IMPROVED HAR INFERENCE
USING POWER KERNELS WITHOUT TRUNCATION

By

Peter C.B. Phillips, Yixiao Sun and Sainan Jin

June 2005

COWLES FOUNDATION DISCUSSION PAPER NO. 1513

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Improved HAR Inference Using Power Kernels without Truncation*

Peter C. B. Phillips
Cowles Foundation, Yale University,
University of Auckland & University of York

Yixiao Sun
Department of Economics
University of California, San Diego

and

Sainan Jin
Guanghua School of Management
Peking University

November 29, 2004

*Phillips thanks the NSF for partial research support under Grant No. SES 04-142254.

ABSTRACT

Employing power kernels suggested in earlier work by the authors (2003), this paper shows how to refine methods of robust inference on the mean in a time series that rely on families of untruncated kernel estimates of the long-run parameters. The new methods improve the size properties of heteroskedastic and autocorrelation robust (HAR) tests in comparison with conventional methods that employ consistent HAC estimates, and they raise test power in comparison with other tests that are based on untruncated kernel estimates. Large power parameter (ρ) asymptotic expansions of the nonstandard limit theory are developed in terms of the usual limiting chi-squared distribution, and corresponding large sample size and large ρ asymptotic expansions of the finite sample distribution of Wald tests are developed to justify the new approach. Exact finite sample distributions are given using operational techniques. The paper further shows that the optimal ρ that minimizes a weighted sum of type I and II errors has an expansion rate of at most $O\left(T^{1/2}\right)$ and can even be $O(1)$ for certain loss functions, and is therefore slower than the $O\left(T^{2/3}\right)$ rate which minimizes the asymptotic mean squared error of the corresponding long run variance estimator. A new plug-in procedure for implementing the optimal ρ is suggested. Simulations show that the new plug-in procedure works well in finite samples.

JEL Classification: C13; C14; C22; C51

Keywords: Asymptotic expansion, consistent HAC estimation, data-determined kernel estimation, exact distribution, HAR inference, large ρ asymptotics, long run variance, loss function, power parameter, sharp origin kernel.

1 Introduction

Seeking to robustify inference, many practical methods in econometrics now make use of heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimates. Most commonly used HAC estimates are formulated using conventional kernel smoothing techniques (for an overview, see den Haan and Levin (1997)), although quite different approaches like wavelets (Hong and Lee (2001)) and direct regression methods (Phillips (2004)) have recently been explored. While appealing in terms of their asymptotic properties, consistent HAC estimates provide only asymptotic robustness in econometric testing and finite sample performance is known to be unsatisfactory in many cases, but especially when there is strong autocorrelation in the data. HAC estimates are then biased downwards and the associated tests are liberal-biased. These size distortions in testing are often substantial and have been discussed extensively in recent work (e.g., Kiefer and Vogelsang (2003) and Sul, Phillips and Choi (2003)).

Robustification in regression testing is achieved by the use of a test statistic that is asymptotically pivotal under a general maintained hypothesis for the regression components. Consistent HAC estimation is not necessary for this purpose and, indeed, any procedure that scales out the effects of the nuisance parameters in the test statistics will work. Kiefer, Vogelsang and Bunzel (2000, hereafter KVB) suggested the use of untruncated, inconsistent kernel estimates in the construction of test statistics and showed that the limit theory is nuisance parameter free but no longer standard normal or chi-squared. Work on related procedures has been done by Vogelsang (2003), Kiefer and Vogelsang (2002a, 2002b, 2003; hereafter KV) and by the present authors (2003a & 2003b; hereafter PSJ_a & PSJ_b). These techniques may be grouped with conventional HAC procedures as having the same goals of robust inference and the term heteroskedastic and autocorrelation robust (HAR) methods has been used to collectively describe them (Phillips, 2004).

Inconsistent covariance matrix estimates play an interesting role in improving the size properties of tests, essentially because they preserve in the limit theory the finite sample randomness of the denominator in the conventional t-ratio. In this respect, these tests behave in large samples more like their finite sample analogues than the conventional asymptotic normal and chi-squared tests, for which the denominator is non-random. In the case of the Gaussian location model, Jansson (2004) showed that the KVB test statistic is closer to its limit distribution in the precise sense that the error in the rejection probability (ERP) is of order $O\left(T^{-1}\log T\right)$ for sample size T under the null, whereas the corresponding ERP for a test based on a conventional consistent HAC estimate is at most of order $O(T^{-1/2})$, as shown in Velasco and Robinson (2001). While tests such as KVB typically have better size than those that use HAC estimators, there is also a clear and compensating reduction in power. The challenge is to develop test procedures with size improvements like those of KVB, while retaining the good power properties of conventional tests based on HAC estimators.

The present paper confronts this challenge by developing a procedure that combines the use of untruncated kernels, as in KVB, with a refinement that enables the use of critical values that appropriately correct those of the limit theory for conventional tests based on consistent HAC estimators, while at the same time enhancing the test power of the KVB test. The class of HAR tests considered here involve the use of a power kernel suggested by the authors in other work (2003a) and this class includes both consistent and inconsistent HAC estimates, depending on whether the power parameter, ρ, is fixed or passes to infinity as $T \to \infty$. When $\rho \to \infty$, the first order limit theory corresponds to that of a test based on conventional consistent HAC estimation, whereas for ρ fixed, the limit theory is nonstandard, as in the case of the KVB test. The mechanism for making improvements in both size (compared with asymptotic normal tests) and power (compared with the KVB test) is to use a test statistic for a moderate value of ρ for which the critical values can be obtained from the appropriate nonstandard limit distribution, which is nuisance parameter free. It is shown here how these critical values may be very well approximated using an asymptotic expansion of the limit distribution about its limiting chi-squared distribution. This version of the procedure has the advantage of being easily implemented and does not require the use of tables of nonstandard distributions.

This refinement improves test size in the same manner as the KVB test, and is justified in the present paper by asymptotic expansions of both the non-standard limit distribution as $\rho \to \infty$ and the finite sample distribution as $T \to \infty$ and $\rho \to \infty$. The first expansion can be regarded as a high order expansion under the sequential limit in which $T \to \infty$ first followed by $\rho \to \infty$. The second expansion is a high order expansion under the joint limit in which $T \to \infty$ and $\rho \to \infty$ simultaneously. Corresponding asymptotic expansions of the power functions indicate that for typical economic time series test power increases as ρ increases. Finite sample improvements in test power over other tests with untruncated kernels like the KVB test have been noted in simulations reported in other work by the authors (2003a, 2003b) and in independent work by Ravikumar, Ray and Savin (2004) using the methods of PSJ_b. The asymptotic expansions given in the present paper help to explain these power improvements.

A further contribution of the present paper is to use these asymptotic expansions to suggest a practical procedure for test implementation which optimally balances the type I and type II errors. The type I error is measured by using the first correction term in the asymptotic expansion of the finite sample distribution of the test statistic about its nonstandard limit distribution. This term is of order $O(\rho/T)$ and it increases in magnitude as ρ increases for any given T. Similarly, the expansions under the local alternative reveal that in general the type II error decreases as ρ increases. Thus, to this order in the asymptotic expansion, increasing ρ reduces the type II error but also increases the type I error. Since the desirable effects on the two types of errors generally work in opposing directions, we construct a loss function criterion by taking a weighted sum of the two types of errors and show how ρ may be selected in such a way as to optimize the criterion. This approach gives an optimal ρ which generally has an expansion rate of at most $\rho_{\text{opt}} = O\left(T^{1/2}\right)$ and which can even be $O(1)$ for certain loss functions. This rate is less than the optimal rate of $O\left(T^{2/3}\right)$ that applies when minimizing the asymptotic mean squared error of the corresponding HAC variance estimate (c.f., PSJ_a). Thus, optimal values of ρ for HAC standard error estimation are larger as $T \to \infty$ than those which are most suited for statistical testing. The fixed ρ rule is obtained by attaching substantially higher weight to the type I error in the construction of the loss function. This theory therefore provides some insight into the type of loss function for which there is a decision theoretic justification for the use of fixed ρ rules in econometric testing. These conclusions are also relevant to the use of untruncated kernel estimates in econometric testing of the type suggested in KV (2003).

The plan of the paper is as follows. Section 2 overviews the class of power kernels that will be used in the present paper's development and reviews some first order limit theory for Wald type tests as $T \to \infty$ with the power parameter ρ fixed and as $\rho \to \infty$. Section 3 derives an exact distribution theory using operational techniques. Section 4 develops an asymptotic expansion of the non-standard limit distribution under the null hypothesis as the power parameter $\rho \to \infty$ about the usual limiting chi-squared distribution. The second order term in this asymptotic expansion delivers a correction term that can be used to adjust the critical values in the usual chi-squared test. An asymptotic expansion of the local power function is also given. Section 5 develops comparable expansions of the finite sample distribution of the statistic as $T \to \infty$ for a fixed ρ and as both $T \to \infty$ and $\rho \to \infty$. This expansion validates the use of the corrected critical values in practical work. Section 6 proposes a selection rule for ρ that is suitable for implementation in semiparametric testing. This criterion optimizes a loss function that is constructed to balance higher order approximations to the type I and type II errors. Section 7 reports some simulation evidence on the performance of the new procedures. Section 8 concludes and discusses the implications of the results for applied work. Proofs and additional technical results are in the Appendix.

2 HAR Inference for the Mean

Throughout the paper, we focus on the inference about β in the location model:

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

where u_t is zero mean process with a nonparametric autocorrelation structure. The non-standard limiting distribution in this section and its asymptotic expansion in Section 4 apply to general regression models under certain conditions on the regressors, see PSJ$_a$. However, the asymptotic expansion of the finite sample distribution in Section 5 applies only to the location model. A possible extension is discussed in Section 8.

The OLS estimation of β gives

$$\hat{\beta} = \bar{Y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

and the scaled estimation error is

$$\sqrt{T}(\hat{\beta} - \beta) = \frac{1}{\sqrt{T}} S_T, \tag{2}$$

where $S_t = \sum_{\tau=1}^{t} u_\tau$. Let $\hat{u}_\tau = y_\tau - \hat{\beta}$ be the demeaned time series and $\hat{S}_t = \sum_{\tau=1}^{t} \hat{u}_\tau$ be the corresponding partial sum process.

The following condition is commonly used to facilitate the limit theory (e.g., KVB, PSJ_a, and Jansson, 2004).

Assumption 1 $S_{[Tr]}$ *satisfies the functional law*

$$T^{-1/2}S_{[Tr]} \Rightarrow \omega W(r), \quad r \in [0,1]$$

where ω^2 is the long run variance of u_t and $W(r)$ is the standard Brownian motion.

Under Assumption 1,

$$T^{-1/2}\hat{S}_{[Tr]} \Rightarrow \omega V(r), \quad r \in [0,1], \tag{3}$$

where V is a standard Brownian bridge process, and

$$\sqrt{T}(\hat{\beta} - \beta) \Rightarrow \omega W(1) = N(0, \omega^2), \tag{4}$$

which provides the usual basis for robust testing about β. It is the standard practice to estimate ω^2 using kernel-based nonparametric HAC estimators that involve smoothing and truncation lag covariances. When u_t is stationary with spectral density $f_{uu}(\lambda)$, the long run variance (LRV) of u_t is

$$\omega^2 = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma(j) = 2\pi f_{uu}(0), \tag{5}$$

where $\gamma(j) = E(u_t u_{t-j})$. HAC estimates of ω^2 typically have the following form

$$\hat{\omega}^2(M) = \sum_{j=-T+1}^{T-1} k(\frac{j}{M})\hat{\gamma}(j), \ \hat{\gamma}(j) = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-j} u_{t+j}u_t & \text{for } j \geq 0 \\ \frac{1}{T}\sum_{t=-j+1}^{T} u_{t+j}u_t & \text{for } j < 0 \end{cases} \tag{6}$$

involving the sample covariances $\hat{\gamma}(j)$. In (6), $k(\cdot)$ is some kernel function, M is a bandwidth parameter, and consistency of $\hat{\omega}^2(M)$ requires $M \to \infty$ and $M/T \to 0$ as $T \to \infty$ (e.g. Andrews (1991), Andrews and Monahan (1992), Hansen (1992), Newey and West (1987,1994), de Jong and Davidson (2000)). Jansson (2002) provides a recent overview and weak conditions for consistency of such estimates.

To test the null $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$, the standard approach relies on a t-ratio statistic of the form

$$t_{\hat{\omega}(M)} = T^{1/2}(\hat{\beta} - \beta_0)/\hat{\omega}(M) \tag{7}$$

which is asymptotically $N(0,1)$. Use of $t_{\hat{\omega}(M)}$ is convenient empirically and therefore widespread in practical work, in spite of well-known problems of size distortion in inference.

In a series of papers, KVB and KV propose the use of kernel-based estimators of ω^2 in which M is set equal to the sample size T or proportional to T. These estimates are inconsistent and tend to random quantities instead of ω^2, so the limit distribution of (7) is no longer standard normal. Nonetheless, use of these estimates results in valid asymptotically similar tests.

In related work, PSJ$_a$ and PSJ$_b$ propose the use of estimates of ω^2 based on power kernels without truncation, so that $M = T$ again. For instance, in PSJ$_a$ a class of sharp origin kernels were constructed in this way by taking an arbitrary power $\rho \geq 1$ of the usual Bartlett kernel, giving

$$k_\rho(x) = \begin{cases} (1-|x|)^\rho, & |x| \leq 1 \\ 0, & |x| > 1 \end{cases} \quad \text{for } \rho \in \mathbb{Z}^+. \tag{8}$$

We will focus on the sharp origin kernels in the rest of the paper. Using k_ρ in (6) and letting $M = T$ gives HAC estimates of the form

$$\hat{\omega}_\rho^2 = \sum_{j=-T+1}^{T-1} k_\rho\left(\frac{j}{T}\right)\hat{\gamma}(j). \tag{9}$$

Under Assumption 1, $\hat{\omega}_\rho^2 \Rightarrow \omega^2 \Xi_\rho$, where $\Xi_\rho = \int_0^1 \int_0^1 k_\rho(r-s)dV(r)dV(s)$.

The associated t statistic is given by

$$t^*\left(\hat{\omega}_\rho\right) = T^{1/2}(\hat{\beta} - \beta_0)/\hat{\omega}_\rho. \tag{10}$$

When the power parameter ρ is fixed as $T \to \infty$, PSJ$_a$ showed that under Assumption 1 the t^*-statistic has the nonstandard limit distribution:

$$t^*\left(\hat{\omega}_\rho\right) \Rightarrow W(1)\Xi_\rho^{-1/2} \tag{11}$$

under the null and

$$t^*\left(\hat{\omega}_\rho\right) \Rightarrow \left(\delta + W(1)\right)\Xi_\rho^{-1/2}, \tag{12}$$

under the local alternative $H_1 : \beta = \beta_0 + cT^{-1/2}$, where $\delta = c/\omega$.

When ρ is sample size dependent and satisfies $1/\rho + \left(\rho \log T\right)/T \to 0$, PSJ$_a$ showed that $\hat{\omega}_\rho$ is consistent. In this case, the t^*-statistic has conventional normal limits: under the null $t^*\left(\hat{\omega}_\rho\right) \Rightarrow W(1) =_d N(0,1)$; and under the local alternative $t^*\left(\hat{\omega}_\rho\right) \Rightarrow \delta + W(1)$.

Thus, the t^*-statistic has nonstandard limit distributions arising from the random limit of the HAC estimate $\hat{\omega}_\rho$ when ρ is fixed as $T \to \infty$, just as the KVB and KV tests do. However, as ρ increases, the effect of this randomness diminishes, and when $\rho \to \infty$ the limit distributions approach those of conventional regression tests with consistent HAC estimates.

The mechanism we develop for making improvements in size without sacrificing much power, is to use a test statistic constructed with $\hat{\omega}_\rho$ based on a moderate value of ρ. The critical values of this test can be obtained from the fixed ρ limit theory given above. Alternatively, they can be based on an accurate but simple asymptotic expansion of that distribution about its limiting chi-squared distribution that applies as $\rho \to \infty$. This expansion is developed in Section 4.

3 Probability Densities of the Nonstandard Limit Distribution and the Finite Sample Distribution

This section develops some useful formulae for the probability densities of the fixed ρ limit theory and the exact distribution of the test statistic.

First note that in the limit theory of the t-ratio test, $W(1)$ is independent of Ξ_ρ, so the conditional distribution of $W(1)\Xi_\rho^{-1/2}$ given Ξ_ρ is normal with zero mean and variance Ξ_ρ^{-1}. We can write $\Xi_\rho = \Xi_\rho(\mathcal{V})$ where the process $\mathcal{V}$ has probability measure $P(\mathcal{V})$. The pdf of $t = W(1)\Xi_\rho^{-1/2}$ can then be written in the mixed normal form as

$$p_t(z) = \int_{\Xi_\rho(\mathcal{V})>0} N\left(0, \Xi_\rho^{-1}\right) dP(\mathcal{V}). \tag{13}$$

For the finite sample distribution of $t_T = t^*(\hat{\omega}_\rho)$, we assume that u_t is a Gaussian process. Since u_t is in general autocorrelated, $\sqrt{T}(\hat{\beta}-\beta)$ and $\hat{\omega}$ are statistically dependent. To find the exact finite sample distribution of the t-statistic, we decompose $\hat{\beta}$ and $\hat{\omega}$ into statistically independent components. Let $u = (\mu_1, ...u_T)'$, $y = (y_1, ..., y_T)$, $l_T = (1, ..., 1)^T$ and $\Omega_T = var(u)$. Then the GLS estimator of β is $\tilde{\beta} = \left(l_T'\Omega_T^{-1}l_T\right)^{-1} l_T'\Omega_T^{-1}y$ and

$$\hat{\beta} - \beta = \tilde{\beta} - \beta + \left(l_T'l_T\right)^{-1} l_T'\tilde{u} \tag{14}$$

where $\tilde{u} = (I - l_T\left(l_T'\Omega_T^{-1}l_T\right)^{-1} l_T'\Omega_T^{-1})u$, which is statistically independent of $\tilde{\beta} - \beta$.

Therefore the t-statistic can be written as

$$t_T = \frac{\sqrt{T}(\tilde{\beta} - \beta)}{\hat{\omega}_\rho(\hat{u})} + \frac{l_T'\tilde{u}}{\sqrt{T}\hat{\omega}_\rho(\tilde{u})} \tag{15}$$

It is easy to see that $\hat{u} = \left(I - l_T(l_T'l_T)^{-1}l_T'\right)u = \left(I - l_T(l_T'l_T)^{-1}l_T'\right)\tilde{u}$. As consequence, the conditional distribution of t_T given $\tilde{u}$ is

$$N\left(\frac{l_T'\tilde{u}}{\sqrt{T}\hat{\omega}_\rho(\tilde{u})}, \frac{T\left(l_T'\Omega_T^{-1}l_T\right)^{-1}}{(\hat{\omega}_\rho(\tilde{u}))^2}\right). \tag{16}$$

Letting $P(\tilde{u})$ be the probability measure of $\tilde{u}$, we deduce that the probability density of t_T is

$$\begin{aligned} p_{t_T}(z) &= \int N\left(\frac{l_T'\tilde{u}}{\sqrt{T}\hat{\omega}_\rho(\tilde{u})}, \frac{T\left(l_T'\Omega_T^{-1}l_T\right)^{-1}}{(\hat{\omega}_\rho(\tilde{u}))^2}\right) dP(\tilde{u}) \\ &= E\left\{N\left(\frac{l_T'\tilde{u}}{\sqrt{T}\hat{\omega}_\rho(\tilde{u})}, \frac{T\left(l_T'\Omega_T^{-1}l_T\right)^{-1}}{(\hat{\omega}_\rho(\tilde{u}))^2}\right)\right\}, \end{aligned} \tag{17}$$

which is a mean and variance mixture of normal distributions.

Using $\tilde{u} \sim N\left(0, \Omega_T - l_T\left(l_T'\Omega_T^{-1}l_T\right)^{-1}l_T'\right)$ and employing operational techniques along the lines developed in Phillips (1993), we can write expression (17) in the form

$$\begin{aligned} p_{t_T}(z) &= \left[N\left(\frac{l_T'\partial q}{\sqrt{T}\hat{\omega}_\rho(\partial q)}, \frac{T\left(l_T'\Omega_T^{-1}l_T\right)^{-1}}{(\hat{\omega}_\rho(\partial q))^2}\right)\int e^{z'\tilde{u}}dP(\tilde{u})\right]_{q=0} \\ &= \left[N\left(\frac{l_T'\partial q}{\sqrt{T}\hat{\omega}_\rho(\partial q)}, \frac{T\left(l_T'\Omega_T^{-1}l_T\right)^{-1}}{(\hat{\omega}_\rho(\partial q))^2}\right)e^{q'\left\{\Omega_T - l_T\left(l_T'\Omega_T^{-1}l_T\right)^{-1}l_T'\right\}q}\right]_{q=0}. \end{aligned} \tag{18}$$

This provides a general expression for the finite sample distribution of the test statistic t_T under Gaussianity.

4 Expansion of the Nonstandard Limit Theory

This section develops asymptotic expansions of the limit distributions given in (11) and (12) as the power parameter $\rho \to \infty$. These expansions can be taken about the relevant central and noncentral chi-squared limit distributions that apply when $\rho \to \infty$, corresponding to the null and alternative hypotheses.

The expansions of the nonstandard limit distributions are of some independent interest. For instance, they can be used to deliver correction terms to the limit distributions under the null, thereby providing a mechanism for adjusting the nominal critical values provided by the usual chi-squared distribution. The latter correspond to the critical values that would be used for tests based on conventional consistent HAC estimates. As we shall see, when the $O(1/\rho)$ correction on the nominal chi-squared asymptotic critical value is implemented using this asymptotic expansion, the resulting expression provides an asymptotic justification for the continued fraction approximation suggested in PSJ$_a$ for practical testing situations.

Let $D(\cdot)$ be the cdf of a χ_1^2 variate, then

$$P\left\{\left|W(1)\Xi_\rho^{-1/2}\right| \le z\right\} = P\left\{W^2(1) \le z^2\Xi_\rho\right\} = E\left\{D(z^2\Xi_\rho)\right\}. \tag{19}$$

Observe that Ξ_ρ is a quadratic functional of a Gaussian process whose moments exist to all orders. It follows that we may develop an expansion of $E\left\{D(z^2\Xi_\rho)\right\}$ in terms of the moments of $\Xi_\rho - \mu_\rho$ where $\mu_\rho = E\left(\Xi_\rho\right) = \rho/(\rho+2)$, and $\sigma_\rho^2 = var\left(\Xi_\rho\right)$. In particular, we have

$$ED(z^2\Xi_\rho) = D(\mu_\rho z^2) + \frac{1}{2}D''(\mu_\rho z^2)z^4 E\left(\Xi_\rho - \mu_\rho\right)^2 + O\left(E(\Xi_\rho - \mu_\rho)^3\right) \tag{20}$$

as $\rho \to \infty$, where the $O(\cdot)$ term holds uniformly for any $z \in [M_l, M_u] \subset \mathbb{R}^+$ and M_l and M_u may be chosen arbitrarily small and large, respectively. As shown in Lemma 7 in the appendix, $E(\Xi_\rho - \mu_\rho)^j = O(1/\rho^{j-1})$ as $\rho \to \infty$, so that (20) gives an asymptotic series representation in increasing powers of ρ^{-1} of the limit distribution (19).

In fact, as shown in PSJ$_a$, $\Xi_\rho = \int_0^1\int_0^1 k_\rho^*(r,s)dW(r)dW(s)$, where $k_\rho^*(r,s)$ is defined by

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

The function $k_\rho(z)$ is continuous, symmetric and positive semi-definite, which guarantees the positive semi-definiteness of kernel HAC estimators defined as in (9), c.f. Newey and West (1987), Andrews (1991). The positive semi-definiteness of $k_\rho(z)$ inherits from that of $k(z)$, see Sun (2004) for a proof. The positive semi-definiteness enables the use of Mercer's theorem (e.g., see Shorack and Wellner (1986)) so that $k_\rho(r-s)$ can be represented as $k_\rho(r-s) = \sum_{n=1}^\infty \lambda_n f_n(r) f_n(s)$, where $\lambda_n > 0$ are the eigenvalues of the kernel and $f_n(x)$ are the corresponding eigenfunctions, i.e. $\lambda_n f_n(s) = \int_0^1 k(r-s)f_n(r)dr$.

Now with $\lambda_n > 0$, $k_\rho^*(r,s)$ is also positive semi-definite. This is because $k_\rho^*(r,s)$ can be written as

$$k_\rho^*(r,s) = \sum_{n=1}^{\infty} \lambda_n g_n(r) g_n(s) \text{ for any } (r,s) \in [0,1] \times [0,1]. \tag{21}$$

where $g_n(r) = f_n(r) - \int_0^1 f_n(\tau)d\tau$, and λ_n and $f_n(\cdot)$ are eigenvalues and eigenfunctions of $k_\rho(r-s)$. As a consequence, for any function $q(x) \in L^2[0,1]$, we have

$$\int_0^1 \int_0^1 q(r) k^*(r,s) q(s) dr ds = \sum_{n=1}^{\infty} \lambda_n \left(\int_0^1 g_n(r) q(r) dr \right)^2 \geq 0. \tag{22}$$

Thus, by Mercer's theorem, $k_\rho^*(r,s)$ has the representation $k_\rho^*(r,s) = \sum_{n=1}^{\infty} \lambda_n^* f_n^*(r) f_n^*(s)$ in terms of $\lambda_n^* > 0$, the eigenvalues of $k_\rho^*(r,s)$, and $f_n^*(x)$, the corresponding eigenfunctions. Using this representation, we can easily show that $\Xi_\rho = \sum_{n=1}^{\infty} \lambda_n^* Z_n^2$, where $Z_n \sim iidN(0,1)$ for $n \geq 1$. Therefore, the characteristic function of $\Xi_\rho - \mu_\rho$ is given by

$$\phi(t) = E\left\{ e^{it(\Xi_\rho - \mu_\rho)} \right\} = e^{-it\mu_\rho} \Pi_{n=1}^{\infty} \left\{ 1 - 2i\lambda_n^* t \right\}^{-1/2}. \tag{23}$$

Let $\kappa_1, \kappa_2, \kappa_3, \ldots$ be the cumulants of $\Xi_\rho - \mu_\rho$. Then

$$\kappa_1 = 0 \text{ and } \kappa_m = 2^{m-1}(m-1)! \sum_{n=1}^{\infty} (\lambda_n^*)^m \text{ for } m \geq 2. \tag{24}$$

Some algebraic manipulations show that for $m \geq 2$

$$\kappa_m = 2^{m-1}(m-1)! \int_0^1 \ldots \int_0^1 \left(\prod_{j=1}^{m} k_\rho^*(\tau_j, \tau_{j+1}) \right) d\tau_1 \cdots d\tau_m, \tag{25}$$

where $\tau_1 = \tau_{m+1}$.

With these preliminaries, we are able to develop an asymptotic expansion of $P\left\{ \left| W(1) \Xi_\rho^{-1/2} \right| < z \right\}$ as the power parameter $\rho \to \infty$. In fact, a full series expansion is possible using this method, but our purpose here requires only the leading term in the expansion.

Theorem 1 *The nonstandard limiting distribution under the null hypothesis satisfies*

$$F(z) = P\left\{ \left| W(1) \Xi_\rho^{-1/2} \right| < z \right\} = D(z^2) + \left[D''(z^2) z^4 - 2D'(z^2) z^2 \right] / \rho + O\left(1/\rho^2\right) \tag{26}$$

as $\rho \to \infty$, where the $O\left(1/\rho^2\right)$ term holds uniformly for any $z \in [M_l, M_u]$ with $0 < M_l < M_u < \infty$.

For any $\alpha \in (0,1)$, let $z_\alpha^2 \in \mathbb{R}^+$, $z_{\alpha,\rho}^2 \in \mathbb{R}^+$ such that $D(z_\alpha^2) = 1-\alpha$ and $F(z_{\alpha,\rho}) = 1-\alpha$. Then, using a power series expansion, we have

$$z_{\alpha,\rho}^2 = z_\alpha^2 + \frac{1}{2\rho}(5z_\alpha^2 + z_\alpha^4) + O\left(1/\rho^2\right) \tag{27}$$

and

$$F\left(z_\alpha^2 + \frac{1}{2\rho}(5z_\alpha^2 + z_\alpha^4)\right) = \alpha + O\left(1/\rho^2\right), \tag{28}$$

which we formalize in the following corollary.

Corollary 2 *Second order corrected critical values based on the expansion (26) are as follows:*

$$z_{\alpha,\rho}^2 = z_\alpha^2 + \frac{1}{2\rho}(5z_\alpha^2 + z_\alpha^4) + O\left(1/\rho^2\right), \tag{29}$$

and

$$z_{\alpha,\rho} = z_\alpha + \frac{1}{4\rho}(5z_\alpha + z_\alpha^3) + O\left(1/\rho^2\right), \tag{30}$$

for asymptotic chi-square and normal tests, respectively, where z_α is the nominal critical value from the standard normal distribution.

Consider as an example the case where $\alpha = 0.05$, $z_\alpha = 1.96$ and $P(W^2(1) \leq (1.96)^2) = 0.95$. Thus, for a two-sided $t^*(\hat{\omega}_\rho)$ test, the corrected critical value to the order $O\left(\rho^{-1}\right)$ at the 5% level is

$$z_{\alpha,\rho} = 1.96 + \frac{1}{4\rho}(5 \times 1.96 + 1.96^3) = 1.96 + \frac{4.3325}{\rho}. \tag{31}$$

This is also the critical value for the one-sided test (>) at the 2.5% level.

In PSJ_a, the critical values for the one-sided test were represented in terms of a hyperbola taking the following form: $z_{\alpha,\rho} = c + b/(\rho + a)$, where c is the critical value from the standard normal and a and b are constants that were computed by simulation in PSJ_a. For the 2.5% level one-sided test, the fitted curve had the form

$$z_{\alpha,\rho} = 1.96 + \frac{4.329}{\rho + 0.469} = 1.96 + \frac{4.329}{\rho} + O\left(1/\rho^2\right), \tag{32}$$

upon expansion. Clearly, (32) is remarkably close to the asymptotic expansion (31). Some calculations show that correspondingly close results hold for other significance levels.

Higher order continued fraction approximants may also be obtained in a similar way. Calculations indicate that expressions (29) and (30) are quite accurate for moderate values of ρ ($\rho \geq 5$, say). Since the limiting distributions (11) and (12) are valid for general regression models under certain conditions on the regressors (see PSJ_a), the corrected critical values $z_{\alpha,\rho}$ and $z^2_{\alpha,\rho}$ may be used for hypothesis testing in a general regression framework.

We now develop a local asymptotic power analysis using the nonstandard limit theory. Under the local alternative $H_1 : \beta = \beta_0 + cT^{-1/2}$, the limiting distribution of the test statistic $t_T = t^*(\hat{\omega}_\rho)$ for fixed ρ is $(\delta + W(1))\,\Xi_\rho^{-1/2}$. Let $G_\lambda = G(\cdot;\lambda^2)$ be the cdf of a non-central $\chi^2_1(\lambda^2)$ variate with noncentrality parameter λ^2, then we can measure the local asymptotic power by $P\{(\delta + W(1))^2 > z^2_{\alpha,\rho}\Xi_\rho\} = 1 - EG_\delta(z^2_{\alpha,\rho}\Xi_\rho)$ and develop an asymptotic approximation to this quantity. Using a Taylor series expansion similar to (20), we can prove the following theorem.

Theorem 3 *The nonstandard limiting distribution under the local alternative hypothesis* $H_1 : \beta = \beta_0 + cT^{-1/2}$ *satisfies*

$$P\left\{\left|(\delta + W(1))\,\Xi_\rho^{-1/2}\right| > z_{\alpha,\rho}\right\} = 1 - G_\delta(z^2_\alpha) - z^4_\alpha K_\delta\left(z^2_\alpha\right)/\rho + O(1/\rho^2), \tag{33}$$

as $\rho \to \infty$ *where*

$$K_\delta(z) = \sum_{j=0}^{\infty} \frac{\left(\delta^2/2\right)^j}{j!} e^{-\delta^2/2} \frac{z^{j-1/2} e^{-z/2}}{\Gamma(j+1/2) 2^{j+1/2}} \frac{j}{z} \tag{34}$$

is positive for all all z_α *and* δ.

According to Theorem 3, asymptotic test power, as measured by $P\{(\delta + W(1))^2 > z^2\Xi_\rho\}$, increases monotonically with ρ when ρ is large. Fig. 1 graphs the function $f(z_\alpha, \delta) = z^4_\alpha K_\delta\left(z^2_\alpha\right)$ for different values of z_α and δ. For a given critical value, $f(z_\alpha, \delta)$ achieves its maximum around $\delta = 2$, implying that the power increase from choosing a large ρ is greatest when the local alternative is in an intermediate neighborhood of the null hypothesis. For any given local alternative hypothesis, the function is monotonically increasing in z_α. Therefore, the power improvement due to the choice of a large ρ increases with the confidence level $1 - \alpha$.

5 Expansions of the Finite Sample Distribution

This section develops a finite sample expansion for the simple location model (c.f., Jansson, 2004). This development, like that of Section 3 and Jansson (2004), relies on Gaussianity,

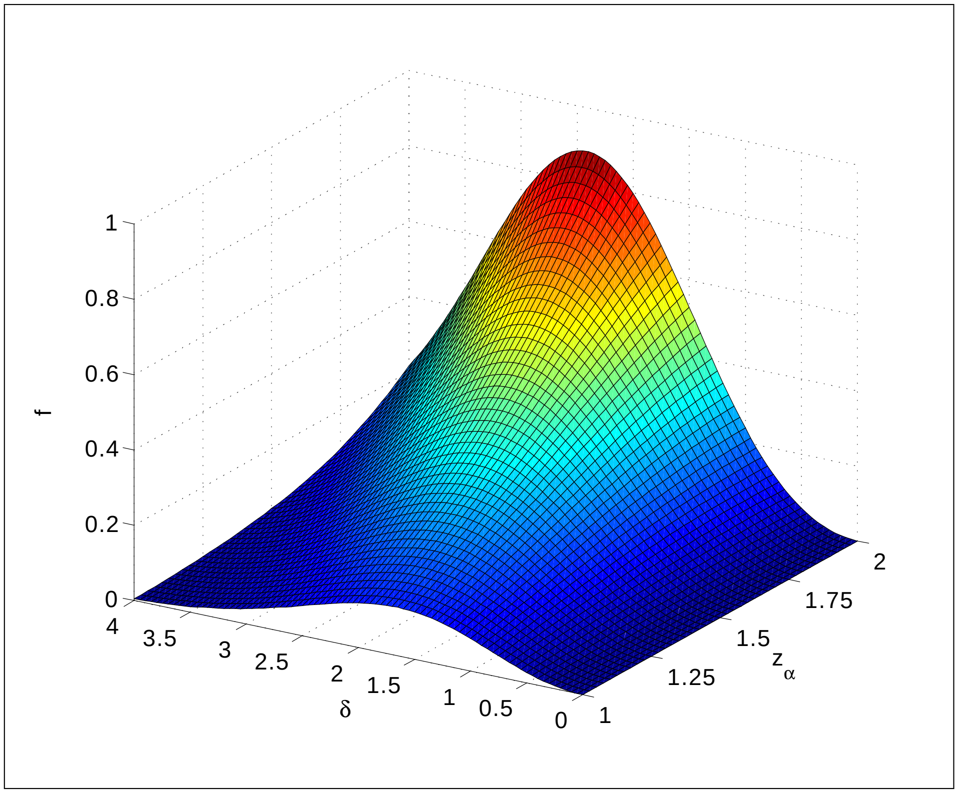

Figure 1: The graph of $f(z_\alpha, \delta) = z_\alpha^4 K_\delta\left(z_\alpha^2\right)$ as a function of z_α and δ.

which facilitates the derivations. The assumption could be relaxed by taking distributions based (for example) on Gram-Charlier expansions, but at the cost of much greater complexity (see, for example, Phillips (1980), Taniguchi and Puri (1996), Velasco and Robinson (2001)).

The following assumption on u_t facilitates the development of the higher order expansion.

Assumption 2 *u_t is a mean zero stationary Gaussian process with*

$$\sum_{h=-\infty}^{\infty} h^2 \left|\gamma(h)\right| < \infty, \tag{35}$$

where $\gamma(h) = Eu_t u_{t-h}$.

We consider the asymptotic expansion of $P\{\left|\sqrt{T}(\hat{\beta} - \beta_0)/\hat{\omega}\right| \leq z\}$ for $\hat{\omega} = \hat{\omega}_\rho$ and $\beta = \beta_0 + c/\sqrt{T}$. Depending on whether c is zero or not, such an expansion can be used to approximate the size and power of the t-test.

Recall that $T^{-1/2}l_T'\tilde{u} = \sqrt{T}(\hat{\beta}-\beta) - \sqrt{T}(\tilde{\beta}-\beta)$. But

$$\omega_T^2 := var\left(\sqrt{T}(\hat{\beta}-\beta)\right) = T^{-1}l_T'\Omega_T l_T = \omega^2 + O\left(T^{-1}\right)$$

and it follows from Grenander and Rosenblatt (1957) that

$$\tilde{\omega}_T^2 := var\left(\sqrt{T}(\tilde{\beta}-\beta)\right) = T\left(l_T'\Omega_T^{-1}l_T\right)^{-1} = \omega^2 + O\left(T^{-1}\right).$$

Therefore $T^{-1/2}l_T'\tilde{u} = N(0, O(1/T))$. Combining this and independence between $\tilde{\beta}$ and $\tilde{u}$, we have

$$\begin{aligned}
&P\left\{\sqrt{T}\left(\hat{\beta}-\beta_0\right)/\hat{\omega} \le z\right\} \\
&= P\left\{\sqrt{T}\left(\tilde{\beta}-\beta\right)/\hat{\omega} + \sqrt{T}\left(\beta-\beta_0\right)/\hat{\omega} + T^{-1/2}l_T'\tilde{u}/\hat{\omega} \le z\right\} \\
&= P\left\{\sqrt{T}\left(\tilde{\beta}-\beta\right)/\tilde{\omega}_T + c/\tilde{\omega}_T \le z\hat{\omega}/\tilde{\omega}_T - T^{-1/2}l_T'\tilde{u}/\tilde{\omega}_T\right\} \\
&= E\Phi\left(z\hat{\omega}/\tilde{\omega}_T - c/\tilde{\omega}_T - T^{-1/2}l_T'\tilde{u}/\tilde{\omega}_T\right) \\
&= E\Phi\left(z\hat{\omega}/\tilde{\omega}_T - c/\tilde{\omega}_T\right) - T^{-1/2}E\varphi\left(z\hat{\omega}/\tilde{\omega}_T - c/\tilde{\omega}_T\right)l_T'\tilde{u}/\tilde{\omega}_T + O\left(1/T\right) \\
&= P\left\{\sqrt{T}\left(\tilde{\beta}-\beta\right)/\tilde{\omega}_T + c/\tilde{\omega}_T \le z\hat{\omega}/\tilde{\omega}_T\right\} + O\left(1/T\right), \qquad (36)
\end{aligned}$$

where Φ and φ are the cdf and pdf of the standard normal distribution, respectively. The second to last equality follows because $\hat{\omega}^2$ is quadratic in $\tilde{u}$ and thus $E\varphi\left(z\hat{\omega}/\tilde{\omega}_T - c/\tilde{\omega}_T\right)l_T'\tilde{u} = 0$. In a similar fashion we find that

$$P\left\{\sqrt{T}\left(\hat{\beta}-\beta_0\right)/\hat{\omega} \le -z\right\} = P\left\{\sqrt{T}\left(\tilde{\beta}-\beta\right)/\tilde{\omega}_T + c/\tilde{\omega}_T \le -z\hat{\omega}/\tilde{\omega}_T\right\} + O\left(1/T\right).$$

Therefore

$$\begin{aligned}
F_T(z) &:= P\left\{\left|\sqrt{T}\left(\hat{\beta}-\beta_0\right)/\hat{\omega}\right| \le z\right\} \\
&= P\left\{\left[\sqrt{T}\left(\tilde{\beta}-\beta\right)/\tilde{\omega}_T + c/\tilde{\omega}_T\right]^2 \le z^2\hat{\omega}^2/\tilde{\omega}_T^2\right\} + O\left(1/T\right) \\
&= E\left\{G_\delta(z^2\hat{\omega}^2/\tilde{\omega}_T^2)\right\} = E\left\{G_\delta(z^2\varsigma_{\rho T})\right\} + O\left(1/T\right), \qquad (37)
\end{aligned}$$

where $\varsigma_{\rho T} := \left(\hat{\omega}/\omega_T\right)^2$ converges weakly to Ξ_ρ.

Since $\hat{\omega}^2 = T^{-1}\hat{u}'W_\rho\hat{u} = T^{-1}u'A_TW_\rho A_Tu$, where W_ρ is $T \times T$ with (j,s)-th element $k_\rho((j-s)/T)$ and $A_T = I_T - l_Tl_T'/T$, $\varsigma_{\rho T}$ is a quadratic form in a Gaussian vector. To evaluate $E\left\{G_\delta(z^2\varsigma_{\rho T})\right\}$, we proceed to compute the cumulants of $\varsigma_{\rho T}-\mu_{\rho T}$ for $\mu_{\rho T}:=E\varsigma_{\rho T}$. It is easy to show that the characteristic function of $\varsigma_{\rho T}-\mu_{\rho T}$ is given by

$$\phi_{\rho T}(t)=\left|I-2it\frac{\Omega_T A_T W_\rho A_T}{T\omega_T^2}\right|^{-1/2}\exp\left\{-it\mu_{\rho T}\right\},$$

where $\Omega_T=E(uu')$ and the cumulant generating function is

$$\ln\left(\phi_{\rho T}(t)\right)=-\frac{1}{2}\log\det\left(I-2it\frac{\Omega_T A_T W_\rho A_T}{T\omega_T^2}\right)-it\mu_{\rho T}:=\sum_{m=1}^{\infty}\kappa_{m,T}\frac{(it)^m}{m!},\tag{38}$$

where the $\kappa_{m,T}$ are the cumulants of $\varsigma_{\rho T}-\mu_{\rho T}$. It follows from (38) that $\kappa_{1,T}=0$ and

$$\kappa_{m,T}=2^{m-1}(m-1)!T^{-m}\left(\omega_T^2\right)^{-m}\text{Trace}\left[\left(\Omega_T A_T W_\rho A_T\right)^m\right]\text{ for }m\geq 2.\tag{39}$$

By proving $\kappa_{m,T}$ is close to κ_m in the precise sense given in Lemma 8 in the appendix, we can establish the following theorem, which gives the order of magnitude of the error in the nonstandard limit distribution of t_T as $T\to\infty$ with fixed ρ. The requirement $\rho\geq 16z^2$ on ρ that appears in the statement of the result is a technical condition in the proof that facilitates the use of a power series expansion. The requirement can be relaxed but at the cost of more tedious calculations.

Theorem 4 *Let Assumption 2 hold. If $\rho\geq 16z^2$, then*

$$F_T(z)=P\left\{\left|\left(W(1)+\delta\right)\Xi_\rho^{-1/2}\right|\leq z\right\}+O\left(1/T\right),\tag{40}$$

when $T\to\infty$ with fixed ρ.

Under the null hypothesis $H_0:\beta=\beta_0$, we have $\delta=0$. In this case, Theorem 4 is comparable to that of Jansson (2004), which was also obtained for the Gaussian location model and for kernels related to the Bartlett kernel ($\rho=1$) but with an error of $O\left(\log T/T\right)$. Theorem 4 indicates that the error in the rejection probability for tests with ρ fixed and using critical values obtained from the nonstandard limit distribution of $W(1)\Xi_\rho^{-1/2}$ is $O\left(T^{-1}\right)$. As in Jansson (2004), this represents an improvement over conventional tests based on consistent HAC estimates. Under the alternative hypothesis, $1-F_T(z)$ gives the power of the test. Theorem 4 shows that the power of the test can be approximated by $P\left\{\left|\left(\delta+W(1)\right)\Xi_\rho^{-1/2}\right|>z\right\}$ with an error of order $O(1/T)$.

Combined with Theorems 1 and 3, Theorem 4 characterizes the size and power properties of the test under the sequential limit in which T goes to infinity first for a fixed ρ and then ρ goes to infinity. Under the sequential limit theory, the size distortion of the t-test based on the corrected critical values is

$$P\left\{\left|\sqrt{T}\left(\hat{\beta}-\beta_0\right)/\hat{\omega}\right| \le z_{\alpha,\rho}\right\} - \alpha = O\left(1/\rho^2 + 1/T\right)$$

and the corresponding local asymptotic power is

$$P\left\{\left|\sqrt{T}\left(\hat{\beta}-\beta_0\right)/\hat{\omega}\right| > z_{\alpha,\rho}\right\} = 1 - G_\delta(z_\alpha^2) - z_\alpha^4 K_\delta\left(z_\alpha^2\right)/\rho + O\left(1/\rho^2 + 1/T\right).$$

To evaluate the order of size distortion, we have to compare the orders of magnitude of $1/\rho^2$ and $1/T$. Such a comparison jeopardizes the sequential nature of the limiting directions and calls for higher order approximation that allows $T \to \infty$ and $\rho \to \infty$ simultaneously.

The next theorem gives a higher order expansion of the finite sample distribution for the case where $T \to \infty$ and $\rho \to \infty$ at the same time. This expansion validates the use of the corrected critical values given in the previous section which were derived there on the basis of an expansion of the (nonstandard) limit distribution.

Theorem 5 *Let Assumption 2 hold. If* $1/\rho + \rho/T \to 0$ *as* $T \to \infty$, *then*

$$F_T(z) = G_\delta(z^2) + \left[G_\delta''(z^2)z^4 - 2G_\delta'(z^2)z^2\right]\frac{1}{\rho} - d_{\gamma T}G_\delta'(z^2)z^2\frac{\rho}{T} + O\left(\frac{1}{T} + \frac{\rho^2}{T^2} + \frac{1}{\rho^2}\right), \quad (41)$$

where $d_{\gamma T} = \omega_T^{-2}\sum_{h=-T+1}^{T-1}|h|\,\gamma(h)$.

As shown in PSJ$_a$, the bias in the HAC estimate $\hat{\omega}^2$ is of order $O(\rho/T)$ when $1/\rho + \rho \log T/T \to 0$ as $T \to \infty$, and this bias depends on the coefficient $\omega^{(1)} = \sum_{h=-\infty}^{\infty}|h|\,\gamma(h)$, which is the limit of $\sum_{h=-T+1}^{T-1}|h|\,\gamma(h)$. As is apparent from (41), the bias in estimating ω^2 manifests itself in the limiting distribution of the test statistic under both the null and local alternative hypotheses.

Under the null hypothesis, $\delta = 0$ and $G_\delta(\cdot) = D(\cdot)$, so

$$F_T(z) = D(z^2) + \left[D''(z^2)z^4 - 2D'(z^2)z^2\right]\frac{1}{\rho} - d_{\gamma T}D'(z^2)z^2\frac{\rho}{T} + O\left(\frac{1}{T} + \frac{\rho^2}{T^2} + \frac{1}{\rho^2}\right) \quad (42)$$

Note that the leading two terms (up to order $O(1/\rho)$) in this expansion are the same as those in the corresponding expansion of the limit distribution $F(z)$ given in (26) above.

Thus, use of the corrected critical values given in (29) and (30), which take account of terms up to order $O(1/\rho)$, should lead to size improvements when $\rho^2/T \to 0$, in a similar way to those attained by a KVB type test with fixed ρ, as shown in Theorem 4 above and Jansson (2004).

The third term in the expansion (42) is $O\left(T^{-1}\right)$ when ρ is fixed. When ρ increases with T, this term provides an asymptotic measure of the size distortion in tests based on the use of the first two terms of (42), or equivalently those based on the nonstandard limit theory, at least to order $O\left(\rho^{-1}\right)$. Thus, the third term of (42) approximately measures how satisfactory the corrected critical values given by (29) and (30) are for any given values of ρ and T.

Under the local alternative hypothesis, the power of the test based on the corrected critical value is $1 - F_T(z_{\alpha,\rho})$. Theorem 5 shows that $F_T(z_{\alpha,\rho})$ can be approximated by

$$G_\delta(z_{\alpha,\rho}^2) + \left[G_\delta''(z_{\alpha,\rho}^2)z_{\alpha,\rho}^4 - 2G_\delta'(z_{\alpha,\rho}^2)z_{\alpha,\rho}^4\right]\frac{1}{\rho} - d_{\gamma T}G_\delta'(z_{\alpha,\rho}^2)z_{\alpha,\rho}^2\frac{\rho}{T}$$

with an approximation error of order $O\left(1/T + \rho^2/T^2 + 1/\rho^2\right)$.

We formalize the results on the size distortion and local power expansion in the following corollary.

Corollary 6 *Let Assumption 2 hold. If* $1/\rho + \rho/T \to 0$ *as* $T \to \infty$*, then*

(a) the size distortion of the t-test based on the second order corrected critical values is

$$(1 - F_T(z_{\alpha,\rho})) - \alpha = d_{\gamma T}D'(z^2)z^2\frac{\rho}{T} + O\left(\frac{1}{T} + \frac{\rho^2}{T^2} + \frac{1}{\rho^2}\right). \tag{43}$$

(b) under the local alternative $H_1 : \beta = \beta_0 + c/\sqrt{T}$*, the power of the t-test based on the second order corrected critical values is*

$$\begin{aligned} &P\left(\left|\sqrt{T}(\hat{\beta} - \beta_0)/\hat{\omega}\right| \geq z_{\alpha,\rho}\right) \\ &= 1 - G_\delta(z_\alpha^2) - z_\alpha^4 K_\delta\left(z_\alpha^2\right)\frac{1}{\rho} + d_{\gamma T}G_\delta'(z_\alpha^2)z_\alpha^2\frac{\rho}{T} + O\left(\frac{1}{T} + \frac{\rho^2}{T^2} + \frac{1}{\rho^2}\right). \end{aligned} \tag{44}$$

It is clear from the proof of the theorem that the size distortion of the t-test based on the nonstandard limiting theory can also be approximated by $d_{\gamma T}D'(z^2)z^2\rho/T$ with an approximation error of order $O\left(1/T + \rho^2/T^2 + 1/\rho^2\right)$. Therefore, the critical values from the nonstandard limiting distribution provide a second order correction on the critical values from the standard normal distribution. By mimicking the randomness of the denominator of the t-statistic, the nonstandard limit theory provides a more accurate approximation to the finite sample distribution. However, just as with the standard limit theory, the nonstandard limit theory does not deal with the bias problem of long run variance estimation.

Comparing (44) with (33), we get an additional term which arises from the asymptotic bias of the long run variance estimator. For economic time series, it is typical that $d_{\gamma T} > 0$, as discussed below. So this additional term also increases monotonically with ρ, thereby increasing power. Of course, size distortion also tends to increase with ρ as is apparent in (43), so we now need find a value of ρ to balance size distortion with increasing power. Practical suggestions for choosing ρ are given in the next section.

6 Optimal Choice of ρ

When estimating the long run variance, PSJ$_a$ show there is an optimal choice of ρ which minimizes the asymptotic mean squared error of the estimate and gives an optimal expansion rate of $O\left(T^{2/3}\right)$ for ρ in terms of the sample size T. Developing an optimal choice of ρ for semiparametric testing is not as straightforward. In what follows we provide one possible approach to constructing an optimizing criterion that is based on balancing the type I and type II errors induced by various choices of ρ.

Using the expansion (43), the type I error for a nominal size α test can be expressed as

$$1 - F_T(z_{\alpha,\rho}) = \alpha + d_{\gamma T} D'(z_\alpha^2) z_\alpha^2 \frac{\rho}{T} + o\left(\frac{1}{\rho} + \frac{\rho}{T}\right), \tag{45}$$

Similarly, from (44), the type II error has the form

$$G_\delta(z_\alpha^2) + z_\alpha^4 K_\delta(z_\alpha^2) \frac{1}{\rho} - d_{\gamma T} G'_\delta(z_\alpha^2) z_\alpha^2 \frac{\rho}{T} + o\left(\frac{1}{\rho} + \frac{\rho}{T}\right). \tag{46}$$

A loss function for the test may be constructed based on the following three factors: (i) The magnitude of the type I error, as measured by the second term of (45); (ii) The magnitude of the type II error, as measured by the $O\left(1/\rho\right)$ and $O(\rho/T)$ terms in (46); and (iii) the relative importance of type I and type II errors. These two types of errors are related to the size and power of the test and we use both sets of terminology which shall not cause any confusion.

For most economic time series we can expect that $d_{\gamma T} > 0$ and then both $d_{\gamma T} D'(z_\alpha^2) z_\alpha^2 > 0$ and $d_{\gamma T} G'_\delta(z_\alpha^2) z_\alpha^2 > 0$. Hence, the ρ/T term in (45) typically leads to upward size distortion (large type I error) in testing, as found in simulations by PSJ$_a$ and this upward distortion corresponds to that found in work by KV and others on the use of conventional HAC estimates in testing. On the other hand, the ρ/T term in (46) indicates that there is a corresponding increase in power of a similar magnitude by virtue of the third term of (46) as the type II error is correspondingly reduced. Indeed, for $\delta > 0$ this power increase will generally exceed the upward size distortion from $d_{\gamma T} D'(z_\alpha^2) z_\alpha^2$ because $G'_\delta(z_\alpha^2) > D'(z_\alpha^2)$ for $\delta \in (0, 7.5)$ and $z_\alpha = 1.645, 1.960$ or 2.580. Fig. 2 graphs the ratio $G'_\delta(z^2)/D'(z^2)$ against δ for different values of z, illustrating the relative magnitude of $G'_\delta(z_\alpha^2)$ and $D'(z_\alpha^2)$. The situation is further complicated by the fact that there is an additional term in the type II error of $O(1/\rho)$ that affects power. As we have seen earlier, $K_\delta(z_\alpha^2) > 0$ so that the second term of (46) leads to an increase in power (or a reduction in the type II error) as ρ increases. Thus, power generally increases with ρ for two reasons – one from the nonstandard limit theory and the other from the (typical) downward bias in estimating the long run variance.

The case of $d_{\gamma T} < 0$ usually arises where there is negative serial correlation in the errors and so tends to be less typical for economic time series. In such a case, (45) shows that type I error is capped at the nominal level α, at least up to an error of $o(\rho/T)$. Test size is then conservative and the goal in selection of ρ is to cap the type I error while attempting to reduce the type II error as much as possible. In this case, we have $d_{\gamma T}\left\{D'(z_\alpha^2) - G'_\delta(z_\alpha^2)\right\} > 0$ provided that δ is not too large (Fig. 2).

These considerations suggest that a loss function may be constructed by taking a suitable weighted average of the type I and type II errors given in (45) and (46). The loss function below distinguishes the two cases where $d_{\gamma T} > 0$ and $d_{\gamma T} < 0$ in terms of the weights employed. We define

$$\begin{aligned} & L(\rho; \delta, T, z_\alpha) \\ & = A_T^I\left(\alpha + d_{\gamma T} D'(z_\alpha^2) z_\alpha^2 \frac{\rho}{T}\right) + A_T^{II}\left(G_\delta(z_\alpha^2) + \frac{1}{\rho} z_\alpha^4 K_\delta(z_\alpha^2) - \frac{\rho}{T} d_{\gamma T} G'_\delta(z_\alpha^2) z_\alpha^2\right) \\ & = \begin{cases} d_{\gamma T}\left\{A_T^I D'(z_\alpha^2) - A_T^{II} G'_\delta(z_\alpha^2)\right\} z_\alpha^2 \frac{\rho}{T} + A_T^{II} z_\alpha^4 K_\delta(z_\alpha^2) \frac{1}{\rho} + C_T^+, & d_{\gamma T} > 0 \\ d_{\gamma T}\left\{D'(z_\alpha^2) - G'_\delta(z_\alpha^2)\right\} z_\alpha^2 \frac{\rho}{T} + z_\alpha^4 K_\delta(z_\alpha^2) \frac{1}{\rho} + C^-, & d_{\gamma T} < 0 \end{cases} \end{aligned} \tag{47}$$

where $C_T^+ = A_T^I \alpha + A_T^{II} G_\delta(z_\alpha^2)$ and $C^- = \alpha + G_\delta(z_\alpha^2)$ do not depend on ρ, and A_T^I and A_T^{II}

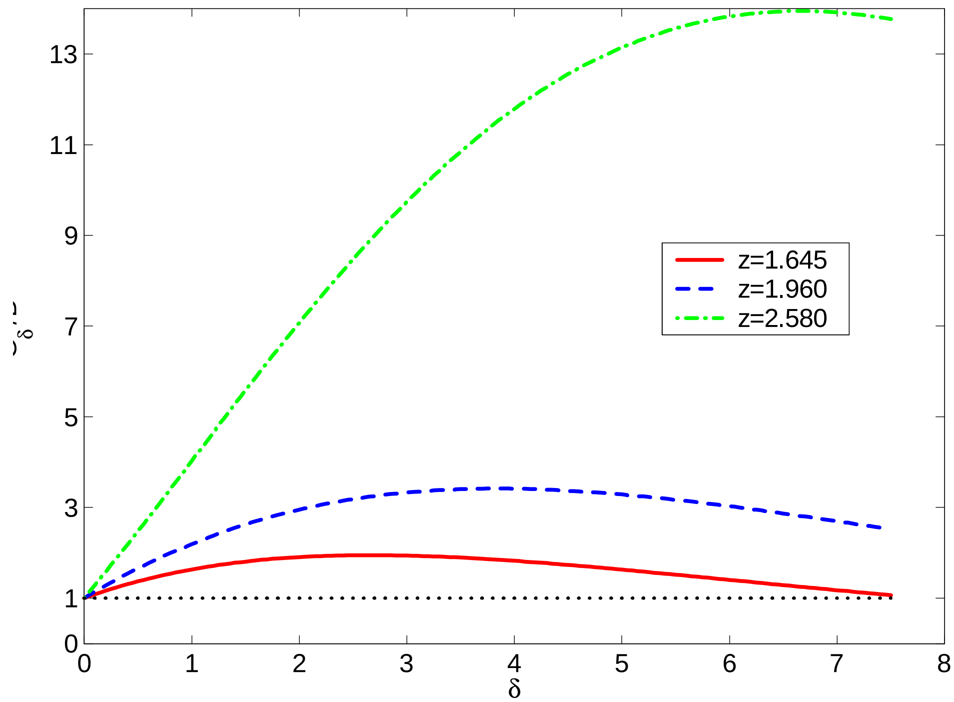

Figure 2: The graph of $G'_\delta(z^2)/D'_\delta(z^2)$ as a function of δ for different values of z

are weights on the type I and II errors that are permitted to be functions of the sample size T. Obviously, the loss $L(\rho;\delta,T,z_\alpha)$ is specified for a particular value of δ and this function could be adjusted in a simple way so that the type II error is averaged over a range of values of δ with respect to some (prior) distribution over alternatives.

The idea behind the form of the loss function (47) is that priority may be placed on capping size in testing, so that when the type I error is distorted toward over-rejection (as it is when $d_{\gamma T} > 0$) weights are introduced to amplify the loss from the type I error relative to the loss from the type II error, leading to the inequality $A_T^I > A_T^{II}$. When $A_T^I D'(z_\alpha^2) - A_T^{II} G'_\delta(z_\alpha^2) > 0$, the loss function $L(\rho;\delta,T,z_\alpha)$ is then minimized for the following choice of ρ

$$\rho_{\text{opt}} = \begin{cases} \left[\left(\frac{A_T^{II} z_\alpha^2 K_\delta(z_\alpha^2)}{d_{\gamma T}\left\{A_T^I D'(z_\alpha^2) - A_T^{II} G'_\delta(z_\alpha^2)\right\}}\right)^{1/2} T^{1/2}\right] \mathbf{1}\left\{d_{\gamma T} > 0\right\}, \\ \left[\left(\frac{z_\alpha^2 K_\delta(z_\alpha^2)}{d_{\gamma T}\left\{D'(z_\alpha^2) - G'_\delta(z_\alpha^2)\right\}}\right)^{1/2} T^{1/2}\right] \mathbf{1}\left\{d_{\gamma T} < 0\right\}, \end{cases} \tag{48}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. Since $d_{\gamma T} = d_\gamma(1 + O(1/T))$ where

$d_\gamma := \sum_{h=-\infty}^{\infty} |h|\,\gamma(h)/\omega^2$, d_γ approximately measures the effects of the bias in long run variance estimation on the size and power of the t-test. As a result, $d_{\gamma T}$ in (48) can be replaced by d_γ when d_γ is available.

If the relative weight satisfies

$$a_T = \frac{A_T^I}{A_T^{II}} \to \infty \text{ as } T \to \infty, \tag{49}$$

we then have

$$\rho_{\text{opt}} = \left(\frac{z_\alpha^2 K_\delta(z_\alpha^2)}{d_{\gamma T} D'(z_\alpha^2)} \right)^{1/2} \left(\frac{T}{a_T} \right)^{1/2} [1 + o(1)], \text{ for } d_{\gamma T} > 0. \tag{50}$$

Fixed ρ rules may then be interpreted as assigning relative weight $a_T = O(T)$ in the loss function so that the emphasis in tests based on such rules is size accuracy, at least when we expect the size distortion to be toward over-rejection. This gives us an interpretation of fixed ρ rules in terms of the loss perceived by the econometrician in this case. Similar considerations would apply in a development along these lines for the fixed bandwidth rules suggested in KV for untruncated conventional kernel estimates.

Otherwise, when a_T is large enough to ensure $A_T^I D'(z_\alpha^2) - A_T^{II} G'_\delta(z_\alpha^2) > 0$, (50) leads to the expansion rate $\rho_{\text{opt}} = O\left(T^{1/2}\right)$. Fig. 2 shows that when $a_T \geq 14$ and the significance level is less than 1%, $A_T^I D'(z_\alpha^2) - A_T^{II} G'_\delta(z_\alpha^2)$ is positive for a broad range of values of δ. Within this general framework, ρ may be fixed or expand with T up to an $O\left(T^{1/2}\right)$ rate corresponding to the relative importance that is placed in the loss function on size and power. Also, according to formula (48) for the case $d_{\gamma T} > 0$, ρ_{opt} decreases as size distortion (measured by $d_{\gamma T} z_\alpha^2 D'(z_\alpha^2)$ or the parameter $d_{\gamma T}$) increases. So, again, a smaller ρ is preferred when size distortion becomes more important given the specific autocorrelation structure measured via its effect on $d_{\gamma T}$.

Observe that when $\rho = O(T/a_T)^{1/2}$, size distortion is of order $O(T a_T)^{-1/2}$ rather than $O\left(T^{-1}\right)$, as it is when ρ is fixed. Thus, the use of $\rho = \rho_{\text{opt}}$ for a finite a_T involves some compromise by allowing the error order in the rejection probability to be somewhat larger in order to achieve higher power. Such compromise is an inevitable consequence of balancing the two elements in the loss function (47).

In cases where size is expected to be conservatively biased (i.e., when $d_{\gamma T} < 0$), the rule in (48) balances size distortion and power reduction with the same weights in the loss function. That is, $A_T^I = A_T^{II} = 1$ in this case. This weighting might be justified by the argument that in the case of a conservative bias, test size is effectively capped and so additional weighting on size distortion is not required. In this case, it seems worthwhile to take advantage of the extra gains in power from increasing ρ. Correspondingly, the expansion rate for ρ_{opt} in (48) in this case turns out to be $O\left(T^{1/2}\right)$.

The formula for ρ_{opt} involves the unknown parameter d_γ, which can be estimated nonparametrically or by a standard plug-in procedure based on a simple model like an AR(1). Both methods achieve a valid order of magnitude and the procedure is obviously analogous to conventional data-driven methods for HAC estimation.

To sum up, the value of ρ which minimizes size distortion in conjunction with raising power as much as possible has an expansion rate of $O\left(T/a_T\right)^{1/2}$, which is at most $O\left(T^{1/2}\right)$ This rate may be compared with the optimal rate of $O\left(T^{2/3}\right)$ which applies when minimizing the mean squared error of estimation of the corresponding HAC estimate, $\hat{\omega}^2$, itself (PSJ$_a$). Thus, the MSE optimal values of ρ for HAC estimation are much larger as $T \to \infty$ than those which are most suited for statistical testing. In effect, optimal HAC estimation tolerates more bias in estimation in order to reduce variance in estimation. In contrast, optimal ρ selection in HAR testing undersmooths the long run variance estimate to reduce bias and allows for greater variance in long run variance estimation through higher order adjustments to the nominal asymptotic critical values or by direct use of the nonstandard limit distribution.

7 Simulation Evidence

In this section, we first provide some simulation evidence on the accuracy of the size approximation given in Corollary 6 and then investigate the performance of the t-test based on the plug-in procedure that optimizes the loss function constructed in the previous section.

7.1 Estimation of the ERP

We consider the simple location model $y_t = \beta_0 + u_t$ as the data generating process, where β_0 is set to be zero without the loss of generality and u_t follows the Gaussian ARMA(1,1) process

$$u_t = \phi u_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1},\ \varepsilon_t \sim iidN(0, \sigma^2). \tag{51}$$

In the simulations, we take processes corresponding to all possible combinations of the following parameter choices:

$$\phi = 0.9, 0.6, 0.3, 0, -0.3, -0.6, -0.9 \text{ and}$$
$$\theta = 0.9, 0.6, 0.3, 0, -0.3, -0.6, -0.9$$

We consider the sample sizes $T = 50, 100, 200$ and $50,000$ replications are used in all cases. We compute the ERP or size distortion (empirical size − nominal size) of the t^*-test for testing the null hypothesis that $\beta = 0$ against the alternative that $\beta \neq 0$. To illustrate how well the asymptotic theory works as the power parameter ρ varies in this finite sample, we compute the size distortion for $\rho = 1, 2, ..., 50$. We set the asymptotic significance level to 10%, and compute the critical values using our hyperbola formula obtained in PSJ$_a$

$$z_{\alpha,\rho} = 1.645 + \frac{3.127}{\rho + 0.457}. \tag{52}$$

Using the critical value

$$z_{\alpha,\rho} = 1.645 + \frac{3.169}{\rho}$$

as given by Corollary 2 produces essentially the same result.

To compare the results of our t^*-test with those of the conventional t-test, we also compute the size distortion for the conventional t-test constructed using sharp origin kernels. In this case, we use the usual standard normal critical value of 1.645 for all values of ρ. Note that we use the same statistic for the t^*-test and t-test. The only difference is that the t^*-test uses critical values from the preceding hyperbola formula while the t-test uses critical values from the standard normal.

We report only the results for sample size $T = 50$ as the results for other sample sizes are qualitatively similar. The results are displayed in Figs. 3–7, which graph the size distortion against the power parameter ρ. We present only a few cases for illustration. There are several noticeable patterns. First, the size distortion curve for the t-test is always above that for the t^*-test. As a result, when the size distortion for the t*-test is positive, the new fixed-ρ asymptotics provides a better approximation to the null distribution than the conventional large-ρ asymptotics. When the size distortion for the t*-test is negative, the fixed-ρ asymptotics continue to give a better approximation when ρ is small but its performance is slightly worse than the large-ρ asymptotics when ρ is large.

This finding confirms the implication of Theorem 4. Second, Corollary 6 provides very good approximations to the finite sample size distortions. For the $ARMA(1,1)$ process (51), we have

$$\sum_{h=-\infty}^{\infty} |h| \, \gamma(h) = \frac{2(1+\phi\theta)(\phi+\theta)}{(1-\phi)^3(1+\phi)}\sigma^2,$$

and $\omega^2 = (1+\theta)^2 (1-\phi)^{-2} \sigma^2$. Thus, according to Corollary 6, size distortion is approximated by

$$\frac{2(1+\phi\theta)(\phi+\theta)}{(1-\phi^2)(1+\theta)^2} D'(z^2) z^2 \frac{\rho}{T} + O\left(\frac{1}{T} + \frac{\rho^2}{T^2} + \frac{1}{\rho^2}\right). \tag{53}$$

When $\theta = 0$, the coefficient for $D'(z^2)z^2\rho/T$ becomes $2\phi/(1-\phi^2)$, indicating that the direction of the size distortion depends on the sign of ϕ and its absolute value grows dramatically with $|\phi|$. Similarly, when $\phi = 0$, the direction of the size distortion depends on the sign of θ and its absolute value grows with $|\theta|$. These theoretical predictions are supported by the simulation results.

In view of the form of (53), for given T size distortion can be approximated to the first order as a linear function of the power parameter ρ of the form

$$erp = c_0 + \rho c_1,$$

which is conformable with Corollary 6. Table 1 gives OLS estimates of c_0 and c_1, the standard error and the R^2 of the linear approximation. Apparently, the linear function fits the size distortion quite well, even for persistent error processes. Note that the ERP has a lower bound -0.10. For AR and MA processes with large negative AR or MA parameters and some ARMA processes, the ERP reaches the lower bound for large values of ρ. Simulation results show that for these cases the ERP curve is approximately linear for small ρ and then becomes flat for large ρ. This nonlinear feature renders linear fitting less satisfactory.

Table 1 and Fig. 4 reveal that for an AR(1) process with a large absolute AR parameter there is more curvature in the finite sample size distortion as ρ increases. This is because higher order terms in (53) become more important in such cases. As is clear from Corollary 6, the error in the size distortion formula (53) involves additional terms of order $O\left(1/\rho^2\right)$ and $O\left(\rho^2/T^2\right)$. The latter terms become important for large values of ρ, indicating that the approximation suggested by (43) is most likely to be appropriate when ρ is moderately valued. Calculations reveal that when $\rho = O\left(\sqrt{T}\right)$, the formula $2(1+\phi\theta)(\phi+\theta)\left(\left(1-\phi^2\right)(1+\theta)^2\right)^{-1} D'(z^2)z^2/T$ provides a good approximation of the slope.

Table 1: OLS Estimation of the Linear Function $erp = c_0 + \rho c_1$
for ARMA(1,1) Processes with AR parameter ϕ and MA parameter θ

(ϕ, θ)	$(0.9, 0)$	$(0.6, 0)$	$(0.3, 0)$	$(-0.3, 0)$	$(-0.6, 0)$	$(-0.9, 0)$
c_0	0.2516	0.0502	0.0157	−0.0178	−0.0439	−0.0859
c_1	0.0067	0.0037	0.0015	−0.0008	−0.0011	−0.0004
$s.e.$	0.0309	0.0070	0.0024	0.0020	0.0053	0.0037
R^2	0.91	0.98	0.99	0.98	0.90	0.66
(ϕ, θ)	$(0, 0.9)$	$(0, 0.6)$	$(0, 0.3)$	$(-0.3, 0)$	$(-0.6, 0)$	$(-0.9, 0)$
c_0	0.0108	0.0105	0.0077	−0.0303	−0.0841	−0.1000
c_1	0.0015	0.0013	0.0009	−0.0009	−0.0004	0.0000
$s.e.$	0.0011	0.0012	0.0010	0.0034	0.0040	0.0000
R^2	1.00	1.00	0.99	0.93	0.68	1.00
(ϕ, θ)	$(-0.6, 0.3)$	$(0.3, -0.6)$	$(0.3, 0.3)$	$(0, 0)$	$(0.6, -0.3)$	$(-0.3, 0.6)$
c_0	−0.0125	−0.0568	0.0207	−0.0010	0.0411	0.0042
c_1	−0.0008	−0.0007	0.0021	0.0001	0.0026	0.0006
$s.e.$	0.0002	0.0041	0.0028	0.0004	0.0068	0.0006
R^2	0.98	0.87	0.99	0.85	0.97	1.00

7.2 Performance of the Plug-in Procedure

We provide some brief illustrations on the new plug-in procedure for selecting ρ in practical work. We employ the AR(1) plug-in procedure, which for the process $v_t = \phi v_{t-1} + e_t$, leads to $d_\gamma = \omega^{-2} \sum_{h=-\infty}^{+\infty} |h|\,\gamma(h) = 2\phi/(1-\phi^2)$ as shown in (53). We consider the simple local model with Gaussian ARMA(1,1) errors:

$$y = \beta + c/\sqrt{T} + u_t$$

where $c = 0$ or 4.1075 and

$$u_t = \phi u_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}, \varepsilon_t \sim iidN(0,1). \tag{54}$$

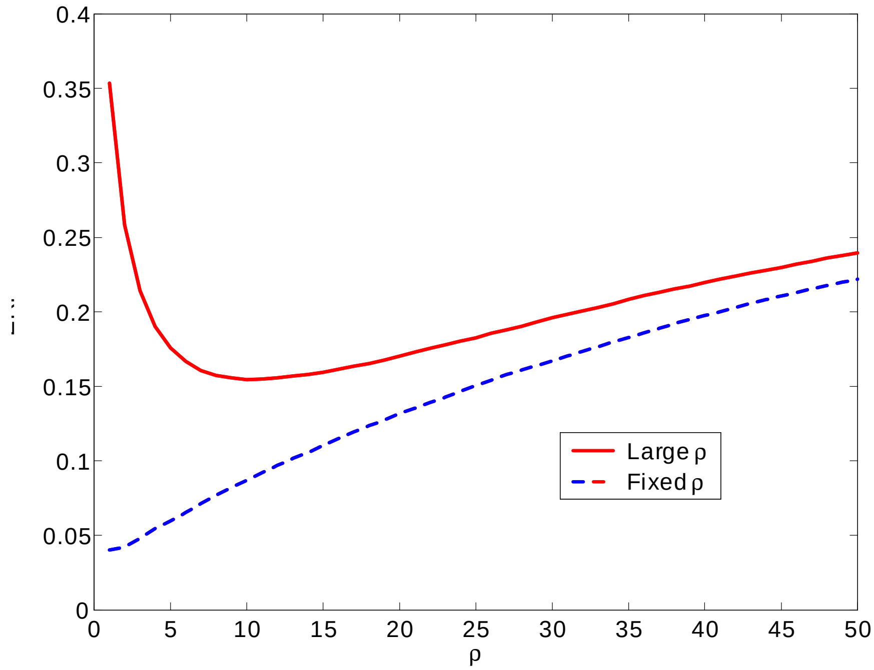

Figure 3: Size distortion for ARMA(1,1) Errors with $(\phi, \theta) = (0.6, 0)$ and $T = 50$

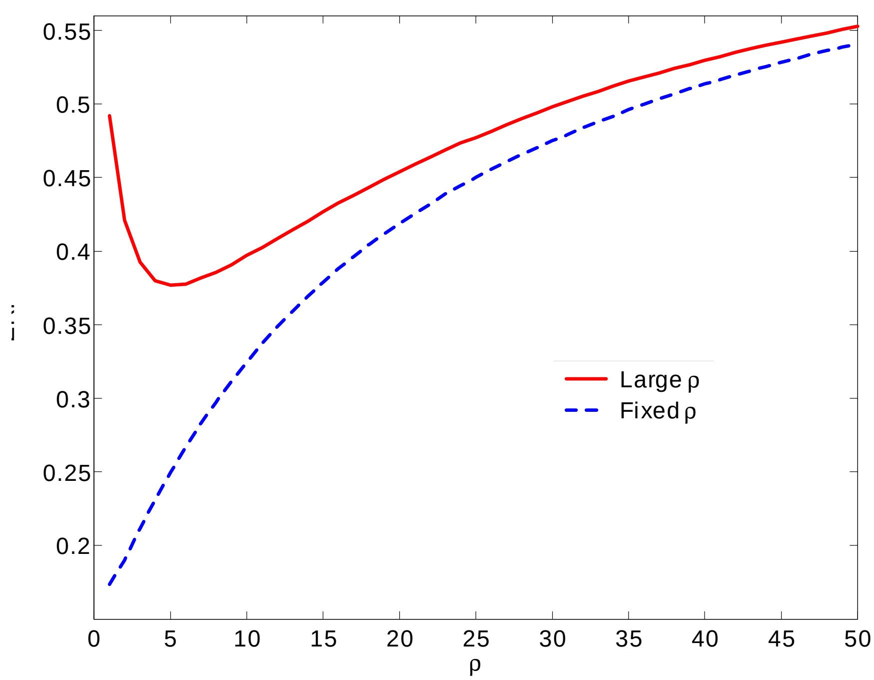

Figure 4: Size distortion for ARMA(1,1) Errors with $(\phi, \theta) = (0.9, 0)$ and $T = 50$

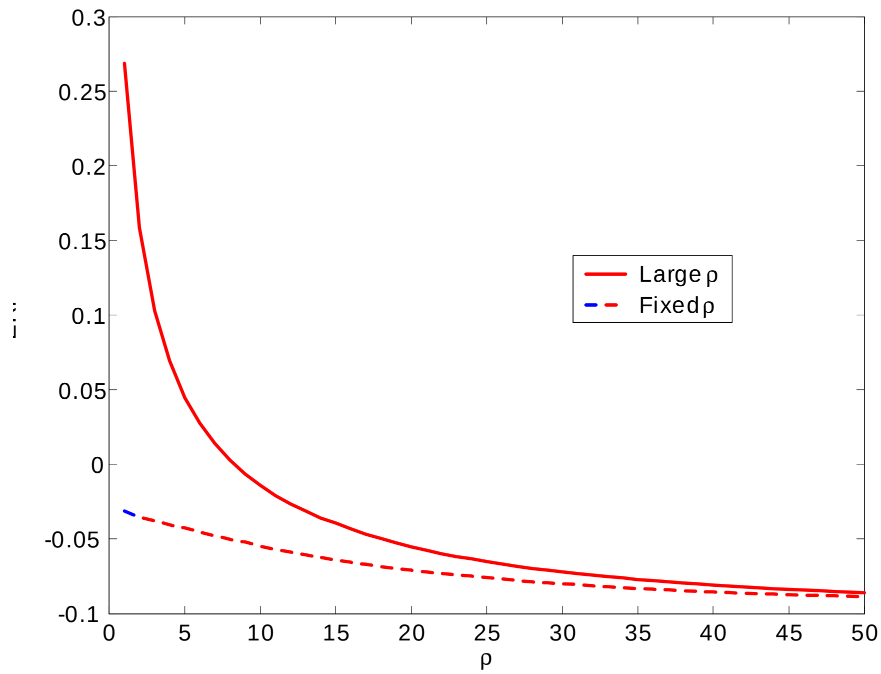

Figure 5: Size distortion for ARMA(1,1) Errors with $(\phi, \theta) = (-0.6, 0)$ and $T = 50$

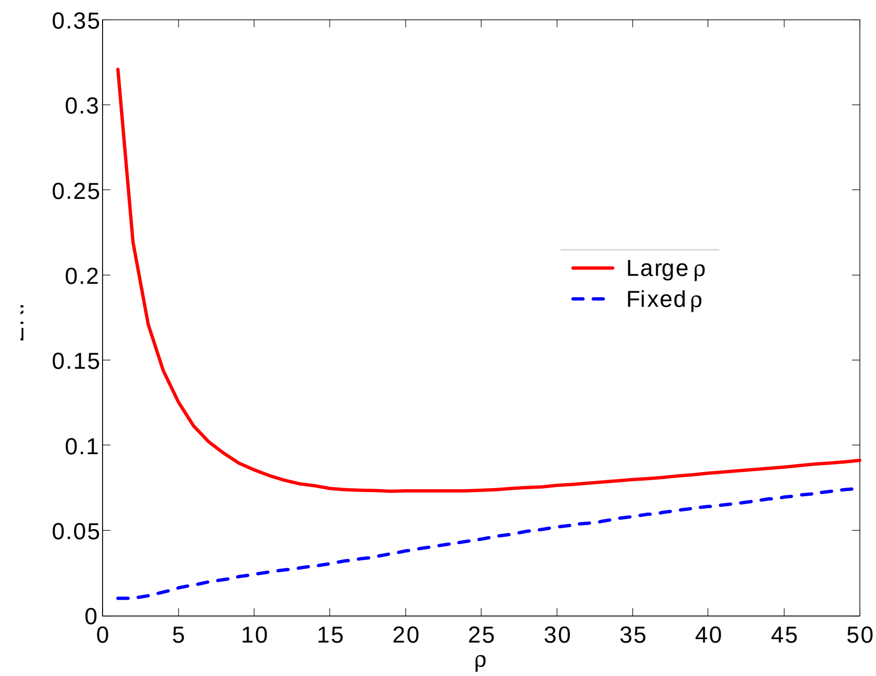

Figure 6: Size distortion for ARMA(1,1) Errors with $(\phi, \theta) = (0.0, 0.6)$ and $T = 50$

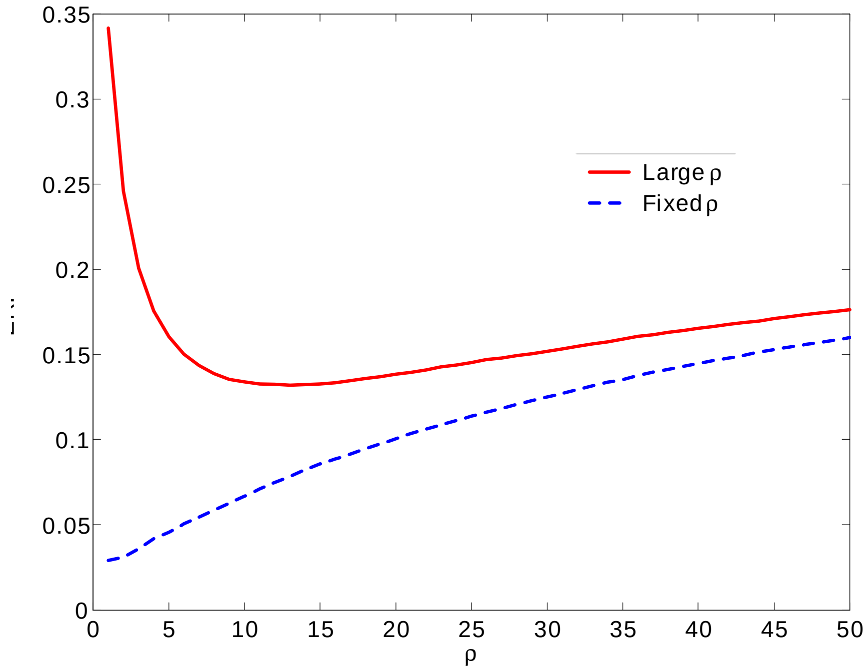

Figure 7: Size distortion for ARMA(1,1) Errors with $(\phi, \theta) = (0.6, -0.3)$ and $T = 50$

Under the null $c = 0$ and under the local alternative $c = 4.1075$. The latter value of c is chosen such that when $\phi = 0.6$ and $\theta = 0$, the local asymptotic power of the t-test is 50%. In other words, $c = 4.1075$ solves $P\{|(c(1-\phi) + W(1))| \geq 1.645\} = 50\%$ for $\phi = 0.6$. As before, we consider three sample sizes $T = 50, 100$ and 200.

For each data generating process, we obtain an estimate $\hat{\phi}$ of the AR coefficient by fitting an AR(1) model to the demeaned time series. Given the estimate $\hat{\phi}$, $d_{\gamma T}$ can be estimated by $\hat{d}_\gamma = \left(2\hat{\phi}\right)/(1-\hat{\phi}^2)$. As suggested in Section 6, we use symmetric weights $A_T^I = A_T^{II} = 1$ when $\hat{d}_\gamma < 0$ and use asymmetric weights $A_T^I = 10, 20$, or T and $A_T^{II} = 1$ when $\hat{d}_\gamma > 0$. We set the significance level to be $\alpha = 10\%$ and the corresponding nominal critical value for the two sided test is $z_\alpha = 1.645$. For all DGPs, we let $\delta = 2$ in computing the optimal power parameter. More specifically,

$$\hat{\rho}_{\text{opt}} = \begin{cases} \left[\left(\frac{(1-\hat{\phi}^2)}{(2\hat{\phi})} \frac{z_\alpha^2 K_\delta(z_\alpha^2)}{\left\{a_T D'(z_\alpha^2) - G'_\delta(z_\alpha^2)\right\}}\right)^{1/2} T^{1/2}\right] \mathbf{1}\left\{\hat{d}_\gamma > 0\right\}, \\ \left[\left(\frac{(1-\hat{\phi}^2)}{(2\hat{\phi})} \frac{z_\alpha^2 K_\delta(z_\alpha^2)}{\left\{D'(z_\alpha^2) - G'_\delta(z_\alpha^2)\right\}}\right)^{1/2} T^{1/2}\right] \mathbf{1}\left\{\hat{d}_\gamma < 0\right\}, \end{cases} \quad (55)$$

for $z_\alpha = 1.645, \delta = 2, a_T = 10, 20$, or T. For each choice of a_T, we obtain $\hat{\rho}_{\text{opt}}$ and use it to construct the LRV estimate and corresponding t*-statistic. We reject the null hypothesis if

$$|t^*| \geq 1.645 + 3.169/\hat{\rho}_{\text{opt}},$$

the corrected critical value from Corollary 2. We may use the critical value from the hyperbola formula as given in (52) but the results are essentially the same. Using 50000 replications, we computed the empirical type I errors (when $c = 0$) and type II errors (when $c = 4.1075$). Depending on whether the true d_γ is positive or not, we calculate the empirical loss by taking a weighted average of the type I and type II errors. When $d_\gamma > 0$, the weights associated with the type I and II errors are $a_T/(1+a_T)$ and $1/(1+a_T)$, respectively. When $d_\gamma > 0$, we use equal weights so that the weights are 50% for both types of errors.

For comparative purposes, we also compute the empirical loss function when the power parameter is the 'optimal' one that minimizes the asymptotic mean squared errors of the LRV estimator. The formula for this power parameter is given in PSJ$_a$ and a plug-in version is

$$\hat{\rho}_{\text{MSE}} = \left[\left(\frac{(1-\hat{\phi}^2)^2}{4\hat{\phi}^2} \right)^{1/3} T^{2/3} \right].$$

Table 2 reports the empirical loss only for the sample size $T = 100$, as it is representative of other sample sizes. It is clear that the new plug-in procedure incurs a significantly smaller loss than the conventional plug-in procedure when $d_\gamma > 0$, which is typical for economic time series. This is true for all values of a_T and parameter combinations considered. Simulation results not reported show that the superior performance of the new procedure also holds for smaller values of a_T such as $a_T = 2$, although the advantage of the new procedure is reduced. When $d_\gamma < 0$, the new plug-in procedure is slightly outperformed by the conventional plug-in procedure. In this case, the reduction in the type II error from choosing a large value of ρ outweighs the increase in the type I error, as the type I error is capped by the nominal size.

In sum, the simulation results in this and previous subsections show that the size and power (or type I and type II errors) expansions given in Corollary 6 provide satisfactory approximations to the finite sample size and power. The simulation results also reveal that the new plug-in procedure works well in terms of incurring a smaller loss than the conventional plug-in procedure.

Table 2: Empirical Loss Using Different Plug-in ρ's for ARMA(1,1) Processes with AR parameter ϕ and MA parameter θ

	$a_T = 10$		$a_T = 20$		$a_T = T$	
(ϕ, θ)	$\hat{\rho}_{\text{opt}}$	$\hat{\rho}_{\text{MSE}}$	$\hat{\rho}_{\text{opt}}$	$\hat{\rho}_{\text{MSE}}$	$\hat{\rho}_{\text{opt}}$	$\hat{\rho}_{\text{MSE}}$
$(0.9, 0)$	0.2515	0.3900	0.2069	0.3789	0.1363	0.3691
$(0.6, 0)$	0.1815	0.2450	0.1448	0.2251	0.0994	0.2078
$(0.3, 0)$	0.1546	0.1864	0.1288	0.1709	0.0953	0.1575
$(-0.3, 0)$	0.1490	0.1421	0.1491	0.1421	0.1491	0.1421
$(-0.6, 0)$	0.0885	0.0852	0.0885	0.0852	0.0885	0.0852
$(-0.9, 0)$	0.0693	0.0584	0.0693	0.0584	0.0693	0.0584
$(0, 0.9)$	0.1593	0.1953	0.1280	0.1737	0.0851	0.1550
$(0, 0.6)$	0.1539	0.1863	0.1256	0.1677	0.0858	0.1514
$(0, 0.3)$	0.1451	0.1686	0.1224	0.1538	0.0925	0.1409
$(0, -0.3)$	0.1225	0.1178	0.1225	0.1178	0.1225	0.1178
$(0, -0.6)$	0.0155	0.0156	0.0155	0.0156	0.0155	0.0156
$(0, -0.9)$	0.0001	0.0000	0.0001	0.0000	0.0001	0.0000
$(-0.6, 0.3)$	0.1661	0.1584	0.1661	0.1584	0.1662	0.1584
$(0.3, -0.6)$	0.0831	0.0804	0.0830	0.0804	0.0829	0.0804
$(0.3, 0.3)$	0.1626	0.2061	0.1313	0.1865	0.0888	0.1694
$(0, 0)$	0.2389	0.2276	0.2419	0.2276	0.2509	0.2276
$(0.6, -0.3)$	0.1756	0.2296	0.1451	0.2141	0.1059	0.2007
$(-0.3, 0.6)$	0.1396	0.1573	0.1192	0.1429	0.0919	0.1303

8 Conclusion and Extensions

The size distortion that arises in nonparametrically studentized testing where consistent HAC estimates are used is now well documented. Reductions in this size distortion may be achieved by the use of inconsistent untruncated HAC estimates in the construction of these tests which in turn rely on nonstandard limit distributions for the critical values.

However, these improvements in size typically come at the cost of substantial reductions in test power.

The solution to this problem that is suggested in this paper involves a compromise, whereby untruncated kernels are still employed but the exponent in the power kernel is chosen so as to control size distortion and to maintain power. The criterion function used here is based on asymptotic expansions of the distribution of the test under both the null and alternative hypotheses. The rule for selecting the optimal exponent in the power kernel generally has an expansion rate of $O\left(T^{1/2}\right)$ or less, and is slower than the rate $O\left(T^{2/3}\right)$ for optimizing the asymptotic mean squared error in HAC estimation. Thus, optimal exponent selection (or, in the terminology of conventional HAC estimation, bandwidth selection) to improve test size and power in HAR inference is different from optimal exponent selection for HAC estimation. HAR testing along these lines actually undersmooths the long run variance estimate to reduce bias and allows for greater variance in long run variance estimation as it is manifested in the test statistic by means of higher order adjustments to the nominal asymptotic critical values or by direct use of the nonstandard limit distribution.

The asymptotic expansions of the finite sample distribution of $\hat{\beta}$ could be extended to the regression model of the form: $y_t = \beta + x_t'\gamma + u_t$ where x_t is a strongly exogenous mean zero vector process. In this case, the OLS and GLS estimators of β satisfy $var\left(\sqrt{T}(\hat{\beta}-\beta)-\sqrt{T}(\tilde{\beta}-\beta)\right) = O(1/T)$ and $\sqrt{T}(\hat{\beta}-\beta)-\sqrt{T}(\tilde{\beta}-\beta)$ is independent of $\sqrt{T}(\tilde{\beta}-\beta)$. These properties ensure that $F_T\left(z\right) = E\left\{G_\delta(z^2\varsigma_{\rho T})\right\} + O\left(1/T\right)$, a crucial step in establishing the asymptotic expansions. Replacing u by $u^* = (I - X(X'X)^{-1}X')u$ for $X = (x_1, ..., x_T)'$ in Assumption 2 and using the same proofs, we can easily establish the asymptotic expansions in Section 5 conditioning on X.

The analysis here could be extended to apply to robust tests where other (positive) kernels are used as the mother kernel prior to exponentiation (PSJ$_b$), or where existing HAC estimation procedures are employed with bandwidth proportional to the sample size (KV). While higher order expansions in those cases will be needed to extend the theory, we conjecture that the formulae will end up being very similar to those given here. In particular, we anticipate that the size distortion in testing will depend on the bias in HAC estimation, for which formulae have already been derived for steep origin kernels in PSJ$_b$ and are well known in the spectral analysis literature (e.g. Hannan, 1970) for estimates based on conventional truncated kernels. Expressions for power reductions will be easy to obtain for different mother kernels using the methods of Section 5. These extensions of the present theory will be reported and evaluated elsewhere.

9 Appendix

9.1 Technical Lemmas and Supplements

Lemma 7 *The cumulants of* $\Xi_\rho - \mu_\rho$ *satisfy*

$$|\kappa_m| \leq 2^{m-1}(m-1)!\left(\frac{4}{\rho+1}\right)^{m-1} \tag{56}$$

and the moments $\alpha_m = E\left(\Xi_\rho - \mu_\rho\right)^m$ *satisfy*

$$|\alpha_m| \leq 2^{2m-2}(m-1)!\left(\frac{4}{\rho+1}\right)^{m-1}. \tag{57}$$

Proof of Lemma 7. Note that

$$\begin{aligned}
&\left|\int_0^1 ... \int_0^1 \left(\prod_{j=1}^m k_\rho^*(\tau_j,\tau_{j+1})\right) d\tau_1 \cdots d\tau_m\right| \\
&\leq \int_0^1 ... \int_0^1 \left|k_\rho^*(\tau_1,\tau_2)k_\rho^*(\tau_2,\tau_3)\cdots k_\rho^*(\tau_{m-1},\tau_m)\right| \left|k_\rho^*(\tau_m,\tau_1)\right| d\tau_1 \cdots d\tau_m \\
&\leq \int_0^1 ... \int_0^1 \left|k_\rho^*(\tau_1,\tau_2)k_\rho^*(\tau_2,\tau_3)\cdots k_\rho^*(\tau_{m-1},\tau_m)\right| d\tau_1 \cdots d\tau_m \\
&\leq \sup_{\tau_2}\int_0^1 \left|k_\rho^*(\tau_1,\tau_2)\right| d\tau_1 \int_0^1 \left|k_\rho^*(\tau_2,\tau_3)k_\rho^*(\tau_3,\tau_4)\cdots k_\rho^*(\tau_{m-1},\tau_m)\right| d\tau_2 \cdots d\tau_m \\
&\leq \sup_{\tau_2}\int_0^1 \left|k_\rho^*(\tau_1,\tau_2)\right| d\tau_1 \sup_{\tau_3}\int_0^1 \left|k_\rho^*(\tau_2,\tau_3)\right| d\tau_2 ... \sup_{\tau_m}\int_0^1 \left[k_\rho^*(\tau_{m-1},\tau_m)\right] d\tau_{m-1} \\
&= \left(\sup_s \int_0^1 \left|k_\rho^*(r,s)\right| dr\right)^{m-1}.
\end{aligned} \tag{58}$$

But

$$\begin{aligned}
&\sup_{s\in[0,1]}\int_0^1 \left|k_\rho^*(r,s)\right| dr \\
&= \sup_{s\in[0,1]}\int_0^1 \left|k_\rho(r-s) - \int_0^1 k_\rho(r-p)dp - \int_0^1 k_\rho(s-q)dq + \int_0^1\int_0^1 k_\rho(p-q)dpdq\right| dr \\
&= \sup_{s\in[0,1]}\int_0^1 \left|k_\rho(r-s) - \frac{2-r^{1+\rho}-(1-r)^{1+\rho}}{\rho+1} - \frac{2-s^{1+\rho}-(1-s)^{1+\rho}}{\rho+1} + \frac{2}{\rho+2}\right| dr
\end{aligned}$$

$$\leq \sup_{s\in[0,1]} \left\{ \int_0^1 k_\rho(r-s)dr + \right.$$

$$\left. \int_0^1 \left(\frac{4 - r^{1+\rho} - (1-r)^{1+\rho} - s^{1+\rho} - (1-s)^{1+\rho}}{\rho+1} - \frac{2}{\rho+2} \right) dr \right\}$$

$$= \sup_{s\in[0,1]} 2\frac{2 - s^{1+\rho} - (1-s)^{1+\rho}}{\rho+1} \leq \frac{4}{\rho+1}, \tag{59}$$

using the fact that

$$\int_0^1 \int_0^1 (1 - |r-p|)^\rho \, dpdr = \frac{2}{\rho+2}. \tag{60}$$

Therefore

$$\left| \int_0^1 \ldots \int_0^1 \left(\prod_{j=1}^m k_\rho^*(\tau_j, \tau_{j+1}) \right) d\tau_1 \cdots d\tau_m \right| \leq \left(\frac{4}{\rho+1} \right)^{m-1}, \tag{61}$$

and

$$|\kappa_m| \leq 2^{m-1}(m-1)! \left(\frac{4}{\rho+1} \right)^{m-1}. \tag{62}$$

Note that the moments $\{\alpha_j\}$ and cumulants $\{\kappa_j\}$ satisfy the following recursive relationship:

$$\alpha_1 = \kappa_1, \;\; \alpha_m = \sum_{j=0}^{m-1} \binom{m-1}{j} \alpha_j \kappa_{m-j}. \tag{63}$$

It follows easily by induction from (62), (63) and the identity

$$\sum_{j=0}^{m-1} \binom{m-1}{j} = 2^{m-1}, \tag{64}$$

that

$$|\alpha_m| \leq 2^{2m-2}(m-1)! \left(\frac{4}{\rho+1} \right)^{m-1}. \tag{65}$$

■

Lemma 8 *Let Assumption 2 hold. When $T \to \infty$ with ρ fixed, we have*

(a)

$$\mu_{\rho T} = \mu_\rho + O\left(\frac{1}{T}\right). \tag{66}$$

(b)

$$\kappa_{m,T} = \kappa_m + O\left\{ \frac{m! 2^{m-1}}{T^2} \left(\frac{4}{\rho+1} \right)^{m-2} \right\} \tag{67}$$

uniformly over $m \geq 1$.

(c)

$$\alpha_{m,T} = E\left(\varsigma_{\rho T} - \mu_{\rho T}\right)^m = \alpha_m + O\left\{\frac{2^{2m-1}m!}{T^2}\left(\frac{4}{\rho+1}\right)^{m-2}\right\} \tag{68}$$

uniformly over $m \geq 1$.

Proof of Lemma 8. We first calculate $\mu_{\rho T} = \left(T\omega_T^2\right)^{-1}\text{Trace}(\Omega_T A_T W_\rho A_T)$. Let $W_\rho^* = A_T W_\rho A_T$, then the (i,j)-th element of W_ρ^* is

$$\tilde{k}_\rho\left(\frac{i}{T}, \frac{j}{T}\right) = k_\rho\left(\frac{i-j}{T}\right) - \frac{1}{T}\sum_{m=1}^{T} k_\rho\left(\frac{i-m}{T}\right) - \frac{1}{T}\sum_{k=1}^{T} k_\rho\left(\frac{k-j}{T}\right) + \frac{1}{T^2}\sum_{k=1}^{T}\sum_{m=1}^{T} k_\rho\left(\frac{k-m}{T}\right). \tag{69}$$

So

$$\begin{aligned}
&\text{Trace}\left(\Omega_T A_T W_\rho A_T\right) = \text{Trace}(\Omega_T W_\rho^*) \\
&= \sum_{1\leq r_1, r_2 \leq T}\left\{\gamma(r_1 - r_2)\tilde{k}_\rho\left(\frac{r_1}{T}, \frac{r_2}{T}\right)\right\} = \sum_{r_2=1}^{T}\sum_{h_1=1-r_2}^{T-r_2}\gamma(h_1)\tilde{k}_\rho\left(\frac{r_2+h_1}{T}, \frac{r_2}{T}\right) \\
&= \left(\sum_{h_1=1}^{T-1}\sum_{r_2=1}^{T-h_1} + \sum_{h_1=1-T}^{0}\sum_{r_2=1-h_1}^{T}\right)\gamma(h_1)\tilde{k}_\rho\left(\frac{r_2+h_1}{T}, \frac{r_2}{T}\right).
\end{aligned} \tag{70}$$

But

$$\begin{aligned}
&\sum_{r_2=1}^{T-h_1}\tilde{k}_\rho\left(\frac{r_2+h_1}{T}, \frac{r_2}{T}\right) = \sum_{r_2=1}^{T-h_1} k_\rho\left(\frac{h_1}{T}\right) - \frac{1}{T}\sum_{r_1=1+h_1}^{T}\sum_{m=1}^{T} k_\rho\left(\frac{r_1-m}{T}\right) \\
&- \frac{1}{T}\sum_{r_2=1}^{T-h_1}\sum_{k=1}^{T} k_\rho\left(\frac{k-r_2}{T}\right) + \sum_{r_2=1}^{T-h_1}\frac{1}{T^2}\sum_{k=1}^{T}\sum_{m=1}^{T} k_\rho\left(\frac{k-m}{T}\right) \\
&= -\frac{1}{T}\sum_{r_1=1}^{T}\sum_{m=1}^{T} k_\rho\left(\frac{r_1-m}{T}\right) - \frac{1}{T}\sum_{r_2=1}^{T}\sum_{k=1}^{T} k_\rho\left(\frac{k-r_2}{T}\right) \\
&+ \sum_{r_2=1}^{T}\frac{1}{T^2}\sum_{k=1}^{T}\sum_{m=1}^{T} k_\rho\left(\frac{k-m}{T}\right) + Tk_\rho\left(\frac{h_1}{T}\right) + C(h_1) \\
&= -\frac{1}{T}\sum_{r=1}^{T}\sum_{s=1}^{T} k_\rho\left(\frac{r-s}{T}\right) + Tk_\rho\left(\frac{h_1}{T}\right) + C(h_1) \\
&= \sum_{r_2=1}^{T}\tilde{k}_\rho\left(\frac{r_2}{T}, \frac{r_2}{T}\right) + T\left\{k_\rho\left(\frac{h_1}{T}\right) - k_\rho\left(0\right)\right\} + C(h_1)
\end{aligned} \tag{71}$$

where $C(h_1)$ is a function of h_1 satisfying $|C(h_1)| \leq h_1$. Similarly,

$$\sum_{r_2=1-h_1}^{T} \tilde{k}_\rho\left(\frac{r_2+h_1}{T}, \frac{r_2}{T}\right) = \sum_{r_2=1}^{T} \tilde{k}_\rho\left(\frac{r_2}{T}, \frac{r_2}{T}\right) + T\left\{k_\rho\left(\frac{h_1}{T}\right) - k_\rho(0)\right\} + C(h_1). \quad (72)$$

Therefore, $\text{Trace}(\Omega_T A_T W_\rho A_T)$ is equal to

$$\begin{aligned} &= \sum_{h_1=-T+1}^{T-1} \gamma(h_1) \sum_{r_2=1}^{T} \tilde{k}_\rho\left(\frac{r_2}{T}, \frac{r_2}{T}\right) + T \sum_{h_1=-T+1}^{T-1} \gamma(h_1)\left\{k_\rho\left(\frac{h_1}{T}\right) - k_\rho(0)\right\} + O(1) \\ &= \sum_{h_1=-T+1}^{T-1} \gamma(h_1) \sum_{r_2=1}^{T} \tilde{k}_\rho\left(\frac{r_2}{T}, \frac{r_2}{T}\right) - \rho \sum_{h_1=-T+1}^{T-1} |h_1|\, \gamma(h_1) + O\left(\frac{\rho^2}{T}\right) + O(1), \quad (73) \end{aligned}$$

where we have used the second order Taylor expansion:

$$k_\rho\left(\frac{|h_1|}{T}\right) - k_\rho(0) = -\rho\,|h_1|\,/T + 0.5\rho(\rho-1)\,(\widetilde{x})^{\rho-2}\, h^2/T^2, \quad (74)$$

for some $\widetilde{x}$ between 0 and $|h_1|\,/T$. Using

$$\sum_{h_1=-T+1}^{T-1} \gamma(h_1) = \omega_T^2(1 + O(\frac{1}{T})) \quad (75)$$

and

$$\frac{1}{T}\sum_{r_2=1}^{T} \tilde{k}_\rho\left(\frac{r_2}{T}, \frac{r_2}{T}\right) = \int_0^1 k_\rho^*(r,r)dr + O(\frac{1}{T}), \quad (76)$$

we now have

$$\mu_{\rho T} = \int_0^1 k_\rho^*(r,r)dr - \frac{\rho}{T\omega_T^2} \sum_{h_1=-T+1}^{T-1} |h_1|\,\gamma(h_1) + O\left(\frac{\rho^2}{T^2} + \frac{1}{T}\right). \quad (77)$$

By definition, $\mu_\rho = E\Xi_\rho = \int_0^1 k^*(r,r)dr$ and thus $\mu_{\rho T} = \mu_\rho + O(1/T)$ as desired.

We next approximate $Trace\left[(\Omega_T A_T W_\rho A_T)^m\right]$ for $m > 1$. The approach is similar to the case $m = 1$ but notationally more complicated. Let $r_{2m+1} = r_1$, $r_{2m+2} = r_2$, and $h_{m+1} = h_1$. Then

$$\begin{aligned} &Trace\left[(\Omega_T A_T W_\rho A_T)^m\right] \\ &= \sum_{r_2, r_4, \ldots, r_{2m}=1}^{T} \prod_{j=1}^{m} \gamma(r_{2j-1} - r_{2j}) \tilde{k}_\rho\left(\frac{r_{2j}}{T}, \frac{r_{2j+1}}{T}\right) \end{aligned}$$

$$= \sum_{r_2,r_4,\ldots,r_{2m}=1}^{T} \sum_{h_1=1-r_2}^{T-r_2} \sum_{h_2=1-r_4}^{T-r_4} .. \sum_{h_m=1-r_{2m}}^{T-r_{2m}} \prod_{j=1}^{m} \gamma(h_j)\tilde{k}_\rho\left(\frac{r_{2j}}{T}, \frac{r_{2j+2}+h_{j+1}}{T}\right)$$

$$= \left(\sum_{h_1=1}^{T-1}\sum_{r_2=1}^{T-h_1} + \sum_{h_1=1-T}^{0}\sum_{r_2=1-h_1}^{T}\right)\cdots\left(\sum_{h_m=1}^{T-1}\sum_{r_{2m}=1}^{T-h_m} + \sum_{h_m=1-T}^{0}\sum_{r_{2m}=1-h_m}^{T}\right)$$

$$\prod_{j=1}^{m} \gamma(h_j)\tilde{k}_\rho\left(\frac{r_{2j}}{T}, \frac{r_{2j+2}+h_{j+1}}{T}\right)$$

$$= I + II, \tag{78}$$

where

$$I = \left(\sum_{h_1=1}^{T-1}\sum_{r_2=1}^{T-h_1} + \sum_{h_1=1-T}^{0}\sum_{r_2=1-h_1}^{T}\right)\cdots\left(\sum_{h_m=1}^{T-1}\sum_{r_{2m}=1}^{T-h_m} + \sum_{h_m=1-T}^{0}\sum_{r_{2m}=1-h_m}^{T}\right)$$

$$\prod_{j=1}^{m} \gamma(h_j)\tilde{k}_\rho\left(\frac{r_{2j}}{T}, \frac{r_{2j+2}}{T}\right) \tag{79}$$

and

$$II = O\left\{\left(\sum_{h_1=1}^{T-1}\sum_{r_2=1}^{T-h_1} + \sum_{h_1=1-T}^{0}\sum_{r_2=1-h_1}^{T}\right)\cdots\left(\sum_{h_m=1}^{T-1}\sum_{r_{2m}=1}^{T-h_m} + \sum_{h_m=1-T}^{0}\sum_{r_{2m}=1-h_m}^{T}\right)\right.$$

$$\left.\prod_{j=1}^{m} |\gamma(h_j)|\left(\frac{\rho|h_{j+1}|}{T}\right)\right\}. \tag{80}$$

Here we have used

$$\left|\tilde{k}_\rho\left(\frac{r_{2j}}{T}, \frac{r_{2j+2}+h_{j+1}}{T}\right) - \tilde{k}_\rho\left(\frac{r_{2j}}{T}, \frac{r_{2j+2}}{T}\right)\right| = O\left(\frac{\rho|h_{j+1}|}{T}\right). \tag{81}$$

A similar result is given and proved in (95) below.

The first term (I) can be written as

$$I = \left(\sum_{h_1=1-T}^{T-1}\sum_{r_2=1}^{T} - \sum_{h_1=1}^{T-1}\sum_{r_2=T-h_1+1}^{T} - \sum_{h_1=1-T}^{0}\sum_{r_2=1}^{-h_1}\right)\cdots$$

$$\left(\sum_{h_m=1-T}^{T-1}\sum_{r_{2m}=1}^{T} - \sum_{h_m=1}^{T-1}\sum_{r_{2m}=T-h_m+1}^{T} - \sum_{h_m=1-T}^{0}\sum_{r_{2m}=1}^{-h_m}\right)\prod_{j=1}^{m}\gamma(h_j)\left\{\tilde{k}_\rho\left(\frac{r_{2j}}{T}, \frac{r_{2j+2}}{T}\right)\right\}$$

$$= \sum_{\pi}\sum_{h_1,r_2}\cdots\sum_{h_m,r_{2m}}\prod_{j=1}^{m}\gamma(h_j)\left\{\tilde{k}_\rho\left(\frac{r_{2j}}{T}, \frac{r_{2j+2}}{T}\right)\right\}, \tag{82}$$

where $\sum_{h_j,r_{2j}}$ is one of the three choices $\sum_{h_j=1-T}^{T-1}\sum_{r_{2j}=1}^{T}$, $-\sum_{h_j=1}^{T-1}\sum_{r_{2j}=T-h_j+1}^{T}$, $-\sum_{h_j=1-T}^{0}\sum_{r_{2j}=1}^{-h_j}$ and $\sum_\pi$ is the summation over all possible combinations of $\left(\sum_{h_1,r_2}\cdots\sum_{h_m,r_{2m}}\right)$. The 3^m summands in (82) can be divided into two groups with the first group consisting of the summands all of whose r indices run from 1 to T and the second group consisting of the rest. It is obvious that the first group can be written as

$$\left(\sum_h\prod_{j=1}^m\gamma(h_j)\right)\sum_r\left\{\tilde{k}_\rho\left(\frac{r_{2j}}{T},\frac{r_{2j+2}}{T}\right)\right\}..$$

The dominating terms in the second group are of the forms

$$\sum_{h_j=1-T}^{T-1}\sum_{r_2=1}^{T}\cdots\sum_{h_k=1-T}^{T-1}\sum_{r_{2k}=T-h_k+1}^{T}\cdots\sum_{h_m=1-T}^{T-1}\sum_{r_{2m}=1}^{T}\prod_{j=1}^m\gamma(h_j)\left\{\tilde{k}_\rho\left(\frac{r_{2j}}{T},\frac{r_{2j+2}}{T}\right)\right\}$$

or

$$\sum_{h_j=1-T}^{T-1}\sum_{r_2=1}^{T}\cdots\sum_{h_k=1-T}^{T-1}\sum_{r_{2k}=T-h_k+1}^{T}\cdots\sum_{h_m=1-T}^{T-1}\sum_{r_{2m}=1}^{-h_m}\prod_{j=1}^m\gamma(h_j)\left\{\tilde{k}_\rho\left(\frac{r_{2j}}{T},\frac{r_{2j+2}}{T}\right)\right\},$$

both of which are bounded by

$$\begin{aligned}&\sum_{h_1=1-T}^{T-1}\sum_{r_2=1}^{T}\cdots\sum_{h_k=1-T}^{T-1}\cdots\sum_{h_m=1-T}^{T-1}\sum_{r_{2m}=1}^{-h_m}\prod_{j=1}^m|\gamma(h_j)|\,|h_k|\prod_{j\neq k,m}\left|\tilde{k}_\rho\left(\frac{r_{2j}}{T},\frac{r_{2j+2}}{T}\right)\right|\\ &\leq\left[\sup_{r_4}\sum\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_4}{T}\right)\right]^{m-2}\left(\sum_{h_j}|\gamma(h_j)|\right)^{m-1}\left(\sum_{h_k}|\gamma(h_k)|\,|h_k|\right),\end{aligned}$$

using the same approach as in (58). Approximating the sum by an integral and noting that the second group contains $(m-1)$ terms which are of the same orders of magnitude as the above typical dominating terms, we conclude that the second group is of order $O\left[mT^{m-2}\left(4/(\rho+1)\right)^{m-2}\right]$ uniformly over m. As a consequence,

$$I=\left(\sum_h\prod_{j=1}^m\gamma(h_j)\right)\sum_r\left\{\tilde{k}_\rho\left(\frac{r_{2j}}{T},\frac{r_{2j+2}}{T}\right)\right\}+O\left\{mT^{m-2}\left(\frac{4}{\rho+1}\right)^{m-2}\right\},\qquad(83)$$

uniformly over m.

The second term (II) is easily shown to be of order $o\left(mT^{m-2}\left(1/\rho\right)^{m-2}\right)$ uniformly over m. Therefore

$$\begin{aligned}&Trace\left[(\Omega_TA_TW_\rho A_T)^m\right]\\ &=\left(\sum_h\gamma(h)\right)^m\sum_r\left\{\tilde{k}_\rho\left(\frac{r_{2j}}{T},\frac{r_{2j+2}}{T}\right)\right\}+O\left\{mT^{m-2}\left(\frac{4}{\rho+1}\right)^{m-2}\right\}\end{aligned}\qquad(84)$$

and

$$\begin{aligned}
\kappa_{m,T} &= 2^{m-1}(m-1)!T^{-m}\left(\omega_T^2\right)^{-m}\text{Trace}\left[(\Omega_T A_T W_\rho A_T)^m\right] \\
&= 2^{m-1}(m-1)!\left\{T^{-m}\sum_r \tilde{k}_\rho\left(\frac{r_{2j}}{T},\frac{r_{2j+2}}{T}\right)+O\left[\frac{m}{T^2}\left(\frac{4}{\rho+1}\right)^{m-2}\right]\right\} \\
&= 2^{m-1}(m-1)!\left\{\int\prod_{j=1}^m\int_0^1 k_\rho^*(\tau_j,\tau_{j+1})d\tau_j d\tau_{j+1}+O\left[\frac{m}{T^2}\left(\frac{4}{\rho+1}\right)^{m-2}\right]\right\} \\
&= \kappa_m+O\left\{\frac{m!2^{m-1}}{T^2}\left(\frac{4}{\rho+1}\right)^{m-2}\right\}, \qquad (85)
\end{aligned}$$

uniformly over m.

Finally, we consider $\alpha_{m,T}$. Note that $\alpha_{1,T}=E(\varsigma_{\rho T}-\mu_{\rho T})=0$ and

$$\alpha_{m,T}=\sum_{j=0}^{m-1}\binom{m-1}{j}\alpha_{j,T}\kappa_{m-j,T}. \qquad (86)$$

It follows that

$$\alpha_{m,T}=\sum_\pi\frac{m!}{(j_1!)^{m_1}(j_2!)^{m_2}\cdots(j_k!)^{m_k}}\frac{1}{m_1!m_2!\cdots m_k!}\prod_{j\in\pi}\kappa_{j,T}, \qquad (87)$$

where the sum is taken over the elements

$$\pi=[\underbrace{j_1,\cdots j_1}_{m_1\text{ times}},\underbrace{j_2,\cdots j_2}_{m_2\text{ times}},\cdots\underbrace{j_k,\cdots j_k}_{m_k\text{ times}}] \qquad (88)$$

for some integer k, sequence $\{j_k\}$ such that $j_1>j_2>\cdots>j_k$ and $m=\sum_{i=1}^k m_i j_i$.

Combining the preceding formula with part (b) gives

$$\begin{aligned}
\alpha_{m,T} &= \alpha_m+O\left\{\frac{2^{m-1}}{T^2}\left(\frac{4}{\rho+1}\right)^{m-2}\sum_\pi\frac{m!}{m_1!m_2!\cdots m_k!}\right\} \\
&= \alpha_m+O\left\{\frac{2^{2m-1}m!}{T^2}\left(\frac{4}{\rho+1}\right)^{m-2}\right\}, \qquad (89)
\end{aligned}$$

uniformly over m, where the last line follows because $\sum_\pi\frac{1}{m_1!m_2!\cdots m_k!}<2^m$. ■

Lemma 9 *Let Assumption 2 hold. If $\rho\to\infty$ and $T\to\infty$ such that $\rho/T\to0$, then*

(a)

$$\mu_{\rho T}=\mu_\rho-\frac{\rho}{T\omega_T^2}\sum_{h=1-T}^{T-1}h\Gamma(h)+O\left(\frac{1}{T}\right)+O\left(\frac{\rho^2}{T^2}\right); \qquad (90)$$

(b)

$$\kappa_{2,T} = \frac{2}{\rho+1} + O\left(\frac{1}{T}\right); \tag{91}$$

(c)

$$\kappa_{3,T} = O\left(\frac{1}{\rho^2}\right) + O\left(\frac{1}{T}\right). \tag{92}$$

Proof of Lemma 9. We have proved (90) in the proof of Lemma 8 as equation (77) holds for both fixed ρ and increasing ρ. It remains to consider $\kappa_{m,T}$ for $m = 2$ and 3. We first consider $\kappa_{2,T} = 2T^{-2}\left(\omega_T^{-4}\right)\text{Trace}\left[(\Omega_T A_T W_\rho A_T)^2\right]$. As a first step, we have

$$\begin{aligned}
&\text{Trace}\left[(\Omega_T A_T W_\rho A_T)^2\right] \\
&= \sum_{r_1,r_2,r_3,r_4} \left\{\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_3}{T}\right)\tilde{k}_\rho\left(\frac{r_4}{T},\frac{r_1}{T}\right)\right\}\gamma(r_1-r_2)\gamma(r_3-r_4) \\
&= \sum_{r_2=1}^{T}\sum_{h_1=1-r_2}^{T-r_2}\sum_{r_4=1}^{T}\sum_{h_2=1-r_4}^{T-r_4}\left\{\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_4+h_2}{T}\right)\tilde{k}_\rho\left(\frac{r_4}{T},\frac{r_2+h_1}{T}\right)\right\}\gamma(h_1)\gamma(h_2) \\
&= \left(\sum_{h_1=1}^{T-1}\sum_{r_2=1}^{T-h_1} + \sum_{h_1=1-T}^{0}\sum_{r_2=1-h_1}^{T}\right)\left(\sum_{h_2=1}^{T-1}\sum_{r_4=1}^{T-h_2} + \sum_{h_2=1-T}^{0}\sum_{r_4=1-h_2}^{T}\right) \\
&\left\{\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_4+h_2}{T}\right)\tilde{k}_\rho\left(\frac{r_4}{T},\frac{r_2+h_1}{T}\right)\right\}\gamma(h_1)\gamma(h_2) \\
&:= I_1 + I_2 + I_3 + I_4,
\end{aligned} \tag{93}$$

where

$$I_1 = \sum_{h_1=1}^{T-1}\sum_{h_2=1}^{T-1}\sum_{r_2=1}^{T-h_1}\sum_{r_4=1}^{T-h_2}\left\{\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_4+h_2}{T}\right)\tilde{k}_\rho\left(\frac{r_4}{T},\frac{r_2+h_1}{T}\right)\right\}\gamma(h_1)\gamma(h_2),$$

$$I_2 = \sum_{h_1=1}^{T-1}\sum_{r_2=1}^{T-h_1}\sum_{h_2=1-T}^{0}\sum_{r_4=1-h_2}^{T}\left\{\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_4+h_2}{T}\right)\tilde{k}_\rho\left(\frac{r_4}{T},\frac{r_2+h_1}{T}\right)\right\}\gamma(h_1)\gamma(h_2),$$

$$I_3 = \sum_{h_1=1-T}^{0}\sum_{r_2=1-h_1}^{T}\sum_{r_2=1}^{T-h_1}\sum_{r_4=1}^{T-h_2}\left\{\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_4+h_2}{T}\right)\tilde{k}_\rho\left(\frac{r_4}{T},\frac{r_2+h_1}{T}\right)\right\}\gamma(h_1)\gamma(h_2),$$

and

$$I_4 = \sum_{h_1=1-T}^{0}\sum_{r_2=1-h_1}^{T}\sum_{h_2=1-T}^{0}\sum_{r_4=1-h_2}^{T}\left\{\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_4+h_2}{T}\right)\tilde{k}_\rho\left(\frac{r_4}{T},\frac{r_2+h_1}{T}\right)\right\}\gamma(h_1)\gamma(h_2).$$

We now consider each term in turn. Note that

$$\frac{1}{T}\sum_{k=1}^{T} k_\rho\left(\frac{k-r_4-h_2}{T}\right)=\frac{1}{T}\sum_{k=1-h_2}^{T-h_2} k_\rho\left(\frac{k-r_4}{T}\right)=\frac{1}{T}\sum_{k=1}^{T} k_\rho\left(\frac{k-r_4}{T}\right)+O\left(\frac{|h_2|}{T}\right),$$

and

$$\begin{aligned}&\left|k_\rho\left(\frac{r_2-r_4-h_2}{T}\right)-k_\rho\left(\frac{r_2-r_4}{T}\right)\right|\\&=\rho\left(1-\tilde{x}\right)^{\rho-1}\left|\frac{|r_2-r_4-h_2|}{T}-\frac{|r_2-r_4|}{T}\right|=O\left(\frac{\rho\,|h_2|}{T}\right),\end{aligned}\tag{94}$$

for some $\tilde{x}$ between $|r_2-r_4-h_2|/T$ and $|r_2-r_4|/T$, we have

$$\begin{aligned}&\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_4+h_2}{T}\right)\\&=\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_4}{T}\right)+k_\rho\left(\frac{r_2-r_4-h_2}{T}\right)-k_\rho\left(\frac{r_2-r_4}{T}\right)+O(\frac{|h_2|}{T})\\&=\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_4}{T}\right)+O\left(\frac{\rho\,|h_2|}{T}\right).\end{aligned}\tag{95}$$

Similarly

$$\tilde{k}_\rho\left(\frac{r_4}{T},\frac{r_2+h_1}{T}\right)=\tilde{k}_\rho\left(\frac{r_4}{T},\frac{r_2}{T}\right)+O\left(\frac{\rho\,|h_1|}{T}\right).\tag{96}$$

It follows from (95) and (96) that

$$\begin{aligned}I_1=&\sum_{h_1=1}^{T-1}\sum_{h_2=1}^{T-1}\sum_{r_2=1}^{T-1}\sum_{r_4=1}^{T-1}\left\{\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_4+h_2}{T}\right)\tilde{k}_\rho\left(\frac{r_4}{T},\frac{r_2+h_1}{T}\right)\right\}\gamma(h_1)\gamma(h_2)\\&+O\left(\sum_{h_1=1}^{T-1}\sum_{h_2=1}^{T-1}\left[T\left(|h_1|+|h_2|\right)+|h_1h_2|\right]|\gamma(h_1)\gamma(h_2)|\right)\\=&\sum_{h_1=1}^{T-1}\sum_{h_2=1}^{T-1}\sum_{r_2=1}^{T-1}\sum_{r_4=1}^{T-1}\left\{\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_4+h_2}{T}\right)\tilde{k}_\rho\left(\frac{r_4}{T},\frac{r_2+h_1}{T}\right)\right\}\gamma(h_1)\gamma(h_2)+O(T)\\=&\sum_{h_1=1}^{T-1}\sum_{h_2=1}^{T-1}\sum_{r_2=1}^{T-1}\sum_{r_4=1}^{T-1}\left\{\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_4}{T}\right)\tilde{k}_\rho\left(\frac{r_4}{T},\frac{r_2}{T}\right)\right\}\gamma(h_1)\gamma(h_2)+O(T)\\&+O\left\{\sum_{h_1=1}^{T-1}\sum_{h_2=1}^{T-1}\sum_{r_2=1}^{T-1}\sum_{r_4=1}^{T-1}\left|\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_4}{T}\right)\right|\left(\frac{\rho\left(|h_1|+|h_2|\right)}{T}\right)|\gamma(h_1)\gamma(h_2)|\right\}\\=&\sum_{h_1=1}^{T-1}\sum_{h_2=1}^{T-1}\sum_{r_2=1}^{T-1}\sum_{r_4=1}^{T-1}\left\{\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_4}{T}\right)\tilde{k}_\rho\left(\frac{r_4}{T},\frac{r_2}{T}\right)\right\}\gamma(h_1)\gamma(h_2)+O\left(T\right).\end{aligned}$$

Following the same procedure, we can show that

$$I_2 = \sum_{h_1=1}^{T-1} \sum_{r_2=1}^{T} \sum_{h_2=1-T}^{0} \sum_{r_4=1}^{T} \left\{ \tilde{k}_\rho \left(\frac{r_2}{T}, \frac{r_4}{T} \right) \tilde{k}_\rho \left(\frac{r_4}{T}, \frac{r_2}{T} \right) \right\} \gamma(h_1)\gamma(h_2) + O\left(T\right),$$

$$I_3 = \sum_{h_1=1-T}^{0} \sum_{r_2=1}^{T} \sum_{r_2=1}^{T} \sum_{r_4=1}^{T} \left\{ \tilde{k}_\rho \left(\frac{r_2}{T}, \frac{r_4}{T} \right) \tilde{k}_\rho \left(\frac{r_4}{T}, \frac{r_2}{T} \right) \right\} \gamma(h_1)\gamma(h_2) + O\left(T\right),$$

and

$$I_4 = \sum_{h_1=1-T}^{0} \sum_{r_2=1}^{T} \sum_{h_2=1-T}^{0} \sum_{r_4=1}^{T} \left\{ \tilde{k}_\rho \left(\frac{r_2}{T}, \frac{r_4}{T} \right) \tilde{k}_\rho \left(\frac{r_4}{T}, \frac{r_2}{T} \right) \right\} \gamma(h_1)\gamma(h_2) + O\left(T\right).$$

As a consequence,

$$\begin{aligned} &\text{Trace}\left[\left(\Omega_T A_T W_\rho A_T \right)^2 \right] \\ &= \sum_{r_2, r_4} \left\{ \tilde{k}_\rho \left(\frac{r_2}{T}, \frac{r_4}{T} \right) \right\}^2 \left(\sum_{h=1-T}^{T-1} \gamma(h_1) \right)^2 + O\left(T\right), \end{aligned} \tag{97}$$

and

$$\begin{aligned} \kappa_{2,T} &= 2T^{-2} \left(\omega_T^{-4} \right) Trace \left[\left(\Omega_T A_T W_\rho A_T \right)^2 \right] \\ &= 2 \int_0^1 \left(k_\rho^*(r,s) \right)^2 drds + O\left(\frac{1}{T} \right) = \frac{2}{\rho+1} + O\left(\frac{1}{T} \right). \end{aligned} \tag{98}$$

The proof for $\kappa_{3,T}$ is essentially the same except that we use Lemma 7 to obtain the first term $O\left(1/\rho^2\right)$. Details are omitted. ■

9.2 Proofs of the Main Results

Proof of Theorem 1. Using the formula from (25), we obtain

$$\begin{aligned} &E\left(\Xi_\rho - \mu_\rho\right)^2 = \kappa_2 = 2 \int_0^1 \int_0^1 \left(k_\rho^*(r,t) \right)^2 dtdr \\ &= 2 \left[\left(\frac{2}{\rho+2} \right)^2 + \frac{1}{\rho+1} - \frac{4}{(\rho+1)^2} \left(\frac{4\rho^2 + 7\rho + 2}{(2\rho+3)(\rho+2)} + \frac{\Gamma^2(\rho+2)}{\Gamma(2\rho+4)} \right) \right]. \end{aligned} \tag{99}$$

Following Lemma 9, we have

$$E\left(\Xi_\rho - \mu_\rho\right)^3 = \kappa_3 = O\left(1/\rho^2\right). \tag{100}$$

Combining (20), (99), (100) and $\mu_\rho = \rho/(\rho+2)$ yields

$$\begin{aligned}
F(z) &= P\left\{\left|W(1)\Xi_\rho^{-1/2}\right| < z\right\} \\
&= D(\mu_\rho z^2) + \frac{1}{2}D''(\mu_\rho z^2)z^4\kappa_2 + O\left(1/\rho^2\right) \\
&= D(\mu_\rho z^2) + D''(\mu_\rho z^2)z^4\frac{1}{\rho+1} + O\left(1/\rho^2\right) \\
&= D(z^2) - D'(z^2)z^2\frac{2}{\rho+2} + D''(z^2)z^4\frac{1}{\rho+1} + O\left(1/\rho^2\right) \\
&= D(z^2) + \left[D''(z^2)z^4 - 2D'(z^2)z^2\right]\frac{1}{\rho} + O\left(1/\rho^2\right), \qquad (101)
\end{aligned}$$

where the $O(\cdot)$ term holds uniformly for any $z \in [M_l, M_u]$ where $0 < M_l < M_u < \infty$. ■

Proof of Corollary 2. Using a power series expansion, we have

$$\begin{aligned}
F(z_{\alpha,\rho}) &= D(z_{\alpha,\rho}^2) + \left[D''(z_{\alpha,\rho}^2)z_{\alpha,\rho}^4 - 2D'(z_{\alpha,\rho}^2)z_{\alpha,\rho}^2\right]\frac{1}{\rho} + O\left(1/\rho^2\right) \\
&= D(z_\alpha^2) + \left[D''(z_\alpha^2)z_\alpha^4 - 2D'(z_\alpha^2)z_\alpha^2\right]\frac{1}{\rho} + D'(z_\alpha^2)\left(z_{\alpha,\rho}^2 - z_\alpha^2\right) \\
&+ \frac{1}{\rho}\left[4z_\alpha^3 D''(z_\alpha^2) + D'''(z_\alpha^2)z_\alpha^4 - 2D''(z_\alpha^2)z_\alpha^2 - 2D''(z_\alpha^2)\right] \\
&\times \left(z_{\alpha,\rho}^2 - z_\alpha^2\right) + O\left(1/\rho^2\right), \qquad (102)
\end{aligned}$$

i.e.

$$\begin{aligned}
0 &= \left[D''(z_\alpha^2)z_\alpha^4 - 2D'(z_\alpha^2)z_\alpha^2\right]\frac{1}{\rho} + D'(z_\alpha^2)\left(z_{\alpha,\rho}^2 - z_\alpha^2\right) \\
&+ \frac{1}{\rho}\left[4z_\alpha^3 D''(z_\alpha^2) + D'''(z_\alpha^2)z_\alpha^4 - 2D''(z_\alpha^2)z_\alpha^2 - 2D''(z_\alpha^2)\right] \\
&\times \left(z_{\alpha,\rho}^2 - z_\alpha^2\right) + O\left(1/\rho^2\right). \qquad (103)
\end{aligned}$$

So

$$z_{\alpha,\rho}^2 = z_\alpha^2 - \frac{1}{\rho}\frac{\left[D''(z_\alpha^2)z_\alpha^4 - 2D'(z_\alpha^2)z_\alpha^2\right]}{D'(z_\alpha^2)} + O\left(1/\rho^2\right). \qquad (104)$$

Now

$$D'(z) = \frac{z^{-1/2}e^{-z/2}}{\Gamma(1/2)\sqrt{2}},\ D''(z) = \frac{1}{4\sqrt{\pi}z^2}\left(-\sqrt{2z}e^{-\frac{1}{2}z} - z^{\frac{3}{2}}\sqrt{2}e^{-\frac{1}{2}z}\right), \qquad (105)$$

and thus

$$\frac{D''(z)}{D'(z)} = \frac{\frac{1}{4\sqrt{\pi}z^2}\left(-\sqrt{2z}e^{-\frac{1}{2}z} - z^{\frac{3}{2}}\sqrt{2}e^{-\frac{1}{2}z}\right)}{\frac{z^{-1/2}e^{-z/2}}{\Gamma(1/2)\sqrt{2}}} = \frac{1}{4z^{\frac{3}{2}}}\left(-2\sqrt{z} - 2z^{\frac{3}{2}}\right). \qquad (106)$$

Hence

$$z_{\alpha,\rho}^2 = z_\alpha^2 - \frac{1}{\rho}\left(z_\alpha^4 \frac{1}{4z_\alpha^3}\left(-2z_\alpha - 2z_\alpha^3\right)\right) + \frac{2}{\rho} z_\alpha^2 + O\left(1/\rho^2\right)$$
$$= z_\alpha^2 + \frac{1}{2\rho}(5z_\alpha^2 + z_\alpha^4) + O\left(1/\rho^2\right), \tag{107}$$

from which we get

$$z_{\alpha,\rho} = \left(z_\alpha^2 + \frac{1}{2\rho}(5z_\alpha^2 + z_\alpha^4) + O\left(1/\rho^2\right)\right)^{1/2}$$
$$= z_\alpha\left(1 + \frac{1}{2\rho}(5 + z_\alpha^2) + O\left(1/\rho^2\right)\right)^{1/2}$$
$$= z_\alpha\left(1 + \frac{1}{4\rho}(5 + z_\alpha^2)\right) + O\left(1/\rho^2\right)$$
$$= z_\alpha + \frac{1}{4\rho}\left(5z_\alpha + z_\alpha^3\right) + O\left(1/\rho^2\right), \tag{108}$$

as stated. ■

Proof of Theorem 3. A Taylor series expansion gives

$$1 - EG_\delta(z_{\alpha,\rho}^2 \Xi_\rho)$$
$$= 1 - G_\delta(\mu_\rho z_{\alpha,\rho}^2) + \frac{1}{2} G_\delta''(\mu_\rho z_{\alpha,\rho}^2) z_{\alpha,\rho}^4 E\left(\Xi_\rho - \mu_\rho\right)^2 + O\left(E(\Xi_\rho - \mu_\rho)^3\right)$$
$$= 1 - G_\delta(z_{\alpha,\rho}^2) - \left[G_\delta''(z_{\alpha,\rho}^2) z_{\alpha,\rho}^4 - 2G_\delta'(z_{\alpha,\rho}^2) z_{\alpha,\rho}^2\right]\frac{1}{\rho} + O\left(1/\rho^2\right), \tag{109}$$

as $\rho \to \infty$, where the last equality follows from the same proof as Theorem 1. Since $z_{\alpha,\rho}^2 = z_\alpha^2 + \frac{1}{2\rho}(5z_\alpha^2 + z_\alpha^4) + O\left(1/\rho^2\right)$, we get

$$1 - EG_\delta(z_{\alpha,\rho}^2 \Xi_\rho) = 1 - G_\delta(z_\alpha^2 + \frac{1}{2\rho}(5z_\alpha^2 + z_\alpha^4))$$
$$- \frac{1}{\rho}\left\{G_\delta''(z_\alpha^2 + \frac{1}{2\rho}(5z_\alpha^2 + z_\alpha^4))\left(z_\alpha^2 + \frac{1}{2\rho}(5z_\alpha^2 + z_\alpha^4)\right)^2\right\}$$
$$- \frac{1}{\rho}\left\{-2G_\delta'(z_\alpha^2 + \frac{1}{2\rho}(5z_\alpha^2 + z_\alpha^4))\left(z_\alpha^2 + \frac{1}{2\rho}(5z_\alpha^2 + z_\alpha^4)\right)\right\} + O\left(1/\rho^2\right)$$
$$= 1 - G_\delta(z_\alpha^2) - G_\delta'(z_\alpha^2)\frac{1}{2\rho}(5z_\alpha^2 + z_\alpha^4) - \left[G_\delta''(z_\alpha^2) z_\alpha^4 - 2G_\delta'(z_\alpha^2) z_\alpha^2\right]\frac{1}{\rho} + O\left(1/\rho^2\right)$$
$$= 1 - G_\delta(z_\alpha^2) - \left(\frac{1}{2} G_\delta'(z_\alpha^2)(5z_\alpha^2 + z_\alpha^4) + G_\delta''(z_\alpha^2) z_\alpha^4 - 2G_\delta'(z_\alpha^2) z_\alpha^2\right)\frac{1}{\rho} + O\left(1/\rho^2\right)$$
$$= 1 - G_\delta(z_\alpha^2) - \left(\frac{1}{2} G_\delta'(z_\alpha^2) z_\alpha^4 + G_\delta''(z_\alpha^2) z_\alpha^4 + \frac{1}{2} G_\delta'(z_\alpha^2) z_\alpha^2\right)\frac{1}{\rho} + O\left(1/\rho^2\right). \tag{110}$$

Note that

$$G'_\delta(z) = \sum_{j=0}^{\infty} \frac{\left(\delta^2/2\right)^j}{j!} e^{-\delta^2/2} \frac{z^{j-\frac{1}{2}} e^{-z/2}}{\Gamma(j+1/2)2^{j+1/2}} \tag{111}$$

and

$$\begin{aligned} G''_\delta(z) &= \sum_{j=0}^{\infty} \frac{\left(\delta^2/2\right)^j}{j!} e^{-\delta^2/2} \left((j-\frac{1}{2})\frac{1}{z} \frac{z^{j-\frac{1}{2}} e^{-z/2}}{\Gamma(j+1/2)2^{j+1/2}} - \frac{1}{2} \frac{z^{j-\frac{1}{2}} e^{-z/2}}{\Gamma(j+1/2)2^{j+1/2}} \right) \\ &= \left(-\frac{1}{2z} - \frac{1}{2}\right) G'_\delta(z) + \sum_{j=0}^{\infty} \frac{\left(\delta^2/2\right)^j}{j!} e^{-\delta^2/2} \frac{z^{j-1/2} e^{-z/2}}{\Gamma(j+1/2)2^{j+1/2}} \frac{j}{z} \\ &= -\frac{1}{2} G'_\delta(z) \left(\frac{1}{z} + 1\right) + K_\delta(z), \end{aligned} \tag{112}$$

so

$$\frac{1}{2} G'_\delta(z_\alpha^2) z_\alpha^4 + G''_\delta(z_\alpha^2) z_\alpha^4 + \frac{1}{2} G'_\delta(z_\alpha^2) z_\alpha^2 = z_\alpha^4 K_\delta\left(z_\alpha^2\right), \tag{113}$$

and

$$1 - EG_\delta(z_{\alpha,\rho}^2 \Xi_\rho) = 1 - G_\delta(z_\alpha^2) - z_\alpha^4 K_\delta\left(z_\alpha^2\right)/\rho + O\left(1/\rho^2\right), \tag{114}$$

completing the proof of the theorem. ■

Proof of Theorem 4. First, since $G_\delta(\cdot)$ is a bounded function, we can rewrite (20) as

$$\begin{aligned} & P\left\{ \left| (W(1)+\delta)\, \Xi_\rho^{-1/2} \right| \le z \right\} = \lim_{B\to\infty} EG_\delta(z^2 \Xi_\rho) 1\left\{ |\Xi_\rho - \mu_\rho| \le B \right\} \\ &= \lim_{B\to\infty} E \sum_{m=1}^{\infty} \frac{1}{m!} G_\delta^{(m)}(\mu_\rho z^2)\left(\Xi_\rho - \mu_\rho\right)^m z^{2m} 1\left\{ |\Xi_\rho - \mu_\rho| \le B \right\} \\ &= \lim_{B\to\infty} \sum_{m=1}^{\infty} \frac{1}{m!} G_\delta^{(m)}(\mu_\rho z^2) \alpha_m z^{2m} 1\left\{ |\Xi_\rho - \mu_\rho| \le B \right\}, \end{aligned} \tag{115}$$

where the last line follows because the infinite sum $\sum_{m=1}^{\infty} \frac{1}{m!} G_\delta^{(m)}(\mu_\rho z^2)\alpha_m z^{2m}$ converges uniformly to $G_\delta(z^2 \Xi_\rho)$ when $|\Xi_\rho - \mu_\rho| \le B$. The uniformity holds because $G_\delta(\cdot)$ is infinitely differentiable with bounded derivatives. Using Lemma 7, we have, for some constant $C > 0$

$$\begin{aligned} & \left| \sum_{m=1}^{\infty} \frac{1}{m!} G_\delta^{(m)}(\mu_\rho z^2) \alpha_m z^{2m} \right| \\ &\le C \sum_{m=1}^{\infty} \frac{1}{m!} z^{2m} |\alpha_m| \le C \sum_{m=1}^{\infty} \frac{1}{m!} z^{2m} 2^{2m-1} (m-1)! \left(\frac{4}{\rho+1}\right)^{m-1} \\ &= \frac{C(\rho+1)}{8} \sum_{m=1}^{\infty} \frac{1}{m} \left(\frac{16 z^2}{\rho+1}\right)^m < \infty, \end{aligned} \tag{116}$$

provided that $\rho + 1 > 16z^2$. As a consequence, the limit $\lim_{B\to\infty}$ can be moved inside the summation sign in (115), giving

$$P\left\{\left|(W(1)+\delta)\,\Xi_\rho^{-1/2}\right| \le z\right\} = \sum_{m=1}^{\infty} \frac{1}{m!} G_\delta^{(m)}(\mu_\rho z^2)\alpha_m z^{2m} \tag{117}$$

when $\rho + 1 > 16z^2$.

Second, it follows from (37) that

$$P\left\{\left|\sqrt{T}\left(\hat{\beta}-\beta_0\right)/\hat{\omega}\right| \le z\right\} = E\left\{G_\delta(z^2\varsigma_{\rho T})\right\} + O(1/T). \tag{118}$$

But

$$E\left\{G_\delta(z^2\varsigma_{\rho T})\right\} = \sum_{m=1}^{\infty} \frac{1}{m!} G_\delta^{(m)}(\mu_{\rho T} z^2)\alpha_{m,T} z^{2m}, \tag{119}$$

where the right hand side converges to $E\left\{G_\delta(z^2\varsigma_{\rho T})\right\}$ uniformly over T because $\alpha_{m,T} = \alpha_m + O\left\{\frac{2^{2m-1}m!}{T^2}\left(\frac{4}{\rho+1}\right)^{m-2}\right\}$ uniformly over m by Lemma 8, $G_\delta^{(m)}(\cdot)$ is a bounded function and

$$\sum_{m=1}^{\infty} \frac{1}{m!} z^{2m} \frac{2^{2m-1}m!}{T^2}\left(\frac{4}{\rho+1}\right)^{m-2} = \frac{1}{T^2}\sum_{m=1}^{\infty}\frac{1}{2}(2z)^{2m}\left(\frac{4}{\rho+1}\right)^{m-2} < \infty, \tag{120}$$

when $\rho + 1 > 16z^2$. Therefore

$$P\left\{\left|\sqrt{T}\left(\hat{\beta}-\beta_0\right)/\hat{\omega}\right| \le z\right\} = \sum_{m=1}^{\infty} \frac{1}{m!} G_\delta^{(m)}(\mu_{\rho T} z^2)\alpha_{m,T} z^{2m} + O(\frac{1}{T}). \tag{121}$$

It follows from (117) and (121) that

$$\begin{aligned}
&\left|P\left\{\left|\sqrt{T}\left(\hat{\beta}-\beta_0\right)/\hat{\omega}\right| \le z\right\} - P\left\{\left|(W(1)+\delta)\,\Xi_\rho^{-1/2}\right| < z\right\}\right| \\
&= \left|\sum_{m=1}^{\infty} \frac{1}{m!} G_\delta^{(m)}(\mu_{\rho T} z^2)\alpha_{m,T} z^{2m} - \sum_{m=1}^{\infty} \frac{1}{m!} G_\delta^{(m)}(\mu_\rho z^2)\alpha_m z^{2m}\right| + O(\frac{1}{T}) \\
&= \left|\sum_{m=1}^{\infty} \frac{1}{m!} G_\delta^{(m)}(\mu_\rho z^2)\alpha_{m,T} z^{2m} - \sum_{m=1}^{\infty} \frac{1}{m!} G_\delta^{(m)}(\mu_\rho z^2)\alpha_m z^{2m}\right| + O(\frac{1}{T}) \qquad (122) \\
&= \left|\sum_{m=1}^{\infty} \frac{1}{m!} G_\delta^{(m)}(\mu_\rho z^2)\left(\alpha_{m,T}-\alpha_m\right) z^{2m}\right| + O(\frac{1}{T}) \\
&= O\left\{\frac{1}{T^2}\sum_{m=1}^{\infty} 2^{2m-1}\left(\frac{4}{\rho+1}\right)^{m-2} z^{2m}\right\} + O\left(\frac{1}{T}\right) \\
&= O\left(\frac{1}{T}\right),
\end{aligned}$$

where the second equality holds because $G_\delta^{(j)}(\mu_{\rho T}z^2) = G_\delta^{(j)}(\mu_\rho z^2) + O\left(\frac{1}{T}\right)$, Lemma 8 holds and $\sum_{m=1}^{\infty} \frac{1}{m!} G_\delta^{(m)}(\mu_{\rho T}z^2)\alpha_{m,T}z^{2m} < \infty$ uniformly over T, and the last equality follows from (120). This completes the proof of Theorem 4. ■

Proof of Theorem 5. It follows from Lemma 9 that when $\rho \to \infty$,

$$\alpha_{2,T} = \kappa_{2,T} = \frac{2}{\rho+1} + O\left(\frac{1}{T}\right), \tag{123}$$

$$\alpha_{3,T} = \kappa_{3,T} = O\left(\frac{1}{\rho^2}\right) + O\left(\frac{1}{T}\right), \tag{124}$$

and

$$\mu_{\rho T} = \mu_\rho - \frac{\rho}{T\omega_T^2} \sum_{h=-T+1}^{T-1} h\Gamma(h) + O\left(\frac{1}{T}\right) + O\left(\frac{\rho^2}{T^2}\right). \tag{125}$$

Thus, as $\rho \to \infty$,

$$\begin{aligned} F_T(z) &= P\left\{\left|\sqrt{T}\left(\hat{\beta} - \beta_0\right)/\hat{\omega}\right| \le z\right\} = E\left\{G_\delta(z^2\varsigma_{\rho T})\right\} + O(1/T) \\ &= G_\delta(\mu_{\rho T}z^2) + \frac{1}{2}G_\delta''(\mu_{\rho T}z^2)z^4\alpha_{2,T} + O\left(\alpha_{3,T}\right) \\ &= G_\delta(\mu_{\rho T}z^2) + \frac{1}{2}G_\delta''(\mu_{\rho T}z^2)z^4\left(\frac{2}{\rho+1} + O\left(\frac{1}{T}\right)\right) + O\left(\frac{1}{\rho^2}\right) + O\left(\frac{1}{T}\right) \end{aligned} \tag{126}$$

$$\begin{aligned} &= G_\delta(\mu_\rho z^2) + G_\delta'(\mu_\rho z^2)z^2\left(\mu_{\rho T} - \mu_\rho\right) + \frac{1}{\rho+1}G_\delta''(\mu_\rho z^2)z^4 \\ &+ O\left(\frac{1}{\rho^2}\right) + O\left(\frac{1}{T}\right) + O\left(\frac{\rho^2}{T^2}\right), \end{aligned} \tag{127}$$

using (123) to (125). But

$$\begin{aligned} G_\delta(\mu_\rho z^2) &= G_\delta(z^2) + G_\delta'(z^2)z^2\left(\mu_\rho - 1\right) + O\left(\frac{1}{\rho^2}\right) \\ &= G_\delta(z^2) - 2G_\delta'(z^2)z^2\frac{1}{\rho} + O\left(\frac{1}{\rho^2}\right), \end{aligned} \tag{128}$$

and

$$\begin{aligned} &G_\delta'(\mu_\rho z^2)z^2\left(\mu_{\rho T} - \mu_\rho\right) \\ &= \left(G_\delta'(z^2) + O\left(\frac{1}{\rho}\right)\right)z^2\left(-\frac{\rho}{T\omega_T^2}\sum_{h=1-T}^{T-1} h\Gamma(h) + O\left(\frac{1}{T}\right) + O\left(\frac{\rho^2}{T^2}\right)\right) \\ &= -\frac{\rho}{T\omega_T^2}\sum_{h=1-T}^{T-1} h\Gamma(h)G_\delta'(z^2)z^2 + O\left(\frac{1}{T}\right) + O\left(\frac{\rho^2}{T^2}\right), \end{aligned} \tag{129}$$

so

$$F_T(z) = G_\delta(z^2) + \left(G''_\delta(\mu_\rho z^2) z^4 - 2G'_\delta(z^2)z^2\right)\frac{1}{\rho} - d_{\gamma T} G'_\delta(z^2) z^2 \frac{\rho}{T}$$
$$+ O\left(\frac{1}{T}\right) + O\left(\frac{\rho^2}{T^2}\right) + O\left(\frac{1}{\rho^2}\right), \qquad (130)$$

as desired. ■

Proof of Corollary 6. Part (a). Using Theorem 5, we have, as $1/\rho + 1/T + \rho/T \to 0$

$$\begin{aligned} F_T(z_{\alpha,\rho}) &= D(z^2_{\alpha,\rho}) + \left[D''(z^2_{\alpha,\rho})z^4_{\alpha,\rho} - 2D'(z^2_{\alpha,\rho})z^2_{\alpha,\rho}\right]\frac{1}{\rho} \\ &\quad - d_{\gamma T} D'(z^2_{\alpha,\rho})z^2_{\alpha,\rho}\frac{\rho}{T} + O\left(\frac{1}{T} + \frac{\rho^2}{T^2} + \frac{1}{\rho^2}\right) \qquad (131) \\ &= F(z_{\alpha,\rho}) - d_{\gamma T} D'(z^2_{\alpha,\rho})z^2_{\alpha,\rho}\frac{\rho}{T} + O\left(\frac{1}{T} + \frac{\rho^2}{T^2} + \frac{1}{\rho^2}\right) \\ &= 1 - \alpha - d_{\gamma T} D'(z^2_{\alpha,\rho})z^2_{\alpha,\rho}\frac{\rho}{T} + O\left(\frac{1}{T} + \frac{\rho^2}{T^2} + \frac{1}{\rho^2}\right). \end{aligned}$$

So

$$1 - F_T(z_{\alpha,\rho}) - \alpha = d_{\gamma T} D'(z^2)z^2\frac{\rho}{T} + O\left(\frac{1}{T} + \frac{\rho^2}{T^2} + \frac{1}{\rho^2}\right). \qquad (132)$$

Part (b). Plugging $z^2_{\alpha,\rho}$ into (41) yields

$$\begin{aligned} &P\left(\left|\frac{\sqrt{T}(\hat\beta - \beta_0)}{\hat\omega}\right|^2 \geq z^2_{\alpha,\rho}\right) \\ &= 1 - G_\delta(z^2_{\alpha,\rho}) - \left[G''_\delta(z^2_{\alpha,\rho})z^4_{\alpha,\rho} - 2G'_\delta(z^2_{\alpha,\rho})z^2_{\alpha,\rho}\right]\frac{1}{\rho} \\ &+ d_{\gamma T} G'_\delta(z^2_{\alpha,\rho})z^2_{\alpha,\rho}\frac{\rho}{T} + O\left(\frac{1}{T}\right) + O\left(\frac{\rho^2}{T^2}\right) + O\left(\frac{1}{\rho^2}\right) \\ &= 1 - G_\delta(z^2_\alpha) - z^4_\alpha K_\delta\left(z^2_\alpha\right)\frac{1}{\rho} + d_{\gamma T} G'_\delta(z^2_\alpha)z^2_\alpha\frac{\rho}{T} + O\left(\frac{1}{T} + \frac{\rho^2}{T^2} + \frac{1}{\rho}\right), \qquad (133) \end{aligned}$$

where the last equality follows from the same proof as Theorem 3. ■

References

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

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

[3] de Jong, R. M. and J. Davidson (2000): "Consistency of Kernel Estimators of Heteroskedastic and Autocorrelated Covariance Matrices," *Econometrica,* 68, 407–424.

[4] den Haan, W. J. and A. Levin (1997): "A Practitioners Guide to Robust Covariance Matrix Estimation," in G. Maddala and C. Rao (eds), *Handbook of Statistics: Robust Inference,* Volume 15, Elsevier, New York.

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

[6] Hannan, E. J. (1970): *Multiple Time Series*, New York, Wiley.

[7] Hansen, B. E. (1992): "Consistent Covariance Matrix Estimation for Dependent Heterogenous Processes," *Econometrica,* 60, 967–972.

[8] Hong, Y. and J. Lee (2001): "Wavelet-based Estimation for Heteroskedastic and Autocorrelation Consistent Variance-covariance Matrices," Working paper, Cornell University.

[9] Jansson, M. (2002): "Consistent Covariance Matrix Estimation for Linear Processes," *Econometric Theory*, 18, 1449–1459.

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

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

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

[13] ——— (2002b): "Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation," *Econometrica,* 70, 2093–2095.

[14] ——— (2003): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," Working Paper, Center for Analytic Economics, Cornell University.

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

[16] ——— (1994): "Automatic Lag Selection in Covariance Estimation," *Review of Economic Studies*, 61, 631–654.

[17] Phillips, P. C. B. (1980): "Finite Sample Theory and the Distributions of Alternative Estimators of the Marginal Propensity to Consume," *Review of Economic Studies*, 47, 183–224.

[18] Phillips, P. C. B. (1993): "Operational Algebra and Regression t-Tests," In Peter C.B. Phillips (ed.), *Models, Methods, and Applications of Econometrics.* Basil Blackwell.

[19] Phillips, P. C. B. (2004): "HAC Estimation by Automated Regression," Forthcoming in *Econometric Theory*

[20] Phillips, P. C. B., Y. Sun and S. Jin (2003a): "Consistent HAC Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation," Cowles Foundation Discussion Paper No. 1407, Yale University.

[21] Phillips, P. C. B., Y. Sun and S. Jin (2003b): "Long Run Variance Estimation with Steep Origin Kernels without Truncation," Department of Economics, UCSD.

[22] Ravikumar, B., S. Ray and N. E. Savin (2004): "Robust Wald Testing with Exponentiated Kernels: A CAPM Example," Working paper, Department of Economics, University of Iowa.

[23] Shorack, G. R. and A. J. Wellner (1986): *Empirical Processes with Applications to Statistics*, John Wiley & Sons.

[24] Sul, D., P. C. B. Phillips and C-Y. Choi (2003): "Prewhitening Bias in HAC Estimation," Cowles Foundation Discussion Paper No. 1436, Yale University.

[25] Sun, Y. (2004): "Estimation of the Long-run Average Relationship in Nonstationary Panel Time Series," *Econometric Theory* 20, 1227–1260.

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

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

[28] Vogelsang, T. J. (2003): "Testing in GMM Models Without Truncation," Chapter 10 of *Advances in Econometrics* Volume 17, Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later, ed. by T. B. Fomby and R. C. Hill, Elsevier Science, 199–233.