuous to condition on $\mathcal{E}$ or remove the conditioning, if needed.

Now conditioning $\mathcal{E}$, the numerator of η^2 satisfies:

$$\begin{aligned}
&e_p' \left[I_p + C_{pq}C_{qq}^{-1}C_{qq}^{-1}C_{pq}'\right] e_p \\
&= 1 + \frac{1}{\mu_1^2} e_p' C_{pq} \left[I_q + \frac{C_{qq} - EC_{qq}}{\mu_1}\right]^{-1} \left[I_q + \frac{C_{qq} - EC_{qq}}{\mu_1}\right]^{-1} C_{pq}' e_p \\
&= 1 + \frac{1}{\mu_1^2} e_p' C_{pq} (I_q - M_{qq})(I_q - M_{qq}) C_{pq}' e_p \\
&= 1 + \frac{1}{\mu_1^2} e_p' C_{pq} \tilde{M}_{qq} C_{pq}' e_p
\end{aligned}$$

where M_{qq} is a matrix with elements $M_{qq}(i,j)$ satisfying $|M_{qq}(i,j)| = O(\mu_2^{\varsigma})$ conditional on $\mathcal{E}$, and $\tilde{M}_{qq}$ is a matrix satisfying $\left|\tilde{M}_{qq}(i,j) - 1\{i = j\}\right| = O(\mu_2^{\varsigma})$ conditional on $\mathcal{E}$.

Let

$$C_{p+q,p+q} = \begin{pmatrix} C_{pp} & C_{pq} \\ C'_{pq} & C_{qq} \end{pmatrix}$$

and $e_{q+q,p}$ be the matrix consisting of the p columns of the identity matrix I_{p+q}. To evaluate the denominator of η^2, we note that

$$\begin{aligned} D_{pp}^{-1} &= \frac{1}{\mu_1} e'_{q+q,p} \left(I_{p+q} + \frac{C_{p+q,p+q} - EC_{p+q,p+q}}{\mu_1} \right)^{-1} e_{q+q,p} \\ &= \frac{1}{\mu_1} e'_{q+q,p} \left(I_{p+q} - \frac{C_{p+q,p+q} - EC_{p+q,p+q}}{\mu_1} \right) e_{q+q,p} \\ &\quad + e'_{q+q,p} \left(\frac{\left[C_{p+q,p+q} - EC_{p+q,p+q}\right]\left[C_{p+q,p+q} - EC_{p+q,p+q}\right]}{\mu_1^2} \right) e_{q+q,p} + M_{pp} \\ &= \frac{1}{\mu_1} \left(I_p - \frac{C_{pp} - EC_{pp}}{\mu_1} + \frac{\left[C_{pp} - EC_{pp}\right]\left[C_{pp} - EC_{pp}\right] + C_{pq}C'_{pq}}{\mu_1^2} \right) + M_{pp} \end{aligned}$$

where M_{pp} is a matrix with elements $M_{pp}(i,j)$ satisfying $|M_{pp}(i,j)| = O\left(\mu_2^{3\varsigma}\right) = o(\mu_2)$ for $\varsigma > 1/3$ conditional on $\mathcal{E}$.

For the purpose of proving our result, M_{pp} can be ignored as its presence generates an approximation error of order $o(\mu_2)$, which is the same as the order of the approximation error given in the theorem. More specifically, let

$$\tilde{C}_{pp} = \frac{C_{pq}\tilde{M}_{qq}C'_{pq}}{\mu_1^2}, \quad \tilde{D}_{pp}^{-} = \left(I_p - \frac{C_{pp} - EC_{pp}}{\mu_1} + \frac{\left[C_{pp} - EC_{pp}\right]\left[C_{pp} - EC_{pp}\right] + C_{pq}C'_{pq}}{\mu_1^2} \right)$$

and

$$\tilde{\eta}^2 := \tilde{\eta}^2(\tilde{C}_{pp}, \tilde{D}_{pp}^{-}) = \frac{\mu_1 \left(1 + e'_p \tilde{C}_{pp} e_p\right)}{e'_p \left(I_p + \tilde{C}_{pp}\right) \tilde{D}_{pp}^{-} \left(I_p + \tilde{C}_{pp}\right) e_p},$$

then

$$P\left(pF_\infty < z\right) = E\mathcal{G}_p\left(\eta^2 z\right) = E\mathcal{G}_p\left(\tilde{\eta}^2 z\right) + o\left(\mu_2\right).$$

Note that for any $q \times q$ matrix L_{qq}, we have

$$\begin{aligned} EC_{pq}L_{qq}C'_{pq} &= E\int_0^1\int_0^1 Q_h^*(r_1,s_1)Q_h^*(r_2,s_2)dB_p(r_1)dB_q(s_1)'L_{qq}dB_q(s_2)dB_p(r_1)' \\ &= E\int_0^1\int_0^1 Q_h^*(r_1,s_1)Q_h^*(r_2,s_2)dB_p(r_1)dB_p(r_1)'tr\left(dB_q(s_2)dB_q(s_1)'L_{qq}\right) \\ &= tr\left(L_{qq}\right)\int_0^1\int_0^1 \left[Q_h^*(r,s)\right]^2 drdrI_p = tr\left(L_{qq}\right)\mu_2 I_p. \end{aligned} \quad (41)$$

Taking an expansion of $\tilde{\eta}^2(\tilde{C}_{pp}, \tilde{D}_{pp}^{-})$ around $\tilde{C}_{pp} = q\mu_2/\mu_1^2 I_p$ and $\tilde{D}_{pp}^{-} = I_p$, we obtain

$$\tilde{\eta}^2 = \eta_0^2 + err$$

where err is the approximation error and

$$\eta_0^2 = \mu_1 - \frac{2q\mu_2}{\mu_1} + e'_p\left[C_{pp} - EC_{pp}\right]e_p - \frac{e'_p\left[C_{pp} - EC_{pp}\right]\left[C_{pp} - EC_{pp}\right]e_p}{\mu_1} + \frac{\left\{e'_p\left[C_{pp} - EC_{pp}\right]e_p\right\}^2}{\mu_1}.$$

We keep enough terms in η_0^2 so that $E\mathcal{G}_p\left(\tilde{\eta}^2 z\right) = E\mathcal{G}_p\left(\eta_0^2 z\right) + o\left(\mu_2\right).$

Now we write

$$\begin{aligned} P\left(pF_\infty < z\right) &= E\mathcal{G}_p(\eta_0^2 z) + o\left(\mu_2\right) \\ &= \mathcal{G}_p\left(z\right) + \mathcal{G}_p'\left(z\right) z \left(E\eta_0^2 - 1\right) + \frac{1}{2}\mathcal{G}_p''\left(z\right) z^2 E(\eta_0^2 - 1)^2 + o\left(\mu_2\right). \end{aligned}$$

In view of

$$\begin{aligned} E\eta_0^2 - 1 &= \left(\mu_1 - 1\right) - \frac{2q\mu_2}{\mu_1} - \frac{1}{\mu_1} E e_p' \left[C_{pp} - EC_{pp}\right]\left[C_{pp} - EC_{pp}\right] e_p + E\frac{\left\{e_p'\left[C_{pp} - EC_{pp}\right] e_p\right\}^2}{\mu_1} \\ &= \left(\mu_1 - 1\right) - \frac{2q\mu_2}{\mu_1} - \frac{\mu_2}{\mu_1}\left(p+1\right) + \frac{2\mu_2}{\mu_1} \\ &= \left(\mu_1 - 1\right) - \frac{2q\mu_2}{\mu_1} - \frac{\mu_2}{\mu_1}\left(p-1\right) \end{aligned}$$

and

$$\begin{aligned} &E(\eta_0^2 - 1)^2 \\ &= E\left[\mu_1 - 1 - \frac{2q\mu_2}{\mu_1} + e_p'\left[C_{pp} - EC_{pp}\right] e_p\right]^2 + o\left(\mu_2\right) \\ &= E\left\{e_p'\left[C_{pp} - EC_{pp}\right] e_p\right\}^2 + \left(\mu_1 - 1\right)^2 + o\left(\mu_2\right) \\ &= 2\mu_2 + \left(\mu_1 - 1\right)^2 + o\left(\mu_2\right), \end{aligned}$$

we have

$$\begin{aligned} &P\left(pF_\infty < z\right) \\ &= \mathcal{G}_p\left(z\right) + \mathcal{G}_p'\left(z\right) z\left[\left(\mu_1 - 1\right) - \frac{2q\mu_2}{\mu_1} - \frac{\mu_2}{\mu_1}\left(p-1\right)\right] + \mathcal{G}_p''\left(z\right) z^2 \mu_2 + o\left(\mu_2\right) \\ &= \mathcal{G}_p\left(z\right) + \mathcal{G}_p'\left(z\right) z\left[\left(\mu_1 - 1\right) - \frac{\mu_2}{\mu_1}\left(p + 2q - 1\right)\right] + \mathcal{G}_p''\left(z\right) z^2 \mu_2 + o\left(\mu_2\right). \end{aligned} \tag{42}$$

■

References

[1] Andrews, D.W.K. (1991). "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59, 817–858.

[2] Andrews, D.W.K. and J. C. Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator." *Econometrica* 60(4), 953–966.

[3] Bester, A., Conley, T., Hansen, C., and Vogelsang, T. (2011): "Fixed-b Asymptotics for Spatially Dependent Robust Nonparametric Covariance Matrix Estimators." Working paper, Michigan State University.

[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] Chen, X., Z. Liao and Y. Sun (2012): "Sieve Inference on Possibly Misspecified Semi-nonparametric Time Series Models." Working paper, Yale, UCLA and UCSD.

[6] Gonçlaves, S. (2011): "The moving blocks bootstrap for panel linear regression models with individual fixed effects." *Econometric Theory* 27(5):1048–1082

[7] Gonçlaves, S. and T. Vogelsang (2011): "Block Bootstrap HAC Robust Tests: The Sophistication of the Naive Bootstrap." *Econometric Theory* 27(4), 745–791

[8] Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50, 1029–1054.

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

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

[14] Kim, M. S. and Sun, Y. (2012): "Heteroskedasticity and Spatiotemporal Dependence Robust Inference for Linear Panel Models with Fixed Effects". *Working paper.* Department of Economics, Ryerson University and UC San Diego.

[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. (2005): "HAC Estimation by Automated Regression." *Econometric Theory* 21, 116–142.

[17] Phillips, P.C.B., Y. Sun and S. Jin (2006): "Spectral Density Estimation and Robust Hypothesis Testing using Steep Origin Kernels without Truncation." *International Economic Review* 47, 837–894

[18] Phillips, P.C.B., Y. Sun and S. Jin (2007): "Consistent HAC Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation." *Journal of Statistical Planning and Inference* 137, 985–1023.

[19] Politis, D. (2011): "Higher-order Accurate, Positive Semi-definite Estimation of Large-sample Covariance and Spectral Density Matrices." *Econometric Theory*, 27, 703–744.

[20] Press, J. (2005): *Applied Multivariate Analysis,* Second Edition, Dover Books on Mathematics.

[21] Shao, X. (2010): "A Self-normalized Approach to Confidence Interval Construction in Time Series." *Journal of the Royal Statistical Society* B, 72, 343–366.

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

[23] Sun, Y. (2012): "Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroscedasticity and Autocorrelation Robust Inference." Working paper, UC San Diego.

[24] Sun, Y. (2013): "A Heteroskedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator." *Econometrics Journal*, forthcoming.

[25] Sun, Y. and D. Kaplan (2012): "A new asymptotic theory for vector autoregressive long-run variance estimation and autocorrelation robust testing." Working paper, UC San Diego.

[26] Sun, Y., P.C.B. Phillips and S. Jin (2008). "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing." *Econometrica* 76(1), 175–194.

[27] Sun, Y., P.C.B. Phillips and S. Jin (2011): "Power Maximization and Size Control in Heteroscedasticity and Autocorrelation Robust Tests with Exponentiated Kernels." *Econometric Theory* 27(6), 1320–1368.

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

[29] Sun, Y. and M.S. Kim (2012a): "Simple and Powerful GMM Over-identification Tests with Accurate Size." *Journal of Econometrics* 166(2), 267–281

[30] Sun, Y. and M.S. Kim (2012b): "Asymptotic F Test in a GMM Framework with Cross Sectional Dependence." *Working paper.* Department of Economics, Ryerson University and UC San Diego.

[31] Taniguchi, M. and Puri, M.L. (1996): "Valid Edgeworth Expansions of M-estimators in Regression Models with Weakly Dependent Residuals." *Econometric Theory* 12, 331–346.

[32] Xiao, Z. and O. Linton (2002): "A Nonparametric Prewhitened Covariance Estimator." *Journal of Time Series Analysis* 23, 215–250.

[33] Zhang, X and X. Shao (2013): "Fixed-Smoothing Asymptotics For Time Series." *The Annals of Statistics*, forthcoming.