POWER MAXIMIZATION AND SIZE CONTROL IN HETEROSKEDASTICITY AND AUTOCORRELATION ROBUST TESTS WITH EXPONENTIATED KERNELS

By

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

COWLES FOUNDATION PAPER NO. 1340

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

2011

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

Econometric Theory, **27**, 2011, 1320–1368.
doi:10.1017/S0266466611000077

POWER MAXIMIZATION AND SIZE CONTROL IN HETEROSKEDASTICITY AND AUTOCORRELATION ROBUST TESTS WITH EXPONENTIATED KERNELS

Yixiao Sun
University of California, San Diego

Peter C.B. Phillips
Yale University, University of Auckland, University of Southampton, Singapore Management University

Sainan Jin
Singapore Management University

Using the power kernels of Phillips, Sun, and Jin (2006, 2007), we examine the large sample asymptotic properties of the t-test for different choices of power parameter (ρ). We show that the nonstandard fixed-ρ limit distributions of the t-statistic provide more accurate approximations to the finite sample distributions than the conventional large-ρ limit distribution. We prove that the second-order corrected critical value based on an asymptotic expansion of the nonstandard limit distribution is also second-order correct under the large-ρ asymptotics. As a further contribution, we propose a new practical procedure for selecting the test-optimal power parameter that addresses the central concern of hypothesis testing: The selected power parameter is test-optimal in the sense that it minimizes the type II error while controlling for the type I error. A plug-in procedure for implementing the test-optimal power parameter is suggested. Simulations indicate that the new test is as accurate in size as the nonstandard test of Kiefer and Vogelsang (2002a, 2002b), and yet it does not incur the power loss that often hurts the performance of the latter test. The results complement recent work by Sun, Phillips, and Jin (2008) on conventional and bT HAC testing.

1. INTRODUCTION

Seeking to robustify inference, many practical methods in econometrics now make use of heteroskedasticity and autocorrelation consistent (HAC) covariance

The authors gratefully acknowledge partial research support from NSF under Grant no. SES-0752443 (Sun), SES-0647086 (Phillips), and NSFC under Grant no. 70501001 and 70601001 (Jin). Address correspondence to Yixiao Sun, Department of Economics, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0508; e-mail: yisun@ucsd.edu.

matrix estimates. Most commonly used HAC estimates are formulated using conventional kernel smoothing techniques (see den Haan and Levin (1997) for an overview), although quite different approaches like wavelets (Hong and Lee, 2001) and direct regression methods (Phillips, 2005b) 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, 2005, hereafter KV (2005); Sul, Phillips, and Choi, 2005).

To address the size distortion problem, Kiefer, Vogelsang, and Bunzel (2000) (hereafter KVB), Kiefer and Vogelsang (2002a, 2002b) (hereafter KV), and Vogelsang (2003) suggested setting the bandwidth equal to the sample size (i.e., $M = T$) in the construction of the long-run variance (LRV) estimation. While the resulting LRV estimate is inconsistent, the associated test statistic is asymptotically nuisance parameter-free, and critical values can be simulated from its nonstandard asymptotic distribution. We may therefore regard these procedures as falling within a general class of heteroskedastic and autocorrelation robust (HAR) techniques in econometrics (Phillips, 2005a). KV show by simulation that the nonstandard test has better size properties than the conventional asymptotic normal or chi-squared test. However, the size improvement comes at the cost of a clear power reduction. In order to reduce the power loss while maintaining the good size properties of the KVB test, KV (2005) set the bandwidth to be proportional to the sample size (T), i.e., $M = bT$ for some $b \in (0, 1]$. Their approach is equivalent to contracting traditional kernels $k(\cdot)$ to get $k_b(x) = k(x/b)$ and using the contracted kernels $k_b(\cdot)$ in the LRV estimation without truncation. In other work, Phillips, Sun, and Jin (2006, 2007) (hereafter PSJ) obtain a new class of kernels by exponentiating the conventional kernels using $k_\rho(\cdot) = (k(\cdot))^\rho$, where ρ is a power exponent parameter. For finite b and ρ, both contracted and exponentiated kernels are designed to reduce, but not totally eliminate, the randomness of the denominator in the t-ratio and in doing so help to improve the power of the t-test.

The parameter b in the KV approach and ρ in the PSJ approach are smoothing parameters that play an important role in balancing size distortion and potential power gain. In other words, the smoothing parameters entail some inherent trade-off between type I and type II errors. Both KV and PSJ suggest plug-in procedures that rely on conventional asymptotic theories to select b or ρ. More specifically, the plug-in selection of b or ρ suggested in these papers minimizes the asymptotic mean squared errors (AMSEs) of the underlying LRV estimator. In theory, such formulations ensure that the selected values $b \to 0$ and $\rho \to \infty$ as $T \to \infty$, and thus the fixed b or fixed ρ asymptotic theory is not applicable. However, to maintain good size properties or smaller type I errors in practical testing, KV and PSJ propose using critical values from the fixed b or fixed ρ asymptotic distributions, which are nonstandard, and treating the estimated b or ρ as fixed even though they are delivered by asymptotic formulas. PSJ justify this hybrid procedure on the basis of simulations that show the resulting tests have better size properties than tests that use standard normal asymptotic critical values. However, there are two remaining problems with this procedure. First, the rationale for good test size based on such plug-in values of ρ is not rigorously established. Second, the AMSE-optimal choice of ρ is not necessarily optimal in the context of hypothesis testing.

A primary contribution of the present paper is to provide analytic solutions to both of these problems by means of asymptotic expansions that provide higher-order information about the type I and type II errors. This approach provides a rigorous justification for the recommended procedure. We consider both first-order and second-order power kernels from PSJ (2006, 2007)), which have been found to work very well in simulations and empirical applications (see Ray and Savin, 2008; Ray, Savin, and Tiwari, 2009).

To investigate the size properties of the PSJ test, we consider the Gaussian location model, which is used also in Jansson (2004) and Sun, Phillips, and Jin (2008) (hereafter SPJ). Asymptotic expansions developed here reveal that the PSJ statistic is closer to its limit distribution when ρ is fixed than when ρ increases with T. More specifically, the error in rejection probability (ERP) of the t-test based on the nonstandard limiting distribution is of order $O(1/T)$ while that based on the standard normal is $O(1/T^{q/q+1})$. This result relates to similar results in Jansson and SPJ, who showed that the ERP of the KVB test is of order $O(\log T/T)$ and $O(1/T)$, respectively, while the ERP of the conventional test using the Bartlett kernel is at most $O(1/\sqrt{T})$, as shown in Velasco and Robinson (2001). These findings therefore provide theoretical support for the simulation results reported in PSJ (2006, 2007), Ray and Savin (2008), and Ray et al. (2009).

The PSJ test, which is based on the nonstandard limiting distribution, is not very convenient to use in practice as the critical values have to be simulated. To design an easy-to-implement test, we develop an expansion of the nonstandard limiting distribution about its limiting chi-squared distribution. A Cornish-Fisher-type expansion then leads to second-order corrected critical values. We find that the corrected critical values provide good approximations to the actual critical values of the nonstandard distribution. The PSJ test based on the corrected critical values has the advantage of being easily implemented and does not require the use of tables of nonstandard distributions.

To show that the hybrid PSJ test using a plug-in exponent and nonstandard critical value is generally less size-distorted than the conventional test, we develop a higher-order asymptotic expansion of the finite sample distribution of the t-statistic as T and ρ go to infinity simultaneously. It is shown that the corrected critical values obtained from the asymptotic expansion of the nonstandard distribution are also second-order correct under the conventional limiting theory. This finding provides a theoretical explanation for the size improvement of the hybrid PSJ test compared to the conventional test.

Combining the standard t-statistic and the high-order corrected critical values, we obtain a new t^*-test whose type I and type II errors can be approximately measured using the above asymptotic expansions. The type I and type II errors depend on the power exponent parameter used in the HAC estimation. Following Sun (2009), we propose to choose this parameter to minimize the type II error subject to the constraint that the type I error is bounded. The bound is defined to be $\kappa\alpha$, where α is the nominal type I error and $\kappa > 1$ is a parameter that captures the user's tolerance on the discrepancy between the nominal and true type I errors. The parameter κ is allowed to be sample-size dependent. For a smaller sample size, we may have a higher tolerance, while for larger sample sizes we may have lower tolerance. The new procedure addresses the central concern of classical hypothesis testing, viz., maximizing power subject to controlling size. For convenience we refer to the resulting ρ as the test-optimal ρ.

The test-optimal ρ is fundamentally different from the mean squared error-optimal (MSE-optimal) ρ that applies when minimizing the AMSE of the corresponding HAC variance estimate (cf. PSJ (2006, 2007)). When the tolerance factor κ is small enough, the test-optimal ρ is of smaller order than the MSE-optimal ρ. The test-optimal ρ can even be $O(1)$ for certain choices of the tolerance factor κ. The theory provides some theoretic justification for the use of fixed ρ rules in econometric testing. Simulation results show that the new plug-in procedure suggested in the present paper works remarkably well in finite samples.

To implement the test-optimal ρ for two-sided tests in a location model, users may proceed in the following steps:

1. Specify the null hypothesis $H_0 : \beta = \beta_0$ and an alternative hypothesis $H_1 : |\beta - \beta_0| = c_0 > 0$, where c_0 may reflect a value of scientific interest or economic significance if such a value is available. (In the absence of such a value, we recommend that the user set the default discrepancy parameter value $\delta = 2.3192$ in (1) below. This choice of δ ensures that the first-order power of the test is 75%).
2. Specify the significance level α of the test and the associated two-sided standard normal critical value z_α satisfying $\Phi(z_\alpha) = 1 - \alpha/2$. Specify a tolerance parameter κ so that $\kappa\alpha$ is the intended upper bound for the type I error. We suggest setting $\kappa = 1.1$ as the default value.
3. Estimate the model by ordinary least squares (OLS) and construct the residuals $\hat{u}_t$. Fit an AR(1) model to the estimated residuals and compute

$$\hat{d} = \frac{2\hat{\phi}}{\left(1-\hat{\phi}\right)^2}, \quad \hat{\sigma}^2 = \frac{1}{T^2}\sum_{t=1}^{T}\left(\hat{u}_t - \hat{\phi}\hat{u}_{t-1}\right)^2,$$

$$\delta = \sqrt{T}c_0\left(1-\hat{\phi}\right)/\hat{\sigma}, \tag{1}$$

where $\hat{\rho}$ is the OLS estimator of the autoregression (AR) coefficient. Set $\delta = 2.3192$ as a default value if the user is unsure about the alternative hypothesis.

4. Specify the kernel function to be used in HAC standard error estimation. Among the commonly used positive definite kernels, we recommend a suitable second-order kernel ($q = 2$) such as the Parzen or quadratic spectral (QS) kernel.

5. Compute the automatic power exponent parameter:

$$\hat{\rho}_{opt} = \left(\frac{c\delta^2 G'_\delta(z_\alpha^2, 3)}{2qg\hat{d}\left[\hat{\lambda}_{opt} G'_0(z_\alpha^2, 1) - G'_\delta(z_\alpha^2, 1)\right]} \right)^{q/q+1} T^{q^2/q+1}, \quad \textbf{(2)}$$

where

$$\hat{\lambda}_{opt} = \begin{cases} 0, & \text{if } \hat{d} < 0 \\ \frac{G'_\delta(z_\alpha^2,1)}{G'_0(z_\alpha^2,1)} + \frac{c\delta^2 G'_\delta(z_\alpha^2,3)\left[gdG'_0(z_\alpha^2,1)\right]^{1/q}}{2q[(\kappa-1)\alpha]^{1+1/q}T}\left(z_\alpha^2\right)^{1+1/q}, & \text{if } \hat{d} > 0 \end{cases}, \quad \textbf{(3)}$$

$$g = \begin{cases} 6.000, & \text{Parzen kernel} \\ 1.421, & \text{QS kernel} \end{cases}, \qquad c = \begin{cases} 0.539, & \text{Parzen kernel} \\ 1.000, & \text{QS kernel} \end{cases}, \quad \textbf{(4)}$$

and $G'_\delta(\cdot, k)$ is the probability density function (pdf) of a (non)central chi-squared variate with k degrees of freedom and noncentrality parameter δ^2.

6. Compute the HAC standard error using power parameter $\hat{\rho}_{opt}$ and construct the usual t-statistic $t_{\hat{\rho}_{opt}}$. Reject the null hypothesis if $|t_{\hat{\rho}_{opt}}| \geq \hat{z}_{\alpha,\hat{\rho}_{opt}}$, where

$$\hat{z}_{\alpha,\hat{\rho}_{opt}} = z_\alpha + \frac{1}{2}\left(\frac{\pi}{\hat{\rho}_{opt}g}\right)^{1/2}\left\{\left(1 + \frac{\sqrt{2}}{4}\right)z_\alpha + \frac{\sqrt{2}}{4}z_\alpha^3\right\}. \quad \textbf{(5)}$$

In related work, SPJ (2008) considered conventional kernels and selected the bandwidth to minimize a loss function formed from a weighted average of type I and type II errors. The present paper differs from SPJ in two aspects. First, while SPJ used the contracted kernel method, an exponentiated kernel approach is used here. An earlier version of the present paper followed SPJ and employed a loss function methodology to select the power exponent parameter, finding that for both LRV estimation and hypothesis testing, the finite sample performance of the exponentiated kernel method is similar to and sometimes better than that of the contracted kernel method. So, exponentiated kernels appear to have some natural advantages. A simulation study in the present paper provides some further evidence on these advantages. Second, the procedure for selecting the smoothing parameter is different in the present paper. While SPJ selected the smoothing parameter to minimize loss based on a weighted average of type I and type II errors, we minimize the type II error after controlling for the type I error. In effect, the loss function here is implicitly defined with an endogenous weight given by the Lagrange multiplier, while the loss function in SPJ is explicitly defined and thus requires a user-chosen weight. This requirement can be regarded as a drawback of the explicit loss function approach, especially when it is hard to evaluate the relative importance of type I and type II errors. Furthermore, the implicit loss function approach used here is more in line with the standard econometrics testing literature, where size control is often considered to be a priority.

The rest of the paper is organized as follows. Section 2 overviews the class of power kernels that are used in the present paper 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 nonstandard limit distribution under both the null and alternative hypotheses as the power parameter $\rho \to \infty$. Section 5 develops comparable finite sample expansions of the statistic as $T \to \infty$ for a fixed ρ and as both $T \to \infty$ and $\rho \to \infty$. Section 6 proposes a selection rule for ρ that is suitable for implementation in semiparametric testing. Section 7 reports some simulation evidence on the performance of the new procedure. Section 8 concludes. Proofs and additional technical results are in the Appendix.

2. HETEROSKEDASTICITY AND AUTOCORRELATION ROBUST INFERENCE FOR THE MEAN

Consider the location model

$$y_t = \beta + u_t, \qquad t = 1, 2, \ldots, T, \tag{6}$$

with u_t autocorrelated and possibly heteroskedastic, and $\mathrm{E}(u_t) = 0$. To test a hypothesis about β, we consider the OLS estimator $\hat{\beta} = \bar{Y} = T^{-1}\sum_{t=1}^{T} y_t$. Recentering and normalizing gives us

$$\sqrt{T}(\hat{\beta} - \beta) = T^{-1/2} S_T, \tag{7}$$

where $S_t = \sum_{\tau=1}^{t} u_\tau$.

We impose the following convenient high-level condition (e.g., KVB; PSJ, 2006, 2007; Jansson, 2004).

Assumption 1. The partial sum process $S_{[Tr]}$ satisfies the functional law $T^{-1/2} S_{[Tr]} \Rightarrow \omega W(r),\ r \in [0,1]$, where ω^2 is the long-run variance of u_t and $W(r)$ is the standard Brownian motion.

Thus, $\sqrt{T}(\hat{\beta} - \beta) \Rightarrow \omega W(1) = N(0, \omega^2)$. Let $\hat{u}_\tau = y_\tau - \hat{\beta}$ be the demeaned time series, and define the corresponding partial sum process $\hat{S}_t = \sum_{\tau=1}^{t} \hat{u}_\tau$. Under Assumption 1, we have $T^{-1/2} \hat{S}_{[Tr]} \Rightarrow \omega V(r),\ r \in [0,1]$, where V is a standard Brownian bridge process. When u_t is stationary, the long-run variance of u_t is

$\omega^2 = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma(j)$, where $\gamma(j) = \mathrm{E}(u_t u_{t-j})$. The conventional approach to estimating ω^2 typically involves smoothing and truncation lag covariances using kernel-based nonparametric HAC estimators. HAC estimates of ω^2 typically have the form

$$\hat{\omega}^2(M) = \sum_{j=-T+1}^{T-1} k(\frac{j}{M})\hat{\gamma}(j), \ \hat{\gamma}(j) = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-j} \hat{u}_{t+j}\hat{u}_t & \text{for } j \geq 0 \\ \frac{1}{T}\sum_{t=-j+1}^{T} \hat{u}_{t+j}\hat{u}_t & \text{for } j < 0 \end{cases}, \tag{8}$$

where $\hat{\gamma}(j)$ are sample covariances, $k(\cdot)$ is some kernel function, and M is a bandwidth parameter. Consistency of $\hat{\omega}^2(M)$ requires that M grows with the sample size T but at a slower rate so that $M = o(T)$ (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 nonparametrically studentized t-ratio statistic is of the form

$$t_{\hat{\omega}(M)} = T^{1/2}(\hat{\beta} - \beta_0)/\hat{\omega}(M), \tag{9}$$

which is asymptotically standard normal. Tests based on $t_{\hat{\omega}(M)}$ and critical values from the standard normal are subject to size distortion, especially when there is strong autocorrelation in the time series.

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, taking the form $M = bT$. These (so called fixed-b) estimates are inconsistent and tend to random quantities instead of ω^2, so the limit distribution of (9) is no longer standard normal. Nonetheless, use of these estimates results in valid asymptotically similar tests. For convenience and as in the Introduction, we refer to the standard approach as the contracted kernel approach.

In related work, PSJ (2006, 2007) propose the use of estimates of ω^2 based on power kernels or exponentiated kernels without truncation. The power kernels were constructed by taking an arbitrary integer power $\rho \geq 1$ of conventional kernels. In this paper we consider the power kernels $k_\rho(x) = (k_{BART}(x))^\rho$, $(k_{PR}(x))^\rho$, or $\left(k_{QS}(x)\right)^\rho$, where

$$k_{BART}(x) = \begin{cases} (1-|x|), & |x| \leq 1 \\ 0, & |x| > 1 \end{cases},$$
$$k_{PR}(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & \text{for } 0 \leq |x| \leq 1/2 \\ 2(1-|x|)^3, & \text{for } 1/2 \leq |x| \leq 1 \\ 0, & \text{otherwise} \end{cases},$$
$$k_{QS}(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right)$$

are the Bartlett, Parzen, and QS kernels, respectively. These kernels have a linear or quadratic expansion at the origin:

$$k(x) = 1 - gx^q + o\left(x^q\right), \quad \text{as } x \to 0+,$$

where $g = 1$, $q = 1$ for the Bartlett kernel, $g = 6$, $q = 2$ for the Parzen kernel, and $g = 18\pi^2/125$, $q = 2$ for the QS kernel. For convenience, we call the exponentiated Bartlett kernels the first-order power kernels and the exponentiated Parzen and QS kernels the second-order power kernels.

Using k_ρ in (8) and letting $M = T$ gives HAC estimates of the form

$$\hat{\omega}_\rho^2 = \sum_{j=-T+1}^{T-1} k_\rho\left(\frac{j}{T}\right)\hat{\gamma}(j). \tag{10}$$

The associated t-statistic is given by

$$t^*\left(\hat{\omega}_\rho\right) = T^{1/2}(\hat{\beta} - \beta_0)/\hat{\omega}_\rho. \tag{11}$$

When the power parameter ρ is fixed as $T \to \infty$, PSJ (2006, 2007) show that 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 has the nonstandard limit distribution

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

under the null and

$$t^*\left(\hat{\omega}_\rho\right) \Rightarrow (\delta + W(1))\,\Xi_\rho^{-1/2}, \tag{13}$$

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

When ρ goes to ∞ at a certain rate, PSJ (2006, 2007) further show 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 standard 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 $\hat{\omega}_\rho$ and to employ critical values that are second-order corrected. The correction is first obtained using an accurate but simple asymptotic expansion of the nonstandard distribution about its limiting chi-squared distribution that applies as $\rho \to \infty$. This expansion is developed in Section 4. The correction is further justified by an asymptotic expansion of the finite sample distribution in Section 5.

3. PROBABILITY DENSITIES OF THE NONSTANDARD LIMIT DISTRIBUTION AND THE FINITE SAMPLE DISTRIBUTION

This section develops some useful formulas 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

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

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 $u = (u_1, ...u_T)'$, $y = (y_1, ..., y_T)'$, $l_T = (1, ..., 1)'$, and $\Omega_T = var(u)$. Then the generalized least squares (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 $\hat{u} = \left(I - l_T(l_T'l_T)^{-1}l_T'\right)u = \left(I - l_T(l_T'l_T)^{-1}l_T'\right)\tilde{u}$. In 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

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

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, where $\partial q = \partial/\partial q$,

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

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

Asymptotic expansions of the limit distributions given in (12) and (13) can be obtained as the power exponent parameter $\rho \to \infty$. Moreover, these expansions may be developed for both the central and noncentral chi-squared limit distributions that apply when $\rho \to \infty$, corresponding to the null and alternative hypotheses. The approach adopted here is to work with the expansion of the noncentral distribution and is therefore closely related to the approach used in SPJ (2008) for studying standard and bT type tests.

Let $G_\lambda = G(\cdot;\lambda^2)$ be the cumulative distribution function (CDF) of a noncentral $\chi_1^2(\lambda^2)$ variate with noncentrality parameter λ^2; then $P\left\{\left|(\delta + W(1))\,\Xi_\rho^{-1/2}\right| \le z\right\} = P\left\{(\delta + W(1))^2 \le \Xi_\rho z^2\right\} = \mathrm{E}\left\{G_\delta(\Xi_\rho z^2)\right\}$. An expansion of $\mathrm{E}\left\{G_\delta(\Xi_\rho z^2)\right\}$ can be developed in terms of the moments of $\Xi_\rho - \mu_\rho$, where $\mu_\rho = \mathrm{E}\left(\Xi_\rho\right)$ and $\sigma_\rho^2 = \mathrm{var}\left(\Xi_\rho\right)$. In particular, we have

$$\mathrm{E}G_\delta(\Xi_\rho z^2) = G_\delta(\mu_\rho z^2) + \frac{1}{2}G_\delta''(\mu_\rho z^2)\mathrm{E}\left(\Xi_\rho - \mu_\rho\right)^2 z^4 + \frac{1}{6}\mathrm{E}\left[G_\delta'''(\mu_\rho z^2)\left(\Xi_\rho - \mu_\rho\right)^3 z^6\right] + O\left\{\mathrm{E}\left(\Xi_\rho - \mu_\rho\right)^4\right\}, \quad \textbf{(18)}$$

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.

It is easy to see that $\Xi_\rho = \int_0^1\int_0^1 k_\rho^*(r,s)dW(r)dW(s)$, where $k_\rho^*(r,s)$ is defined by

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

Since $k_\rho^*(r,s)$ is a positive semidefinite kernel (see Sun, 2004, for a proof), it can be represented as $k_\rho^*(r,s) = \sum_{i=1}^\infty \lambda_n^* f_n^*(r) f_n^*(s)$ by Mercer's theorem, where $\lambda_n^* > 0$ are the eigenvalues of the kernel and $f_n^*(r)$ are the corresponding eigenfunctions, i.e., $\lambda_n^* f_n^*(s) = \int_0^1 k_\rho^*(r,s) f_n^*(r)dr$. Using this representation, we can write Ξ_ρ as $\Xi_\rho = \sum_{n=1}^{\infty} \lambda_n^* Z_n^2$, where $Z_n \sim iidN(0,1)$. Therefore, the characteristic function of $\Xi_\rho - \mu_\rho$ is given by $\phi(t) = \mathrm{E}\left\{e^{it(\Xi_\rho - \mu_\rho)}\right\} = e^{-it\mu_\rho}\prod_{n=1}^{\infty}\left\{1-2i\lambda_n^* t\right\}^{-1/2}$.

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

$$\kappa_1 = 0\,,\ \kappa_m = 2^{m-1}(m-1)!\int_0^1 \cdots \int_0^1 \left(\prod_{j=1}^{m} k_\rho^*(\tau_j, \tau_{j+1})\right) d\tau_1 \ldots d\tau_m, \tag{19}$$

where $\tau_1 = \tau_{m+1}$ and $m \geq 2$.

These calculations enable us to develop an asymptotic expansion of $\mathrm{E}\{G_\delta(\Xi_\rho z^2)\}$ as the power parameter $\rho \to \infty$. A full series expansion is possible using this method, but we only require the leading term in the expansion in what follows. Ignoring the technical details, we have, up to smaller-order terms,

$$P\left\{\left|(\delta + W(1))\,\Xi_\rho^{-1/2}\right| \leq z\right\} = G_\delta(z^2) - z^2 G_\delta'\left(z^2\right)\left(\int_0^1\int_0^1 k_\rho(r-s)drds\right) + z^4 G_\delta''(z^2)\left(\int_0^1\int_0^1 k_{2\rho}(r-s)drds\right). \tag{20}$$

Depending on the value of q, the integral $\left(\int_0^1\int_0^1 k_\rho(r-s)drds\right)$ has different expansions as $\rho \to \infty$. For first-order power kernels, direct calculation yields

$$\int_0^1\int_0^1 k_\rho(r-s)drds = \frac{2}{\rho} + O\left(\frac{1}{\rho^2}\right).$$

For second-order power kernels, we use the Laplace approximation and obtain

$$\int_0^1\int_0^1 k_\rho(r-s)drds = \left(\frac{\pi}{\rho g}\right)^{1/2} + O\left(\frac{1}{\rho}\right).$$

It is clear that the rate of decay of the integral depends on the local behavior of the kernel function at the origin.

Plugging these expressions into (20), we obtain the following theorem.

THEOREM 1. *Let* $F_\delta(z) := P\left\{\left|(\delta + W(1))\,\Xi_\rho^{-1/2}\right| \leq z\right\}$ *be the nonstandard limiting distribution. Then as* $\rho \to \infty$,

$$F_\delta(z) = G_\delta(z^2) + \rho^{-1/q} c_q \mathcal{L}_q(G_\delta, z) + O\left(\rho^{-2/q}\right), \tag{21}$$

where

$$\mathcal{L}_q(G_\delta, z) = \left\{G_\delta''(z^2)z^4 - 2G_\delta'(z^2)z^2\right\} I(q=1) + \left\{G_\delta''(z^2)z^4 - \sqrt{2}G_\delta'(z^2)z^2\right\} I(q=2),$$

$$c_q = gI(q=1) + \left(\frac{\pi}{2g}\right)^{1/2} I(q=2),$$

and the remainder $O(\cdot)$ *term holds uniformly for any* $z \in [M_l, M_u]$ *with* $0 < M_l < M_u < \infty$.

When $\delta = 0$, we have

$$F_0(z) = D(z^2) + c_q \rho^{-1/q} \mathcal{L}_q(D, z) + O\left(\rho^{-2/q}\right), \tag{22}$$

where $D(\cdot) = G_0(\cdot)$ is the CDF of χ_1^2 distribution. For any $\alpha \in (0,1)$, let $z_{\alpha,\rho}^2 \in \mathbb{R}^+$, $z_\alpha^2 \in \mathbb{R}^+$, such that $F_0(z_{\alpha,\rho}) = 1-\alpha$, $D(z_\alpha^2) = 1-\alpha$. Then, using a Cornish-Fisher-type expansion, we obtain the following corollary.

COROLLARY 1. *Second-order corrected critical values based on the expansion (22) are as follows:*

(i) For the first-order power kernel,

$$z_{\alpha,\rho} = z_\alpha + \frac{1}{4\rho}(5z_\alpha + z_\alpha^3) + O\left(\frac{1}{\rho^2}\right). \tag{23}$$

(ii) For the second-order power kernel,

$$z_{\alpha,\rho} = z_\alpha + \frac{1}{2}\left(\frac{\pi}{\rho g}\right)^{1/2} \left\{ \left(1 + \frac{\sqrt{2}}{4}\right) z_\alpha + \frac{\sqrt{2}}{4} z_\alpha^3 \right\} + O\left(\frac{1}{\rho}\right), \tag{24}$$

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 values at the 5% level for the Bartlett, Parzen, and QS kernels are

$$z_{\alpha,\rho}^{BART} = 1.96 + \frac{4.3325}{\rho}, \qquad z_{\alpha,\rho}^{PAR} = 1.96 + \frac{1.9230}{\sqrt{\rho}}, \qquad z_{\alpha,\rho}^{QS} = 1.96 + \frac{3.9511}{\sqrt{\rho}}, \tag{25}$$

respectively. These are also the critical values for the one-sided test ($>$) at the 2.5% level. Similarly, the corrected critical values for $\alpha = 0.10$ are given by

$$z_{\alpha,\rho}^{BART} = 1.645 + \frac{3.1691}{\rho}, \qquad z_{\alpha,\rho}^{PAR} = 1.645 + \frac{1.3750}{\sqrt{\rho}}, \qquad z_{\alpha,\rho}^{QS} = 1.645 + \frac{2.8252}{\sqrt{\rho}}. \tag{26}$$

We can evaluate the accuracy of the approximate critical values by comparing them with the exact ones obtained via simulations. In all three cases, the second-order corrected critical values are remarkably close to the exact ones. Details are available upon request.

Since the limiting distributions (12) and (13) are valid for general regression models under certain conditions on the regressors (see PSJ, 2006, 2007), the corrected critical value $z_{\alpha,\rho}$ may be used for hypothesis testing in a general regression framework.

When $\delta \neq 0$ and the corrected critical values are used, we can establish the local asymptotic power in the following corollary.

COROLLARY 2. *The local asymptotic power satisfies*

$$P\left\{\left|(\delta + W(1))\,\Xi_\rho^{-1/2}\right| > z_{\alpha,\rho}\right\} = 1 - G_\delta(z_\alpha^2) - c_q z_\alpha^4 K_\delta\left(z_\alpha^2\right)\rho^{-1/q} + O(\rho^{-2/q}), \tag{27}$$

where

$$K_\delta(z) = G_\delta''(z) - \frac{D''(z)}{D'(z)} G_\delta'(z) = \frac{\delta^2}{2z} G_\delta'(z,3), \tag{28}$$

and $G_\delta'(z,3)$ is the pdf of the noncentral chi-square distribution with 3 degrees of freedom.

Corollary 2 shows that the asymptotic test power increases monotonically with ρ when ρ is large. For a given critical value z_α, the function $f(\delta) = z_\alpha^4 K_\delta\left(z_\alpha^2\right)$ obtains its maximum around $\delta = 2$, implying that the power improvement from choosing a large ρ is greatest when the local alternative is in an intermediate neighborhood of the null. The default value of δ we use in developing a test-optimal smoothing parameter choice is in this intermediate neighborhood.

5. EXPANSIONS OF THE FINITE SAMPLE DISTRIBUTION

Following SPJ (2008), we now develop an asymptotic expansion of the finite sample distribution of the t-statistic in a simple location model. We start with the following weak dependence condition.

Assumption 2. Assume u_t is a mean zero stationary Gaussian process with $\sum_{h=-\infty}^{\infty} h^2 |\gamma(h)| < \infty$, *where* $\gamma(h) = \mathrm{E} u_t u_{t-h}$.

In what follows, we develop an asymptotic expansion of $P\{\left|\sqrt{T}(\hat\beta - \beta_0)/\hat\omega\right| \le z\}$ for $\hat\omega = \hat\omega_\rho$ and for local alternatives of the form $\beta = \beta_0 + c/\sqrt{T}$. A complicating factor in the development is that $\sqrt{T}(\hat\beta - \beta)$ and $\hat\omega$ are in general statistically dependent due to the autocorrelation structure of u_t. To overcome this difficulty, we decompose $\hat{\beta}$ and $\hat{\omega}$ into statistically independent components as in Section 3. After some manipulation, we obtain

$$F_{T,\delta}(z) := P\left\{\left|\sqrt{T}\left(\hat{\beta}-\beta_0\right)/\hat{\omega}\right| \le z\right\} = \mathrm{E}\left\{G_\delta(z^2 \varsigma_{\rho T})\right\} + O\left(T^{-1}\right) \quad \textbf{(29)}$$

uniformly over $z \in \mathbb{R}^+$, where $\varsigma_{\rho T} := \left(\hat{\omega}/\omega_T\right)^2$ converges weakly to Ξ_ρ, and $\omega_T^2 := \mathrm{var}\left(\sqrt{T}(\hat{\beta}-\beta)\right)$.

Since $\hat{\omega}^2 = T^{-1}\hat{u}' W_\rho \hat{u} = T^{-1} u' A_T W_\rho A_T u$, where W_ρ is $T \times T$ with (j,s)-th element $k_\rho((j-s)/T)$ and $A_T = I_T - l_T l_T'/T$, $\varsigma_{\rho T}$ is a quadratic form in a Gaussian vector. To evaluate $\mathrm{E}\left\{G_\delta(z^2\varsigma_{\rho T})\right\}$, we proceed to compute the cumulants of $\varsigma_{\rho T} - \mu_{\rho T}$ for $\mu_{\rho T} := \mathrm{E}\varsigma_{\rho T}$. 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 $\Omega_T = \mathrm{E}(uu')$ and the cumulant generating function is

$$\ln\left(\phi_{\rho T}(t)\right) = -\frac{1}{2}\log\det\left(I - 2it\frac{\Omega_T A_T W_\rho A_T}{T\omega_T^2}\right) - it\mu_{\rho T} := \sum_{m=1}^{\infty} \kappa_{m,T}\frac{(it)^m}{m!}, \quad \textbf{(30)}$$

where the $\kappa_{m,T}$ are the cumulants of $\varsigma_{\rho T} - \mu_{\rho T}$. It follows from (30) that $\kappa_{1,T} = 0$ and

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

By proving that $\kappa_{m,T}$ is close to κ_m in the precise sense given in Lemma A.3 in the Appendix, we can establish the following theorem, which gives the order of magnitude of the error in the nonstandard limit distribution of the t-statistic as $T \to \infty$ with fixed ρ.

THEOREM 2. *Let Assumption 2 hold. If* $\int_0^1 k_\rho(v)dv < 1/\left(16z^2\right)$, *then*

$$F_{T,\delta}(z) = F_\delta(z) + O\left(T^{-1}\right), \quad \textbf{(31)}$$

as $T \to \infty$ *with fixed* ρ.

The requirement $\int_0^1 k_\rho(v)dv < 1/\left(16z^2\right)$ on ρ is a technical condition in the proof. It can be relaxed but at the cost of more tedious calculations. Now $1 - F_{T,\delta}(z)$ gives the power of the test under the alternative hypothesis $H_1 : \beta \neq \beta_0$. Theorem 2 indicates that when ρ is fixed the power of the test can be approximated by $1 - F_\delta(z)$ with an error of order $O(1/T)$. Under the null hypothesis $H_0 : \beta = \beta_0$, $\delta = 0$, Theorem 2 shows that for fixed ρ the ERP for tests using critical values obtained from the nonstandard limit distribution of $W(1)\Xi_\rho^{-1/2}$ is

$O\left(T^{-1}\right)$. This rate is faster than the rate for conventional tests based on consistent HAC estimates.

Combined with Theorem 1, Theorem 2 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 this sequential limit theory, the size distortion of the t-test based on the corrected critical values is

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

and the corresponding local asymptotic power is

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

To evaluate the order of size distortion, we have to compare the orders of magnitude of $\rho^{-2/q}$ and $1/T$. Such a comparison jeopardizes the sequential nature of the limiting directions and calls for a higher-order approximation that allows $T\to\infty$ and $\rho\to\infty$ simultaneously. A corollary to this observation is that fixed ρ asymptotics do not provide an internally consistent framework for selecting the optimal power parameter.

The next theorem establishes the required higher-order expansion.

THEOREM 3. *Let Assumption 2 hold. If* $1/\rho+\rho/T^q\to 0$ *as* $T\to\infty$, *then*

$$F_{T,\delta}(z) = G_\delta(z^2)+c_q\rho^{-1/q}\mathcal{L}_q\left(G_\delta,z\right)-gd_{\gamma T}G_\delta'(z^2)z^2\left(\rho T^{-q}\right)$$
$$+o\left(\rho T^{-q}\right)+O\left(T^{-1}+\rho^{-2/q}\right), \quad \textbf{(32)}$$

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

Under the local alternative hypothesis $H_1 : |\beta-\beta_0| = c/\sqrt{T}$, the power of the test based on the corrected critical values is $1-F_{T,\delta}(z_{\alpha,\rho})$. Theorem 3 shows that $F_{T,\delta}(z_{\alpha,\rho})$ can be approximated by $G_\delta(z^2)+c_q\rho^{-1/q}L_q\left(G_\delta,z\right)-gd_{\gamma T}G_\delta'(z^2)z^2\left(\rho T^{-q}\right)$, and the approximation error is of order $o\left(\rho T^{-q}\right)+O\left(T^{-1}+\rho^{-2/q}\right)$.

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

$$F_{T,0}(z) = D(z^2)+c_q\rho^{-1/q}\mathcal{L}_q\left(D,z\right)-gd_{\gamma T}D'(z^2)z^2\left(\rho T^{-q}\right)$$
$$+o\left(\rho T^{-q}\right)+O\left(T^{-1}+\rho^{-2/q}\right). \quad \textbf{(33)}$$

Importantly, the leading term $D(z^2)+c_q\rho^{-1/q}L_q\left(D,z\right)$ in this expansion is the same as the corresponding expansion of the limit distribution $F_0(z)$ given in (22). A direct implication is that the corrected critical values from the nonstandard limiting distribution are also second-order correct under the conventional large ρ asymptotics. More specifically, up to smaller-order terms, $F_{T,0}(z_{\alpha,\rho}) = 1-\alpha - gd_{\gamma T}D'(z_\alpha^2)z_\alpha^2\left(\rho T^{-q}\right)$. By adjusting the critical values, we eliminate the term $c_q\rho^{-1/q}L_q(D,z)$, which reflects the randomness of the standard error estimator. So if the remaining term $-gd_{\gamma T}D'(z_\alpha^2)z_\alpha^2\left(\rho T^{-q}\right)$ is of smaller order than the eliminated term, the use of the corrected critical values given in Corollary 1 should reduce the size distortion. When ρ increases with T, the remaining term $gd_{\gamma T}D'(z_\alpha^2)z_\alpha^2\left(\rho T^{-q}\right)$ approximately measures the size distortion in tests based on the corrected critical values, or equivalently those based on the nonstandard limit theory, at least to order $O(\rho^{-1/q})$.

If critical values from the standard normal are used, then the ERP is given by the $O\left(\rho^{-1/q}\right)$ and $O(\rho T^{-q})$ terms. To obtain the best rate of convergence of the ERP to zero, we set $\rho = O(T^{q^2/q+1})$ to balance these two terms, and the resulting rate of convergence is of order $O(T^{-q/q+1})$. As we discussed before, this rate is larger than that of the nonstandard test by an order of magnitude. One might argue that this comparison is not meaningful, as the respective orders of the ERP are obtained under different asymptotic specifications of ρ : one for growing ρ and the other one for fixed ρ. To sharpen the comparison, we can compare the ERP under the same asymptotic specification as in SPJ (2008). When ρ is fixed, the ERP of the standard normal test contains an $O(1)$ term, which makes the ERP larger by an order of magnitude than the ERP of the nonstandard test. When ρ grows with T, the ERP of the nonstandard test contains fewer terms than that of the standard normal test. This is usually regarded as an advantage in the econometrics literature.

If we set $\rho = O\left(T^{2q^2/2q+1}\right)$, which is the AMSE-optimal rate for the exponent, then the ERP of the standard normal test is of order $O\left(\rho T^{-q}\right) = O(T^{-q/(2q+1)})$. In this case, the ERP of the nonstandard test or test using the second-order corrected critical values is also of order $O(T^{-q/(2q+1)})$. Therefore, compared with the standard normal test, the hybrid procedure suggested in PSJ (2006, 2007) does not reduce the ERP by an order of magnitude. However, the hybrid procedure eliminates the $O(\rho^{-1/q})$ term from the ERP and thus reduces the overrejection that is commonly seen in economic applications. This explains the better finite sample performances found in the simulation studies of PSJ (2006, 2007), Ray and Savin (2008), and Ray et al. (2009).

For convenience, we refer to the test based on the t^*-statistic $t^*(\hat{\omega}_\rho)$ and the second-order corrected critical values as the t_ρ^*-test. We formalize the results on the size distortion and local power expansion of the t_ρ^*-test in the following corollary.

COROLLARY 3. *Let Assumption 2 hold. If* $1/\rho + \rho/T^q \to 0 \to 0$ *as* $T \to \infty$, *then*

(a) *the size distortion of the* t_ρ^**-test is*

$$gd_{\gamma T}D'(z_\alpha^2)z_\alpha^2\rho T^{-q} + o\left(\rho T^{-q}\right) + O\left(T^{-1} + \rho^{-2/q}\right); \tag{34}$$

(b) under the local alternative, the power of the t_ρ^*-test is

$$1 - G_\delta(z_\alpha^2) - c_q z_\alpha^4 K_\delta\left(z_\alpha^2\right)\rho^{-1/q} + g d_{\gamma T} G'_\delta(z_\alpha^2) z_\alpha^2 \rho T^{-q}$$
$$+ o\left(\rho T^{-q}\right) + O\left(T^{-1} + \rho^{-2/q}\right). \quad \textbf{(35)}$$

Just as in standard limit theory, the nonstandard limit theory does not address the bias problem due to nonparametric smoothing in LRV estimation. Comparing (35) with (27), we get an additional term, reflecting the asymptotic bias of the LRV estimator.

According to Corollary 3 and ignoring the higher-order terms, the type I and type II errors of the t_ρ^*-test can be measured by

$$e_I^\rho = \alpha + g d_{\gamma T} D'(z_\alpha^2) z_\alpha^2 \rho T^{-q}, \quad \textbf{(36)}$$
$$e_{II}^\rho = G_\delta(z_\alpha^2) + c_q z_\alpha^4 K_\delta\left(z_\alpha^2\right)\rho^{-1/q} - g d_{\gamma T} G'_\delta(z_\alpha^2) z_\alpha^2 \rho T^{-q},$$

respectively. SPJ (2008) obtained similar expressions using the contracted kernel approach with bandwidth set equal to $M = bT$. They showed that approximately

$$e_I^b = \alpha + g d_{\gamma T} D'(z_\alpha^2) z_\alpha^2 (bT)^{-q},$$
$$e_{II}^b = G_\delta(z_\alpha^2) + \mu_2 z_\alpha^4 K_\delta(z_\alpha^2) b - g d_{\gamma T} G'_\delta(z_\alpha^2) z_\alpha^2 (bT)^{-q}, \quad \textbf{(37)}$$

where $\mu_2 = \int_{-\infty}^{\infty} k^2(x)dx$. Clearly, the approximations in (36) and (37) are closely related.

To match the orders of magnitude in the above approximations, we can let $\rho = b^{-q}$. In this case, $e_I^\rho = e_I^b$ and $e_{II}^\rho = e_{II}^b + \left(c_q - \int_{-\infty}^{\infty} k^2(x)dx\right) z_\alpha^4 K_\delta(z_\alpha^2) b$. That is, in terms of the type I error, the exponentiated kernel approach and the contracted kernel approach are asymptotically equivalent. However, they differ in the type II error. For the Bartlett, Parzen, and QS kernels, we have, using $2zK_\delta(z) = \delta^2 G'_\delta(z,3)$,

$$e_{II}^\rho = \begin{cases} e_{II}^b + 0.1666 \times \delta^2 z_\alpha^2 G'_\delta(z_\alpha^2, 3) b, \text{ Bartlett kernel} \\ e_{II}^b - 0.0138 \times \delta^2 z_\alpha^2 G'_\delta(z_\alpha^2, 3) b, \text{ Parzen kernel} \\ e_{II}^b + 0.0257 \times \delta^2 z_\alpha^2 G'_\delta(z_\alpha^2, 3) b, \text{ QS kernel} \end{cases}.$$

To calibrate the numerical difference, we take $z_\alpha = 1.645$ and $\delta = 2.3192$. According to the first-order asymptotics, the type II error is 25%. For this choice, $\delta^2 z_\alpha^2 G'_\delta(z_\alpha^2, 3) = 0.99686$. So for the Bartlett kernel, there is some advantage of using the contracted kernel approach. For the Parzen kernel, the exponentiated kernel approach is better, while for the QS kernel the contracted kernel approach is better. However, the difference is small, especially when b is close to zero.

The above comparison is based on higher-order expansions. It may not completely reflect the finite sample performances, especially when the sample size is small. In the simulation study, we will compare the exponential kernel approach with the conventional contracted kernel approach.

6. OPTIMAL EXPONENT CHOICE

When estimating the long-run variance, PSJ (2006, 2007) show there is an optimal choice of ρ that minimizes the AMSE of the estimator and give an optimal expansion rate of $O\left(T^{2q^2/(2q+1)}\right)$ for ρ in terms of the sample size T. The present paper attempts to provide a new approach for optimal exponent selection that addresses the central concern of classical hypothesis testing, which can be expressed as minimizing the type II error subject to controlling the type I error. Sun (2009) employs this constraint minimization approach in related work.

More specifically, we propose to solve

$$\min_{\rho} e_{II}^{\rho}, \quad \text{s.t. } e_I^{\rho} \le \kappa\alpha \ , \tag{38}$$

where e_I^{ρ} and e_{II}^{ρ} are approximate measures of type I and type II errors given in the previous section. Here κ is a constant greater than 1. Ideally, the type I error is less than or equal to the nominal type I error α. In finite samples, there is always some approximation error, and we allow for some discrepancy by introducing the tolerance factor κ. For example, when $\alpha = 5\%$ and $\kappa = 1.2$, we aim to control the type I error such that it is not greater than 6%. Note that κ may depend on the sample size T. For a larger sample size, we may require κ to take smaller values.

The solution to the minimization problem depends on the sign of $d_{\gamma T}$. When $d_{\gamma T} < 0$, the constraint $e_I^{\rho} \le \kappa\alpha$ is not binding, and we have the unconstrained minimization problem $\rho_{\text{opt}} = \arg\min_{\rho} e_{II}^{\rho}$, whose solution for the optimal ρ is

$$\rho_{\text{opt}} = \left(- \frac{c_q z_\alpha^2 K_\delta\left(z_\alpha^2\right)}{q g d_{\gamma T} G_\delta'(z_\alpha^2)} \right)^{q/q+1} T^{q^2/q+1}. \tag{39}$$

When $d_{\gamma T} > 0$, the constraint $e_I^{\rho} \le \kappa\alpha$ may be binding, and we have to use the Kuhn-Tucker theorem to search for the optimum. Let λ be the Lagrange multiplier, and define

$$\begin{aligned} L(\rho,\lambda) &= G_\delta(z_\alpha^2) + c_q z_\alpha^4 K_\delta\left(z_\alpha^2\right)\rho^{-1/q} - g d_{\gamma T} G_\delta'(z_\alpha^2) z_\alpha^2 \rho T^{-q} \\ &\quad + \lambda\left(\left(\alpha + g d_{\gamma T} D'(z_\alpha^2) z_\alpha^2 \rho T^{-q}\right) - \kappa\alpha\right). \end{aligned} \tag{40}$$

It is easy to show that at the optimal ρ, the constraint $e_I^{\rho} \le \kappa\alpha$ is indeed binding and $\lambda > 0$. Hence, the optimal ρ is

$$\rho_{\text{opt}} = \frac{(\kappa - 1)\alpha}{g d_{\gamma T} D'(z_\alpha^2) z_\alpha^2} T^q, \tag{41}$$

and the corresponding Lagrange multiplier is

$$\lambda_{\text{opt}} = \frac{G_\delta'(z_\alpha^2)}{D'(z_\alpha^2)} + \frac{c_q K_\delta\left(z_\alpha^2\right)\left[g d_{\gamma T} D'(z_\alpha^2)\right]^{\frac{1}{q}}}{q\left[(\kappa-1)\alpha\right]^{1+1/q} T} \left(z_\alpha^2\right)^{1/q+2} .$$

Formulas (39) and (41) can be written collectively in the form

$$\rho_{\text{opt}} = \left(\frac{c_q z_\alpha^2 K_\delta \left(z_\alpha^2\right)}{q g d_{\gamma T} \left[\lambda_{\text{opt}} D'(z_\alpha^2) - G_\delta'(z_\alpha^2)\right]} \right)^{q/q+1} T^{q^2/q+1},$$

where

$$\lambda_{\text{opt}} = \begin{cases} 0, & \text{if } d_{\gamma T} < 0 \\ \frac{G_\delta'(z_\alpha^2)}{D'(z_\alpha^2)} + \frac{c_q K_\delta(z_\alpha^2)\left[g d_{\gamma T} D'(z_\alpha^2)\right]^{1/q}}{q[(\kappa-1)\alpha]^{1+1/q} T} \left(z_\alpha^2\right)^{1/q+2}, & \text{if } d_{\gamma T} > 0\,. \end{cases} \tag{42}$$

If we ignore the nonessential constant, then the function $L(\rho, \lambda)$ is a weighted sum of the type I and type II errors with the weight given by the optimal Lagrange multiplier. When $d_{\gamma T} < 0$, the type I error is expected to be capped by the nominal type I error. As a result, the optimal Lagrange multiplier is zero, and we assign all weight to the type II error. This weighting scheme might be justified by the argument that it is worthwhile to take advantage of the extra reduction in the type II error without inflating the type I error. When $d_{\gamma T} > 0$, the type I error is expected to be larger than the nominal type I error. The constraint on the type I error is binding, and the Lagrange multiplier is positive. In this case the loss function is a genuine weighted sum of type I and type II errors. As the tolerance parameter κ decreases toward 1, the weight attached to the type I error increases.

When the nonparametric bias, as measured by $-d_{\gamma T}$, is positive, the optimal ρ grows with T at the rate of $T^{q^2/(q+1)}$, which is slower than $T^{2q^2/(2q+1)}$, the MSE-optimal expansion rate. When the nonparametric bias is negative, which is typical for economic time series, the expansion rate of the optimal ρ depends on the rate at which κ approaches 1. When $\kappa \to 1$ at a rate faster than T^{-q}, the Lagrange multiplier λ_{opt} increases with the sample size at a rate faster than T^q. In this case, the optimal ρ is bounded. Fixed ρ rules may then be interpreted as assigning increasingly larger weight to the type I error. This gives us a practical interpretation of fixed ρ rules in terms of the permitted tolerance of the type I error. When $\kappa \to 1$ at a rate slower than $T^{-q/q+1}$, the Lagrange multiplier λ_{opt} is bounded and the optimal ρ expands with T at a rate faster than $T^{q^2/(q+1)}$. In particular, when $\kappa \to 1$ at the rate of $T^{-q/2q+1}$, the optimal ρ expands with T at the MSE-optimal rate $T^{2q^2/(2q+1)}$. So when $\kappa \to 1$ at a rate faster than $T^{-q/2q+1}$, the optimal ρ has a smaller order of magnitude than the MSE-optimal ρ regardless of the direction of the nonparametric bias.

When $\rho_{\text{opt}} = O\left(\left(T/\sqrt[q]{\lambda_{\text{opt}}}\right)^{q^2/(q+1)}\right)$, the size distortion of the t_ρ^*-test is of order $O((T\lambda_{\text{opt}})^{-q/(q+1)})$. It is apparent that the size distortion becomes smaller if a larger weight is implicitly assigned to the type I error. In particular, when λ_{opt} is finite, the size distortion is of order $O(T^{-q/(q+1)})$, which is larger than $O\left(T^{-q}\right)$, the size distortion for the case $\lambda_{\text{opt}} \sim T^q$. It is obvious that the use of ρ_{opt} involves some trade-off between two elements in the loss function (40). Note that even when λ_{opt} is finite, the size distortion is smaller than $O(T^{-q/(2q+1)})$, which is the size distortion for the conventional t-test using the AMSE-optimal ρ, that is when ρ is set to be $O(T^{2q^2/(2q+1)})$.

The formula for ρ_{opt} involves the unknown parameter $d_{\gamma T}$, which could be estimated nonparametrically or by a standard plug-in procedure based on a simple model like an AR(1) (see Andrews, 1991; Newey and West, 1994). 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 test-optimal ρ that maximizes the local asymptotic power while preserving size in large samples is fundamentally different from the MSE-optimal ρ. The test-optimal ρ depends on the sign of the nonparametric bias and the permitted tolerance for the type I error, while the MSE-optimal ρ does not. When the permitted tolerance becomes sufficiently small, the test-optimal ρ is of smaller order than the MSE-optimal ρ.

7. SIMULATION EVIDENCE

This section presents some simulation evidence on the performance of the t^*-test based on the plug-in implementation of the test-optimal power parameter. We consider the simple location model with Gaussian autoregressive moving average (ARMA)(1,1) errors:

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

where

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

Note that the long-run variance of u_t is $(1+\theta)^2/(1-\phi)^2$. On the basis of the long-run variance, we consider the following c values: $c = \delta_0 (1+\theta)/(1-\phi)$ for $\delta_0 \in [0,5]$. We consider three sets of parameter configurations for ϕ and θ :

$$AR(1) : (\phi,\theta) = (0.9,0), (0.6,0), (0.3,0), (-0.3,0), (-0.6,0), (-0.9,0),$$
$$MA(1) : (\phi,\theta) = (0,0.9), (0,0.6), (0,0.3), (0,-0.3), (0,-0.6), (0,-0.9),$$
$$ARMA(1,1) : (\phi,\theta) = (-0.6,0.3), (0.3,-0.6), (0.3,0.3), (0,0).(0.6,-0.3), (-0.3,0.6),$$

and write $u_t \sim ARMA[\phi,\theta]$.

We consider three sample sizes $T = 100$, 200, and 500. 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 $\hat{d} = 2\hat{\phi}/(1-\hat{\phi})^2$ when $q = 2$ and $\hat{d} = 2\hat{\phi}/(1-\hat{\phi}^2)$ when $q = 1$. We consider the tolerance factors $\kappa = 1.1$ and 1.2. We set the significance level to be $\alpha = 10\%$ and the corresponding nominal critical value for the two-sided test is $z_\alpha = 1.645$. To compute the test-optimal ρ, we need to choose δ for the local alternative hypothesis. In principle, we can estimate δ_0 by OLS, but the OLS estimator is not consistent. Accordingly, we follow standard practice in the optimal testing literature and propose to select δ such that the first-order power is 75%, that is, δ solves $1 - G_\delta(1.645) = 75\%$. The solution is $\delta = 2.3192$, which lies in the intermediate

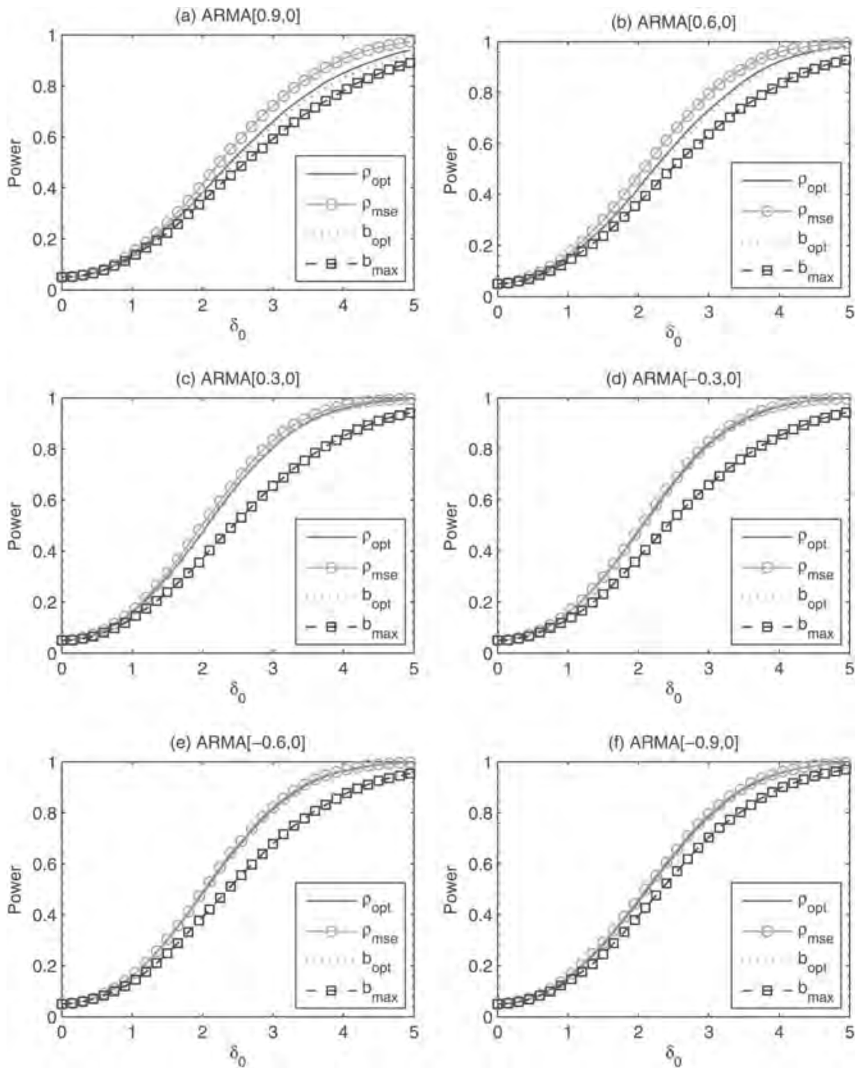

FIGURE 1. Size-adjusted power for different testing procedures under AR(1) errors with Bartlett kernel.

range of alternative hypotheses presented in Figures 1–5. Although this choice of δ may not match the true δ_0 under the local alternative, Monte Carlo experiments show that it delivers a test with good size and power properties.

For each choice of κ, we obtain $\hat{\rho}_{\text{opt}}$ and use it to construct the LRV estimate and corresponding t_ρ^*-statistic. We reject the null hypothesis if $\left|t_\rho^*\right|$ is larger than the corrected critical values given in (26). Using 10,000 replications, we compute the empirical type I error (when $\delta_0 = 0$ and $c = 0$). For comparative purposes, we

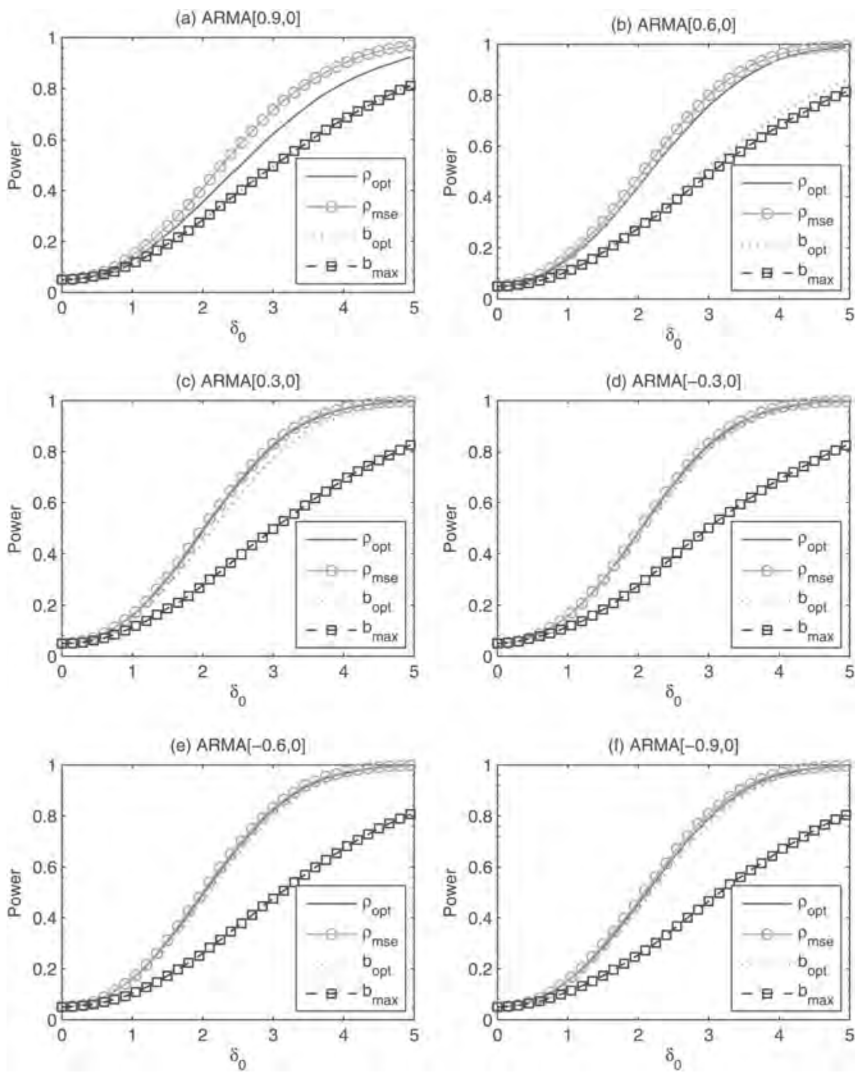

FIGURE 2. Size-adjusted power for different testing procedures under AR(1) errors with Parzen kernel.

also compute the empirical type I error when the power parameter is the ‘optimal’ one that minimizes the asymptotic MSE of the LRV estimate. The formulas for this power parameter are given in PSJ (2006, 2007), and plug-in versions are

$$\hat{\rho}_{\mathrm{MSE}} = \frac{1}{g}\left(\frac{\sqrt{2\pi}}{16}\frac{\left(1-\hat{\phi}\right)^4}{\hat{\phi}^2}\right)^{2/5} T^{8/5} \tag{43}$$

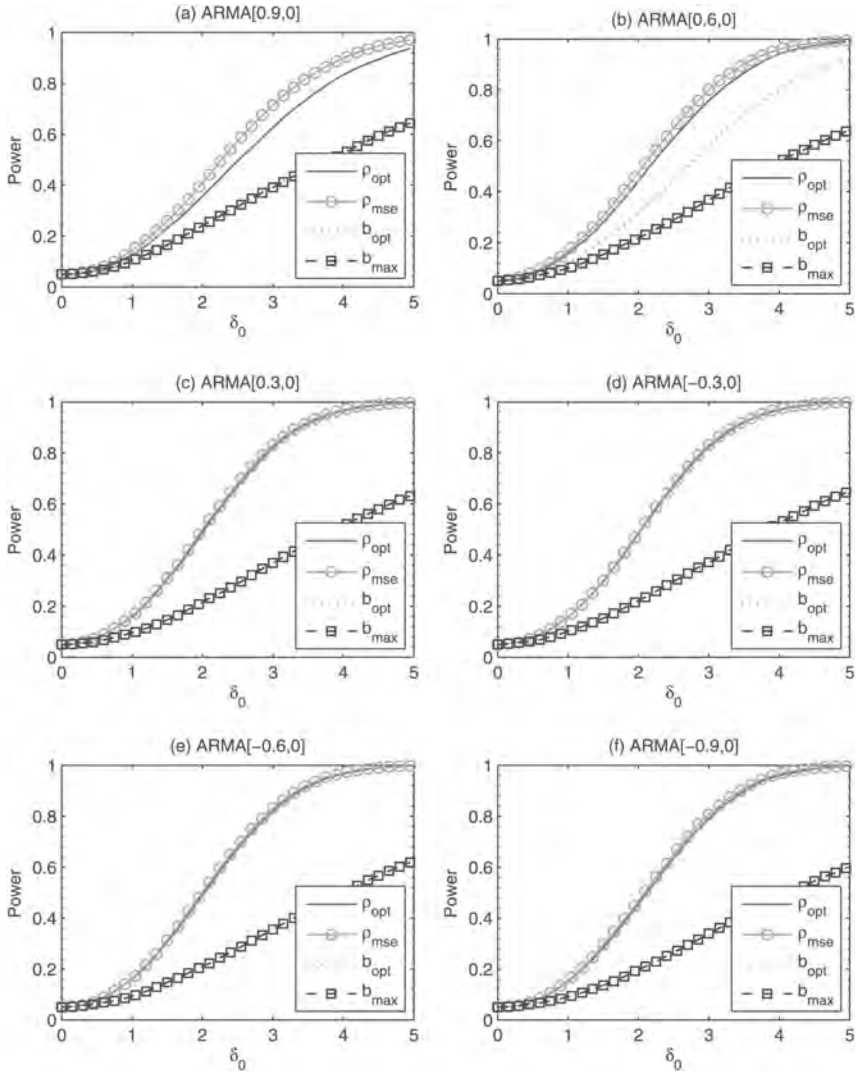

FIGURE 3. Size-adjusted power for different testing procedures under AR(1) errors with QS kernel.

for the second-order power kernels and

$$\hat{\rho}_{\mathrm{MSE}} = \left[\left(\frac{(1-\hat{\phi}^2)^2}{4\hat{\phi}^2} \right)^{1/3} T^{2/3} \right] \tag{44}$$

for the first-order power kernels.

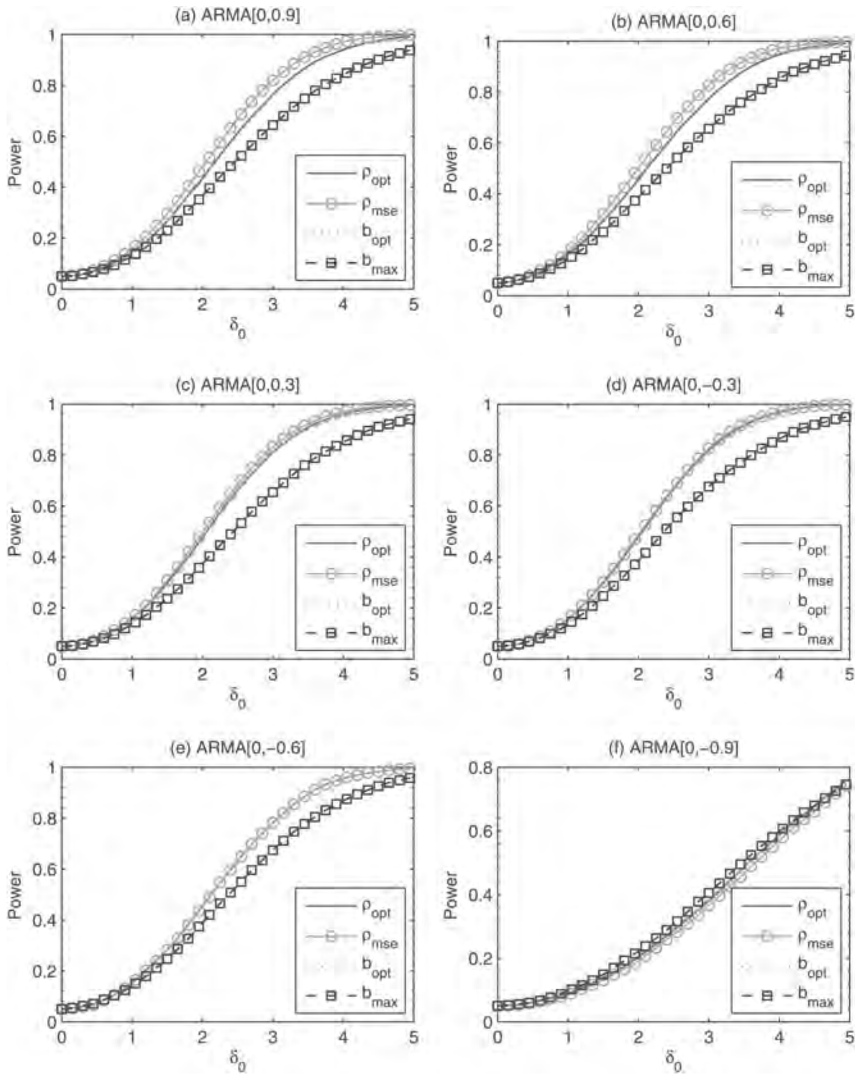

FIGURE 4. Size-adjusted power for different testing procedures under MA(1) errors with Bartlett kernel.

To compare the exponentiated kernel approach with the contracted kernel approach, we include the t^{*}-test based on contracted kernels. It is easy to show that the optimal b that minimizes e_{II}^{b} subject to the constraint that $e_{I}^{b} \leq \kappa\alpha$ is given by

$$b_{\text{opt}} = \left(\frac{\mu_2 z_\alpha^2 K_\delta\left(z_\alpha^2\right)}{q g d_{\gamma T}\left[\tilde{\lambda}_{\text{opt}} D'(z_\alpha^2) - G'_\delta(z_\alpha^2)\right]} \right)^{-1/q+1} T^{-q/q+1}, \tag{45}$$

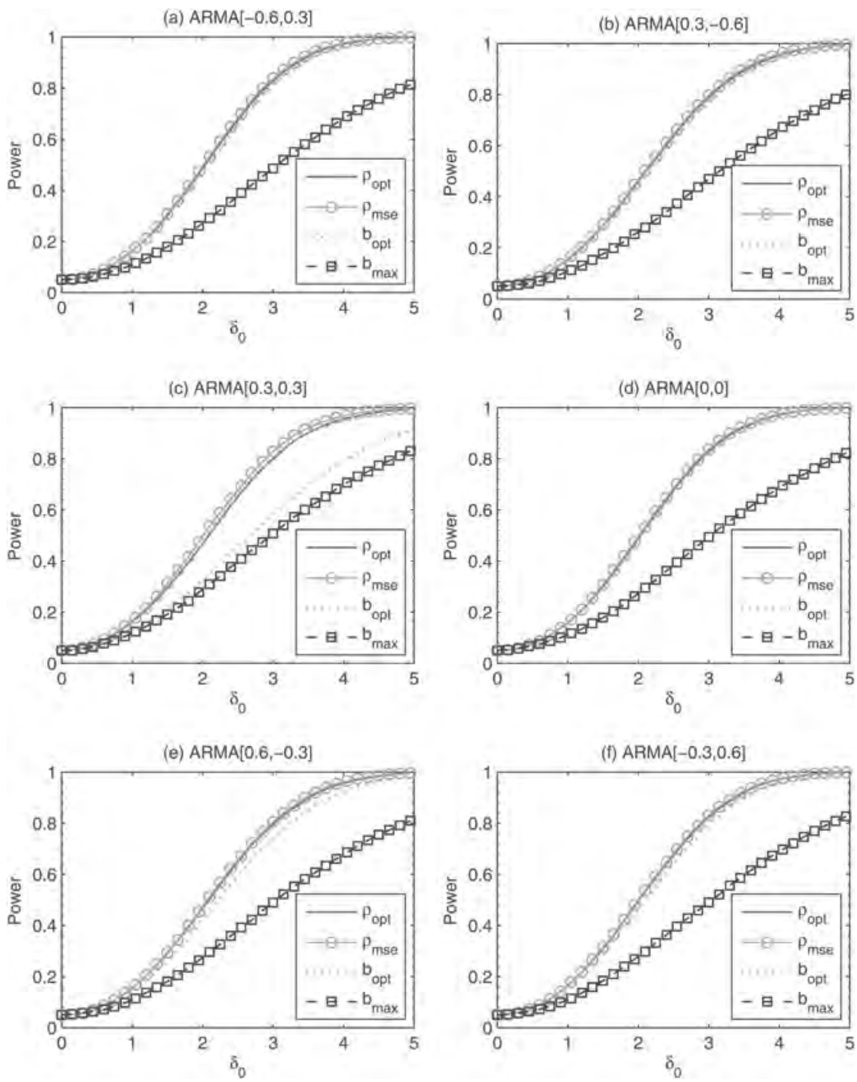

FIGURE 5. Size-adjusted power for different testing procedures under ARMA(1,1) errors with Parzen kernel.

where

$$\tilde{\lambda}_{\text{opt}} = \begin{cases} 0, & \text{if } d_{\gamma T} < 0 \\ \frac{G'_{\delta}(z_{\alpha}^2)}{D'(z_{\alpha}^2)} + \frac{\mu_2 K_{\delta}(z_{\alpha}^2)\left[g d_{\gamma T} D'(z_{\alpha}^2)\right]^{1/q}}{q[(\kappa-1)\alpha]^{1+1/q} T} \left(z_{\alpha}^2\right)^{1/q+2}, & \text{if } d_{\gamma T} > 0. \end{cases} \tag{46}$$

We implement b_{opt} using the same plug-in procedure for ρ_{opt}. In addition, we include the nonstandard test proposed by KV (2002b), which sets b to be 1, the maximum value of b (or equivalently, sets ρ to be 1), and uses nonstandard critical values. We refer to the four testing procedures as $t^*_{\rho_{\text{opt}}}$, $t_{\rho_{\text{mse}}}$, $t^*_{b_{\text{opt}}}$, and $t^*_{b_{\max}}$, respectively.

Tables 1–3 report the empirical type I error and average values of the selected smoothing parameters for the ARMA(1,1) error with sample size $T = 100$, tolerance parameter $\kappa = 1.1$, and significance level $\alpha = 10\%$. Several patterns emerge for the type I error comparison. First, when the empirical type I error is around or greater than the nominal type I error, the testing-optimal plug-in procedure $t^*_{\rho_{\text{opt}}}$ incurs a significantly smaller type I error than the conventional plug-in procedure $t_{\rho_{\text{mse}}}$. In other cases, the type I errors are more or less the same for the two plug-in procedures. Second, compared with the $t^*_{b_{\max}}$-test, the $t^*_{\rho_{\text{opt}}}$-test has similar size distortion except when the error process is highly persistent and second-order

TABLE 1. Finite sample sizes and smoothing parameters of different testing procedures under AR(1) errors $(T = 100, \kappa = 1.1, \alpha = 0.10)$

$[\phi,\theta]$	[0.9,0]	[0.6,0]	[0.3,0]	[-0.3,0]	[-0.6,0]	[-0.9,0]
			Bartlett			
ρ_{opt}	0.1375	0.1132	0.1078	0.0786	0.0626	0.0252
	(1.42)	(4.13)	(13.52)	(18.99)	(10.81)	(5.46)
ρ_{mse}	0.3574	0.2103	0.1554	0.0909	0.0899	0.0863
	(6.84)	(15.93)	(34.76)	(32.14)	(15.17)	(6.06)
b_{opt}	0.2189	0.1373	0.1115	0.081	0.0701	0.0398
	(0.9075)	(0.3007)	(0.1070)	(0.0588)	(0.0990)	(0.2114)
$b_{\max}$	0.1939	0.1214	0.1023	0.0908	0.0852	0.0456
			Parzen			
ρ_{opt}	0.2081	0.1264	0.1073	0.0924	0.0921	0.0948
	(2.06)	(18.67)	(146.42)	(194.07)	(145.97)	(139.21)
ρ_{mse}	0.3221	0.181	0.1399	0.0906	0.0881	0.0882
	(8.04)	(55.88)	(281.87)	(555.34)	(406.20)	(383.65)
b_{opt}	0.2204	0.1717	0.1115	0.0992	0.0992	0.1016
	(1.0000)	(0.9280)	(0.2738)	(0.1168)	(0.1431)	(0.1500)
$b_{\max}$	0.1324	0.0996	0.0922	0.0923	0.0933	0.0877
			QS			
ρ_{opt}	0.204	0.1231	0.1071	0.092	0.0916	0.095
	(7.09)	(77.22)	(616.51)	(817.69)	(614.66)	(586.14)
ρ_{mse}	0.321	0.1807	0.14	0.0906	0.089	0.0903
	(32.33)	(234.30)	(1188.36)	(2342.92)	(1713.24)	(1618.06)
b_{opt}	0.2978	0.1608	0.1098	0.0896	0.0902	0.0953
	(0.9920)	(0.3743)	(0.0649)	(0.0296)	(0.0363)	(0.0380)
$b_{\max}$	0.1282	0.0999	0.0939	0.0945	0.0951	0.0914

Note: Average values of the selected smoothing parameter are in parentheses.
ρ_{opt}: t^*-test with exponentiated kernel, test-optimal rho, and second-order corrected critical values (CV).
ρ_{mse}: t-test with exponentiated kernel, MSE-optimal ρ, and standard normal CV.
b_{opt} : t^*-test with contracted kernel, test-optimal b, and second-order corrected CV.
$b_{\max}$: t^*-test with $b = 1$ or $\rho = 1$, and nonstandard critical values from KV (2002b).

TABLE 2. Finite sample sizes and smoothing parameters of different testing procedures under MA(1) errors $(T = 100, \kappa = 1.1, \alpha = 0.10)$

$[\phi,\theta]$	[0,0.9]	[0,0.6]	[0,0.3]	[0,-0.3]	[0,-0.6]	[0,-0.9]
			Bartlett			
ρ_{opt}	0.0975	0.1005	0.1049	0.0654	0.0179	0.0000
	(5.43)	(6.54)	(15.10)	(19.73)	(14.04)	(12.82)
ρ_{mse}	0.1507	0.1531	0.1389	0.071	0.0184	0.0000
	(19.51)	(22.20)	(38.22)	(34.27)	(21.59)	(19.08)
b_{opt}	0.1104	0.1113	0.1100	0.0679	0.0215	0.0000
	(0.2150)	(0.1782)	(0.0958)	(0.0555)	(0.0752)	(0.0824)
b_{max}	0.1044	0.1052	0.1021	0.0820	0.0478	0.0000
			Parzen			
ρ_{opt}	0.1079	0.1063	0.1030	0.0787	0.0392	0.0000
	(31.37)	(44.15)	(171.67)	(195.92)	(157.05)	(151.95)
ρ_{mse}	0.1326	0.136	0.1245	0.0672	0.0124	0.0000
	(85.94)	(112.99)	(345.02)	(572.76)	(443.68)	(426.30)
b_{opt}	0.1466	0.1394	0.1085	0.0891	0.0684	0.0014
	(0.7546)	(0.5867)	(0.2304)	(0.1127)	(0.1333)	(0.1376)
b_{max}	0.0908	0.0946	0.0960	0.0879	0.0792	0.0183
			QS			
ρ_{opt}	0.1061	0.1050	0.1024	0.0783	0.0387	0.0000
	(130.82)	(184.78)	(723.13)	(825.48)	(661.43)	(639.87)
ρ_{mse}	0.1315	0.1358	0.1244	0.0668	0.0120	0.0000
	(361.18)	(475.40)	(1454.95)	(2416.40)	(1871.46)	(1798.09)
b_{opt}	0.1182	0.1141	0.1044	0.0710	0.0299	0.0000
	(0.1997)	(0.1438)	(0.0546)	(0.0286)	(0.0338)	(0.0349)
b_{max}	0.0935	0.0948	0.0935	0.0891	0.0792	0.0218

Note: Average values of the selected smoothing parameter are in parentheses.
ρ_{opt}: t^*-test with exponentiated kernel, test-optimal rho, and second-order corrected CV.
ρ_{mse}: t-test with exponentiated kernel, MSE-optimal ρ, and standard normal CV.
b_{opt} : t^*-test with contracted kernel, test-optimal b, and second-order corrected CV.
b_{max}: t^*-test with $b = 1$ or $\rho = 1$, and nonstandard critical values from KV (2002b).

kernels are used. Since the bandwidth is set equal to the sample size, the $t^*_{b_{\text{max}}}$-test is designed to achieve the smallest possible size distortion. Given this observation, we can conclude that the $t^*_{\rho_{\text{opt}}}$-test succeeds in controlling for the type I error. Third, compared with the $t^*_{b_{\text{opt}}}$-test, the $t^*_{\rho_{\text{opt}}}$-test has smaller type I error for most of the scenarios considered. Finally, the $t^*_{\rho_{\text{opt}}}$-test based on the Bartlett kernel has the smallest size distortion in an overall sense. This is true even when the error autocorrelation is very high.

Due to the approximation error, the bound we impose on the approximate type I error does not fully control the empirical type I error. This is demonstrated in Tables 1–3. The quality of approximation depends on the persistence of the time series. When the time series is highly persistent, the first-order asymptotic bias of the LRV estimator may not approximate the finite sample bias very well. As a result, the approximate type I error, which is based on the first-order asymptotic

TABLE 3. Finite sample sizes and smoothing parameters of different testing procedures under ARMA(1,1) errors $(T = 100, \kappa = 1.1, \alpha = 0.10)$

$[\phi, \theta]$	[-0.6,0.3]	[0.3,-0.6]	[0.3,0.3]	[0,0]	[0.6,-0.3]	[-0.3,0.6]
			Bartlett			
ρ_{opt}	0.0847	0.0382	0.0996	0.0991	0.1315	0.1015
	(17.34)	(23.63)	(5.21)	(121.72)	(12.04)	(20.32)
ρ_{mse}	0.1016	0.0356	0.1658	0.1093	0.201	0.1295
	(28.70)	(40.12)	(18.89)	(145.63)	(31.57)	(43.86)
b_{opt}	0.0877	0.0379	0.1134	0.1008	0.1365	0.1032
	(0.0634)	(0.0511)	(0.2278)	(0.0271)	(0.1230)	(0.0831)
$b_{\max}$	0.0940	0.0716	0.1064	0.0979	0.1128	0.0996
			Parzen			
ρ_{opt}	0.0973	0.0474	0.1054	0.0999	0.1392	0.0997
	(179.89)	(245.63)	(29.03)	(1849.49)	(124.82)	(254.02)
ρ_{mse}	0.1007	0.0302	0.1435	0.1078	0.1864	0.1155
	(513.77)	(656.61)	(80.49)	(2599.72)	(236.70)	(429.20)
b_{opt}	0.1009	0.0675	0.1558	0.1020	0.1276	0.1032
	(0.1217)	(0.1062)	(0.7950)	(0.0521)	(0.3421)	(0.1876)
$b_{\max}$	0.09	0.0817	0.0948	0.0942	0.0975	0.092
			QS			
ρ_{opt}	0.0973	0.0474	0.1043	0.0999	0.1388	0.0992
	(757.84)	(1035.35)	(120.95)	(7806.44)	(525.37)	(1070.78)
ρ_{mse}	0.1006	0.0299	0.143	0.1079	0.1872	0.1154
	(2167.40)	(2770.40)	(338.19)	(10973.60)	(997.69)	(1810.35)
b_{opt}	0.0957	0.0358	0.1213	0.1048	0.1377	0.0987
	(0.0309)	(0.0269)	(0.2223)	(0.0129)	(0.0820)	(0.0444)
$b_{\max}$	0.0948	0.087	0.093	0.0968	0.0963	0.0925

Note: Average values of the selected smoothing parameter are in parentheses.
ρ_{opt}: t^*-test with exponentiated kernel, test-optimal rho, and second-order corrected CV.
ρ_{mse}: t-test with exponentiated kernel, MSE-optimal ρ, and standard normal CV.
b_{opt} : t^*-test with contracted kernel, test-optimal b, and second-order corrected CV.
$b_{\max}$: t^*-test with $b = 1$ or $\rho = 1$, and nonstandard critical values from KV (2002b).

bias, may not fully capture the empirical type I error. So it is important to keep in mind that the empirical type I error may still be larger than the nominal type I error even if we exert some control over the approximate type I error.

Tables 1–3 show that the optimal ρ decreases as the time series becomes more persistent. This is true for both the test-based criterion and the MSE-based criterion. Comparing $\hat{\rho}_{\text{opt}}$ with $\hat{\rho}_{\text{mse}}$, we find that $\hat{\rho}_{\text{mse}}$ is larger than $\hat{\rho}_{\text{opt}}$. This is consistent with our theory. For hypothesis testing, it is desirable to employ undersmoothing to achieve bias reduction, especially when the time series is persistent. Similarly, the test-optimal b increases as the series correlation becomes stronger.

The above qualitative observations remain valid for other configurations such as different sample sizes and different values of κ. All else being equal, the size distortion of the $t^*_{\rho_{\text{opt}}}$-test and the $t^*_{b_{\text{opt}}}$-test for $\kappa = 1.2$ is slightly larger than that for $\kappa = 1.1$. This is expected, as we have a higher tolerance for the type I error when the value of κ is larger.

Figures 1–3 present finite sample power under AR(1) errors for the three kernels considered. We compute power using the 10% empirical finite sample critical values obtained from the null distribution. So the finite sample power is size-adjusted and power comparisons are meaningful. The parameter configuration is the same as those for Tables 1–3 except that the data generating process is generated under local alternatives. Three observations can be drawn from these figures. First, the $t^*_{\rho_{\text{opt}}}$-test with test-optimal exponent is as powerful as the conventional MSE-based $t_{\rho_{\text{mse}}}$-test. For a given power parameter ρ, using nonstandard critical values or second-order corrected critical values improves the size accuracy of the test but at the cost of clear power reduction. Figures 1–3 show that we can employ the test-optimal ρ to compensate for the power loss. Second, the power of the $t^*_{\rho_{\text{opt}}}$-test is consistently higher than that of the $t^*_{b_{\text{max}}}$-test. The power difference is larger for the second-order kernels than for first-order kernels. Third, the power of the $t^*_{\rho_{\text{opt}}}$-test is either close to or substantially higher than that of the $t^*_{b_{\text{opt}}}$-test. The superior performance of the $t^*_{\rho_{\text{opt}}}$-test occurs when the process is very persistent, e.g., AR(0.6), AR(0.9), and ARMA(0.3,0.3). In those cases, the optimal b is close to 1 on average, and the power of the $t^*_{b_{\text{opt}}}$-test becomes close to that of the $t^*_{b_{\text{max}}}$-test. But the optimal ρ value is larger than 1 on average, leading to higher power than the $t^*_{b_{\text{max}}}$-test, which is the same as the $t^*_{\rho_{\text{min}}}$-test for $\rho_{\text{min}} = 1$. To sum up, the $t^*_{\rho_{\text{opt}}}$-test achieves the same degree of size accuracy as the nonstandard $t^*_{b_{\text{max}}}$-test and yet maintains the power of the conventional $t_{\rho_{\text{mse}}}$-test. The $t^*_{\rho_{\text{opt}}}$-test is as accurate in size as the $t^*_{b_{\text{opt}}}$-test and is more powerful than the latter test when the time series is highly persistent.

Rather than reporting all of the remaining figures for other configurations, we present two representative figures. Figure 4 presents the power curves under the MA(1) error and with the Bartlett kernel, and Figure 5 presents the power curves under the ARMA(1,1) error and with the Parzen kernel. The basic qualitative observations remain the same: The $t^*_{\rho_{\text{opt}}}$-test is as powerful as the standard $t_{\rho_{\text{mse}}}$-test and much more powerful than the nonstandard $t^*_{b_{\text{max}}}$-test.

8. CONCLUSION

Pursuing the same the line of research taken in SPJ (2008) to improve econometric testing where there is nonparametric studentization, the present paper employs the exponentiated kernel approach of PSJ (2006, 2007) and proposes an exponent choice that maximizes the local asymptotic power while controlling for the asymptotic size of the test. This new selection criterion is fundamentally different from the MSE criterion for the point estimation of the long-run variance. When the permitted tolerance on the type I error is low, the expansion rate of the test-optimal exponent is smaller than the MSE-optimal exponent. The fixed exponent rule can be interpreted as exerting increasingly tight control on the type I error.

Monte Carlo experiments show that, when the time series is not highly persistent, the size of the new t^*-test is as accurate as the nonstandard KV test with the bandwidth equal to the sample size. On the other hand, unlike the KV test, which is generally less powerful than the standard t-test, the t^*-test based on the test-optimal power parameter is as powerful as the standard t-test. In addition, Monte Carlo experiments show that the exponentiated kernel approach is at least as competitive as the conventional contracted kernel approach. In some scenarios that are not atypical for economic time series, the exponentiated kernel approach outperforms the contracted kernel approach.

REFERENCES

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–854.

Andrews, D.W.K. & J.C. Monahan (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.

de Jong, R.M. & J. Davidson (2000) Consistency of kernel estimators of heteroskedastic and autocorrelated covariance matrices. *Econometrica* 68, 407–424.

den Haan, W.J. & A. Levin (1997) A practitioner's guide to robust covariance matrix estimation. In G. Maddala and C. Rao (eds.), *Handbook of Statistics: Robust Inference*, vol. 15. Elsevier.

Hansen, B.E. (1992) Consistent covariance matrix estimation for dependent heterogenous processes. *Econometrica* 60, 967–972.

Hong, Y. & J. Lee (2001) Wavelet-Based Estimation for Heteroskedastic and Autocorrelation Consistent Variance-covariance Matrices. Working paper, Cornell University.

Jansson, M. (2002) Consistent covariance matrix estimation for linear processes. *Econometric Theory* 18, 1449–1459.

Jansson, M. (2004) On the error of rejection probability in simple autocorrelation robust tests. *Econometrica* 72, 937–946.

Kiefer, N.M. & T.J. Vogelsang (2002a) Heteroskedasticity-autocorrelation robust standard errors using the bartlett kernel without truncation. *Econometrica* 70, 2093–2095.

Kiefer, N.M. & T.J. Vogelsang (2002b) Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory* 18, 1350–1366.

Kiefer, N.M. & T.J. Vogelsang (2005) A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* 21, 1130–1164.

Kiefer, N.M., T.J. Vogelsang, & H. Bunzel (2000) Simple robust testing of regression hypotheses. *Econometrica* 68, 695–714.

Newey, W.K. & K.D. West (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Newey, W.K. & K.D. West (1994) Automatic lag selection in covariance estimation. *Review of Economic Studies* 61, 631–654.

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.

Phillips, P.C.B. (2005a) Automated discovery in econometrics. *Econometric Theory* 21, 3–20.

Phillips, P.C.B. (2005b) HAC estimation by automated regression. *Econometric Theory* 21, 116–142.

Phillips, P.C.B., Y. Sun, & S. Jin (2006) Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. *International Economic Review* 21, 837–894.

Phillips, P.C.B., Y. Sun, & S. Jin (2007) Long-run variance estimation and robust regression using sharp origin kernels with no truncation. *Journal of Statistical Planning and Inference* 137, 985–1023.

Ray. S. & N.E. Savin (2008) The performance of heteroskedasticity and autocorrelation robust tests: A Monte Carlo study with an application to the three-factor Fama-French asset-pricing model. *Journal of Applied Econometrics* 23, 91–109.

Ray, S., N.E. Savin, & I.A. Tiwari (2009) Testing the CAPM revisited. *Journal of Empirical Finance* 16, 721–733.

Sul, D., P.C.B. Phillips, & C-Y. Choi (2005) Prewhitening bias in HAC estimation. *Oxford Bulletin of Economics and Statistics* 67, 517–546.

Sun, Y. (2004) Estimation of the long-run average relationship in nonstationary panel time series. *Econometric Theory* 20, 1227–1260

Sun, Y. (2009) Robust Trend Inference with Series Variance Estimator and Testing-Optimal Smoothing Parameter. Working paper, UC San Diego.

Sun, Y., P.C.B. Phillips, & S. Jin (2008) Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica* 76, 175–194.

Velasco, C. & P.M. Robinson (2001) Edgeworth expansions for spectral density estimates and studentized sample mean. *Econometric Theory* 17, 497–539.

Vogelsang, T.J. (2003) Testing in GMM models without truncation. In T.B. Fomby & R.C. Hill (eds.), *Advances in Econometrics*, vol. 17, Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later. pp. 199–233. Elsevier Science.

APPENDIX

A.1. Technical Lemmas and Supplements.

LEMMA A.1. *For quadratic power kernels, as $\rho \to \infty$, we have*

(a) $\int_0^1 k_\rho(x)\,dx = O\left(\frac{1}{\sqrt{\rho}}\right)$,

(b) $\int_0^1 (1-x)k_\rho(x)\,dx = \frac{1}{2}\left(\frac{\pi}{\rho g}\right)^{1/2} + O\left(\frac{1}{\rho}\right)$.

Proof of Lemma A.1. We prove part (b) only, as part (a) follows from similar but easier arguments. For both the Parzen and QS kernels, we have for any $\tau > 0$, there exists $\zeta := \zeta(\tau) > 0$ such that $\log k(x) \le -\zeta(\tau)$ for $\tau \le x \le 1$. Therefore, the contribution of the interval $\tau \le x \le 1$ satisfies

$$\int_\tau^1 (1-x)k_\rho(x)\,dx = \int_\tau^1 \exp\{\rho \log k(x) + \log(1-x)\}\,dx$$
$$\le \exp[-(\rho-1)\zeta(\tau)]\int_0^1 k(x)\,dx \le \exp[-(\rho-1)\zeta(\tau)]. \qquad \textbf{(A.1)}$$

We now deal with the integral from $-\tau$ to τ. Both the Parzen and QS kernels exhibit quadratic behavior around the origin in the sense that

$$k(x) = 1 - gx^2 + o\left(x^2\right), \qquad \text{as } x \to 0 \text{ for some } g > 0,$$

which implies $\log k(x) = -gx^2 + o(x^2)$. So, for any given $\varepsilon > 0$, we can determine $\tau > 0$ such that

$$\left|\log k(x) + gx^2\right| \le \varepsilon x^2, \qquad \text{for } 0 \le x \le \tau.$$

In consequence,

$$\int_0^\tau (1-x)\exp\left[-\rho(g+\varepsilon)x^2\right]dx \le \int_0^\tau (1-x)k_\rho(x)\,dx \le (1-x)\int_0^\tau \exp\left[-\rho(g-\varepsilon)x^2\right]dx.$$

Note that

$$\int_0^{\tau} (1-x)\exp\left[-\rho(g+\varepsilon)x^2\right]dx = \int_0^{\infty} (1-x)\exp\left[-\rho(g+\varepsilon)x^2\right] dx - \int_{\tau}^{\infty} (1-x)\exp\left[-\rho(g+\varepsilon)x^2\right].$$

In view of

$$\int_0^{\infty} (1-x)\exp\left[-\rho(g+\varepsilon)x^2\right]dx = \frac{1}{2}\frac{\sqrt{\pi}}{\sqrt{\rho(g+\varepsilon)}} - \frac{1}{2}\frac{1}{\rho(g+\varepsilon)},$$

and

$$\begin{aligned}\left|\int_{\tau}^{\infty} (1-x)\exp\left[-\rho(g+\varepsilon)x^2\right]dx\right| &\leq \int_{\tau}^{\infty} \exp\left[-(\rho-1)(g+\varepsilon)x^2\right]\exp\left[-(g+\varepsilon)x^2\right](|1-x|)\,dx \\ &\leq \exp\left[-(\rho-1)(g+\varepsilon)\tau^2\right]\int_0^{\infty} |1-x|\exp\left[-(g+\varepsilon)x^2\right]dx \\ &= O\left(\exp\left[-(\rho-1)g\tau^2\right]\right),\end{aligned}$$

we have

$$\int_0^{\tau} (1-x)\exp\left[-\rho(g+\varepsilon)x^2\right]dx = \frac{1}{2}\frac{\sqrt{\pi}}{\sqrt{\rho(g+\varepsilon)}} + O\left(\frac{1}{\rho}\right).$$

Similarly,

$$\int_0^{\tau} (1-x)\exp\left[-\rho(g-\varepsilon)x^2\right]dx = \frac{1}{2}\frac{\sqrt{\pi}}{\sqrt{\rho(g-\varepsilon)}} + O\left(\frac{1}{\rho}\right).$$

Therefore,

$$\int_0^{\tau} (1-x)k_{\rho}(x)\,dx = \frac{1}{2}\left(\frac{\pi}{\rho g}\right)^{1/2} + O\left(\frac{1}{\rho}\right). \tag{A.2}$$

Combining (A.1) and (A.2) yields

$$\int_0^{1} (1-x)k_{\rho}(x)\,dx = \frac{1}{2}\left(\frac{\pi}{\rho g}\right)^{1/2} + O\left(\frac{1}{\rho}\right). \tag{A.3}$$

■

LEMMA A.2. *The cumulants of* $\Xi_{\rho} - \mu_{\rho}$ *satisfy*

$$|\kappa_m| \leq 2^{3m-3}(m-1)!\left(\int_0^1 k_{\rho}(v)dv\right)^{m-1},$$

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

$$|\alpha_m| \leq 2^{4m-4}(m-1)!\left(\int_0^1 k_{\rho}(v)dv\right)^{m-1}.$$

Proof of Lemma A.2. Note that

$$
\begin{aligned}
&\left|\int_0^1 \cdots \int_0^1 \left(\prod_{j=1}^m k_\rho^*(\tau_j,\tau_{j+1})\right) d\tau_1\cdots d\tau_m\right| \\
&\quad \le \int_0^1 \cdots \int_0^1 \left|k_\rho^*(\tau_1,\tau_2)k_\rho^*(\tau_2,\tau_3)\cdots k_\rho^*(\tau_{m-1},\tau_m)\right| \left|k_\rho^*(\tau_m,\tau_1)\right| d\tau_1\cdots d\tau_m \\
&\quad \le \int_0^1 \cdots \int_0^1 \left|k_\rho^*(\tau_1,\tau_2)k_\rho^*(\tau_2,\tau_3)\cdots k_\rho^*(\tau_{m-1},\tau_m)\right| d\tau_1\cdots d\tau_m \\
&\quad \le \sup_{\tau_2}\int_0^1 \left|k_\rho^*(\tau_1,\tau_2)\right| d\tau_1 \int_0^1 \left|k_\rho^*(\tau_2,\tau_3)k_\rho^*(\tau_3,\tau_4)\cdots k_\rho^*(\tau_{m-1},\tau_m)\right| d\tau_2\cdots d\tau_m \\
&\quad \le \sup_{\tau_2}\int_0^1 \left|k_\rho^*(\tau_1,\tau_2)\right| d\tau_1 \sup_{\tau_3}\int_0^1 \left|k_\rho^*(\tau_2,\tau_3)\right| d\tau_2 \ldots \sup_{\tau_m}\int_0^1 \left[k_\rho^*(\tau_{m-1},\tau_m)\right] d\tau_{m-1} \\
&\quad = \left(\sup_s \int_0^1 \left|k_\rho^*(r,s)\right| dr\right)^{m-1}. \qquad \textbf{(A.4)}
\end{aligned}
$$

Let $\tau = \tau(r) > 0$ be such that $k_\rho(\tau) = \int_0^1 k_\rho(r-p)dp$. Then, in view of the definition

$$
k_\rho^*(r,s) = k_\rho(r-s) - \int_0^1 k_\rho(r-p)dp - \int_0^1 k_\rho(s-q)dq + \int_0^1\int_0^1 k_\rho(p-q)dpdq,
$$

we have

$$
\begin{aligned}
\sup_s \int_0^1 \left|k_\rho^*(r,s)\right| dr &\le 2\sup_s \int_0^1 \left|k_\rho(r-s) - \int_0^1 k_\rho(r-p)dp\right| dr \\
&= 2\sup_s \int_{|r-s|\le\tau} \left(k_\rho(r-s) - \int_0^1 k_\rho(r-p)dp\right) dr \\
&\quad + 2\sup_s \int_{|r-s|\ge\tau} \left(\int_0^1 k_\rho(r-p)dp - k_\rho(r-s)\right) dr \\
&\le 2\sup_s \int_{|r-s|\le\tau} k_\rho(r-s)dr + 2\sup_s \int_0^1 \left(\int_{|r-s|\ge\tau} k_\rho(r-p)dr\right) dp \\
&\le 2\sup_s\sup_p \int_{|r-s|\le\tau} k_\rho(r-p)dr + 2\sup_s\sup_p \int_{|r-s|\ge\tau} k_\rho(r-p)dr \\
&= 2\sup_p \int_0^1 k_\rho(r-p)dr.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\sup_s \int_0^1 \left|k_\rho^*(r,s)\right| dr &\le 2\sup_s \int_0^1 k_\rho(r-s)dr \\
&= 2\sup_{s\in[0,1]} \left(\int_{-s}^{1-s} k_\rho(v)dv\right) \le 4\int_0^1 k_\rho(v)dv.
\end{aligned}
$$

As a result,

$$
\left|\int_0^1 \cdots \int_0^1 \left(\prod_{j=1}^m k_\rho^*(\tau_j,\tau_{j+1})\right) d\tau_1\cdots d\tau_m\right| \le \left(4\int_0^1 k_\rho(v)dv\right)^{m-1},
$$

and

$$|\kappa_m| \leq 2^{3m-3}(m-1)!\left(\int_0^1 k_\rho(v)dv\right)^{m-1}. \tag{A.5}$$

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

$$\alpha_1 = \kappa_1, \qquad \alpha_m = \sum_{j=0}^{m-1}\binom{m-1}{j}\alpha_j\kappa_{m-j}. \tag{A.6}$$

It follows easily by induction from (A.5), (A.6), and the identity

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

that

$$|\alpha_m| \leq 2^{4m-4}(m-1)!\left(\int_0^1 k_\rho(v)dv\right)^{m-1}. \qquad \blacksquare$$

LEMMA A.3. *Let Assumption 2 hold. When $T \to \infty$ for a fixed ρ, we have*

(a)

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

(b)

$$\kappa_{m,T} = \kappa_m + O\left\{\frac{m!2^{m-1}}{T^2}\left(4\int_0^1 k_\rho(v)dv\right)^{m-2}\right\},$$

uniformly over m ≥ 1.

(c)

$$\alpha_{m,T} = \mathrm{E}\left(\varsigma_{\rho T} - \mu_{\rho T}\right)^m = \alpha_m + O\left\{\frac{2^{2m-1}m!}{T}\left(4\int_0^1 k_\rho(v)dv\right)^{m-2}\right\},$$

uniformly over m ≥ 1.

Proof of Lemma A.3. We first calculate $\mu_{\rho T} = \left(T\omega_T^2\right)^{-1}\mathrm{Trace}\left(\Omega_T A_T W_\rho A_T\right)$. Let $W_\rho^* = A_T W_\rho A_T$, then the (i,j)-th element of W_ρ^* is

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

So

$$\text{Trace}\left(\Omega_T A_T W_\rho A_T\right) = \text{Trace}(\Omega_T W_\rho^*)$$

$$= \sum_{1\le r_1,r_2\le T} \left\{\gamma\left(r_1-r_2\right)\tilde{k}_\rho\left(\frac{r_1}{T},\frac{r_2}{T}\right)\right\} = \sum_{r_2=1}^{T}\sum_{h_1=1-r_2}^{T-r_2}\gamma\left(h_1\right)\tilde{k}_\rho\left(\frac{r_2+h_1}{T},\frac{r_2}{T}\right)$$

$$= \left(\sum_{h_1=1}^{T-1}\sum_{r_2=1}^{T-h_1} + \sum_{h_1=1-T}^{0}\sum_{r_2=1-h_1}^{T}\right)\gamma\left(h_1\right)\tilde{k}_\rho\left(\frac{r_2+h_1}{T},\frac{r_2}{T}\right). \tag{A.7}$$

But

$$\sum_{r_2=1}^{T-h_1}\tilde{k}_\rho\left(\frac{r_2+h_1}{T},\frac{r_2}{T}\right) = \sum_{r_2=1}^{T-h_1}k_\rho\left(\frac{h_1}{T}\right) - \frac{1}{T}\sum_{r_1=1+h_1}^{T}\sum_{m=1}^{T}k_\rho\left(\frac{r_1-m}{T}\right)$$

$$-\frac{1}{T}\sum_{r_2=1}^{T-h_1}\sum_{k=1}^{T}k_\rho\left(\frac{k-r_2}{T}\right) + \sum_{r_2=1}^{T-h_1}\frac{1}{T^2}\sum_{k=1}^{T}\sum_{m=1}^{T}k_\rho\left(\frac{k-m}{T}\right)$$

$$= -\frac{1}{T}\sum_{r_1=1}^{T}\sum_{m=1}^{T}k_\rho\left(\frac{r_1-m}{T}\right) - \frac{1}{T}\sum_{r_2=1}^{T}\sum_{k=1}^{T}k_\rho\left(\frac{k-r_2}{T}\right)$$

$$+\sum_{r_2=1}^{T}\frac{1}{T^2}\sum_{k=1}^{T}\sum_{m=1}^{T}k_\rho\left(\frac{k-m}{T}\right) + Tk_\rho\left(\frac{h_1}{T}\right) + C(h_1)$$

$$= -\frac{1}{T}\sum_{r=1}^{T}\sum_{s=1}^{T}k_\rho\left(\frac{r-s}{T}\right) + Tk_\rho\left(\frac{h_1}{T}\right) + C(h_1)$$

$$= \sum_{r_2=1}^{T}\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_2}{T}\right) + T\left\{k_\rho\left(\frac{h_1}{T}\right) - k_\rho\left(0\right)\right\} + C(h_1), \tag{A.8}$$

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

$$\sum_{r_2=1-h_1}^{T}\tilde{k}_\rho\left(\frac{r_2+h_1}{T},\frac{r_2}{T}\right) = \sum_{r_2=1}^{T}\tilde{k}_\rho\left(\frac{r_2}{T},\frac{r_2}{T}\right) + T\left\{k_\rho\left(\frac{h_1}{T}\right) - k_\rho\left(0\right)\right\} + C(h_1). \tag{A.9}$$

Therefore,

$$\text{Trace}\left(\Omega_T A_T W_\rho A_T\right)$$

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

$$+T\sum_{h_1=-T+1}^{T-1}\gamma\left(h_1\right)\left\{k_\rho\left(\frac{h_1}{T}\right) - k_\rho\left(0\right)\right\} + O\left(1\right)$$

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

$$-\frac{\rho g}{T^{q-1}}\sum_{h_1=-T+1}^{T-1}|h_1|^q\,\gamma\left(h_1\right) + O\left(\frac{\rho}{T^{q-1}}\right) + O(1),$$

where we have used the second-order Taylor expansion

$$k_\rho\left(\frac{|h_1|}{T}\right) - k_\rho(0) = -\rho g\left|h_1^q\right|/T^q + o\left(\rho/T^q\right).$$

Using

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

and

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

we now have

$$\mu_{\rho T} = \int_0^1 k_\rho^*(r,r)dr - \frac{\rho g}{T^q}\frac{1}{\omega_T^2}\sum_{h=-T+1}^{T-1} |h|^q \gamma(h)(1+o(1)) + O\left(\frac{1}{T}\right). \qquad \textbf{(A.10)}$$

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

We next approximate $\mathrm{Trace}\left[\left(\Omega_T A_T W_\rho A_T\right)^m\right]$ 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

$$\begin{aligned}
&\mathrm{Trace}\left[\left(\Omega_T A_T W_\rho A_T\right)^m\right] \\
&\quad= \sum_{r_2,r_4,\ldots,r_{2m}=1}^{T}\prod_{j=1}^{m}\gamma(r_{2j-1}-r_{2j})\tilde{k}_\rho\left(\frac{r_{2j}}{T},\frac{r_{2j+1}}{T}\right) \\
&\quad= \sum_{r_2,r_4,\ldots,r_{2m}=1}^{T}\sum_{h_1=1-r_2}^{T-r_2}\sum_{h_2=1-r_4}^{T-r_4}\cdots\sum_{h_m=1-r_{2m}}^{T-r_{2m}}\prod_{j=1}^{m}\gamma(h_j)\tilde{k}_\rho\left(\frac{r_{2j}}{T},\frac{r_{2j+2}+h_{j+1}}{T}\right) \\
&\quad= \left(\sum_{h_1=1}^{T-1}\sum_{r_2=1}^{T-h_1} + \sum_{h_1=1-T}^{0}\sum_{r_2=1-h_1}^{T}\right)\cdots\left(\sum_{h_m=1}^{T-1}\sum_{r_{2m}=1}^{T-h_m} + \sum_{h_m=1-T}^{0}\sum_{r_{2m}=1-h_m}^{T}\right) \\
&\qquad \prod_{j=1}^{m}\gamma(h_j)\tilde{k}_\rho\left(\frac{r_{2j}}{T},\frac{r_{2j+2}+h_{j+1}}{T}\right) \\
&\quad= I + II,
\end{aligned}$$

where

$$\begin{aligned}
I = &\left(\sum_{h_1=1}^{T-1}\sum_{r_2=1}^{T-h_1} + \sum_{h_1=1-T}^{0}\sum_{r_2=1-h_1}^{T}\right)\cdots\left(\sum_{h_m=1}^{T-1}\sum_{r_{2m}=1}^{T-h_m} + \sum_{h_m=1-T}^{0}\sum_{r_{2m}=1-h_m}^{T}\right) \\
&\prod_{j=1}^{m}\gamma(h_j)\tilde{k}_\rho\left(\frac{r_{2j}}{T},\frac{r_{2j+2}}{T}\right),
\end{aligned}$$

and

$$II = O\left\{\left(\sum_{h_1=1}^{T-1}\sum_{r_2=1}^{T-h_1} + \sum_{h_1=1-T}^{0}\sum_{r_2=1-h_1}^{T}\right)\cdots\left(\sum_{h_m=1}^{T-1}\sum_{r_{2m}=1}^{T-h_m} + \sum_{h_m=1-T}^{0}\sum_{r_{2m}=1-h_m}^{T}\right)\right.$$
$$\left.\prod_{j=1}^{m}\left|\gamma\left(h_j\right)\right|\left(\frac{\rho|h_{j+1}|}{T}\right)\right\}.$$

Here we have used

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

A similar result is given and proved in (A.15) below.

The first term (I) can be written as

$$I = \left(\sum_{h_1=1-T}^{T-1}\sum_{r_2=1}^{T} - \sum_{h_1=1}^{T-1}\sum_{r_2=T-h_1+1}^{T} - \sum_{h_1=1-T}^{0}\sum_{r_2=1}^{-h_1}\right)\cdots$$
$$\left(\sum_{h_m=1-T}^{T-1}\sum_{r_{2m}=1}^{T} - \sum_{h_m=1}^{T-1}\sum_{r_{2m}=T-h_m+1}^{T} - \sum_{h_m=1-T}^{0}\sum_{r_{2m}=1}^{-h_m}\right)$$
$$\prod_{j=1}^{m}\gamma\left(h_j\right)\left\{\tilde{k}_\rho\left(\frac{r_{2j}}{T},\frac{r_{2j+2}}{T}\right)\right\}$$
$$= \sum_{\pi}\sum_{h_1,r_2}\cdots\sum_{h_m,r_{2m}}\prod_{j=1}^{m}\gamma\left(h_j\right)\left\{\tilde{k}_\rho\left(\frac{r_{2j}}{T},\frac{r_{2j+2}}{T}\right)\right\}, \qquad \textbf{(A.11)}$$

where $\sum_{h_j,r_{2j}}$ is one of the three choices $\sum_{h_j=1-T}^{T-1}\sum_{r_{2j}=1}^{T}$, $-\sum_{h_j=1}^{T-1}\sum_{r_{2j}=T-h_j+1}^{T}$, and $-\sum_{h_j=1-T}^{0}\sum_{r_{2j}=1}^{-h_j}$, 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 (A.11) 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

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

The dominating terms in the second group are of the forms

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

or

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

both of which are bounded by

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

using the same approach as in (A.4). Approximating the sum by 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 $O\left[mT^{m-2}\left(4\int_0^1 k_\rho(v)dv\right)^{m-2}\right]$ uniformly over m. As a consequence,

$$I=\left(\sum_h\prod_{j=1}^{m}\gamma(h_j)\right)\sum_r\left\{\tilde{k}_\rho\left(\frac{r_{2j}}{T},\frac{r_{2j+2}}{T}\right)\right\}+O\left\{mT^{m-2}\left(4\int_0^1 k_\rho(v)dv\right)^{m-2}\right\},$$

uniformly over m.

The second term (II) is easily shown to be of smaller order than the first term (I). Therefore

$$\begin{aligned}&\text{Trace}\left[\left(\Omega_T A_T W_\rho A_T\right)^m\right]\\&\quad=\left(\sum_h\gamma(h)\right)^m\sum_r\left\{\tilde{k}_\rho\left(\frac{r_{2j}}{T},\frac{r_{2j+2}}{T}\right)\right\}+O\left\{mT^{m-2}\left(4\int_0^1 k_\rho(v)dv\right)^{m-2}\right\},\end{aligned}$$

and

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

uniformly over m.

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

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

It follows that

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

where the sum is taken over the elements

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

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

$$\alpha_{m,T} = \alpha_m + O\left\{ \frac{2^{m-1}}{T^2} \left(4 \int_0^1 k_\rho(v) dv\right)^{m-2} \sum_\pi \frac{m!}{m_1! m_2! \cdots m_k!} \right\}$$

$$= \alpha_m + O\left\{ \frac{2^{2m-1} m!}{T^2} \left(4 \int_0^1 k_\rho(v) dv\right)^{m-2} \right\}$$

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

LEMMA A.4. *Let Assumption 2 hold. If $\rho \to \infty$ and $T \to \infty$ such that $\rho/T^q \to 0$, then*

(a)

$$\mu_{\rho T} = \mu_\rho - \frac{\rho}{T^q \omega_T^2} \sum_{h=-T+1}^{T-1} |h|^q \gamma(h)(1 + o(1)) + O\left(\frac{1}{T}\right) + o\left(\frac{\rho}{T^q}\right); \quad \textbf{(A.12)}$$

(b)

$$\kappa_{2,T} = 2 \int_0^1 \int_0^1 \left(k_\rho^*(r,s)\right)^2 dr ds + O\left(\frac{1}{T}\right); \quad \textbf{(A.13)}$$

(c) for $m = 3$ and 4,

$$\kappa_{m,T} = O\left(\left(\rho^{-1/q}\right)^{m-1}\right) + O\left(\frac{1}{T}\right). \quad \textbf{(A.14)}$$

Proof of Lemma A.4. We have proved (A.12) in the proof of Lemma A.3, as equation (A.10) holds for both fixed ρ and increasing ρ. It remains to consider $\kappa_{m,T}$ for $m = 2, 3,$ and 4. We first consider $\kappa_{2,T} = 2T^{-2}\left(\omega_T^{-4}\right)\text{Trace}\left[\left(\Omega_T A_T W_\rho A_T\right)^2\right]$. As a first step, we have

$$\text{Trace}\left[\left(\Omega_T A_T W_\rho A_T\right)^2\right]$$

$$= \sum_{r_1, r_2, r_3, r_4} \left\{ \tilde{k}_\rho\left(\frac{r_2}{T}, \frac{r_3}{T}\right) \tilde{k}_\rho\left(\frac{r_4}{T}, \frac{r_1}{T}\right) \right\} \gamma(r_1 - r_2)\gamma(r_3 - r_4)$$

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

$$= \left(\sum_{h_1=1}^{T-1}\sum_{r_2=1}^{T-h_1} + \sum_{h_1=1-T}^{0}\sum_{r_2=1-h_1}^{T}\right)\left(\sum_{h_2=1}^{T-1}\sum_{r_4=1}^{T-h_2} + \sum_{h_2=1-T}^{0}\sum_{r_4=1-h_2}^{T}\right)$$

$$\left\{\tilde{k}_\rho\left(\frac{r_2}{T}, \frac{r_4+h_2}{T}\right)\tilde{k}_\rho\left(\frac{r_4}{T}, \frac{r_2+h_1}{T}\right)\right\}\gamma(h_1)\gamma(h_2)$$

$$:= I_1 + I_2 + I_3 + I_4,$$

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

$$\left|k_\rho\left(\frac{r_2-r_4-h_2}{T}\right) - k_\rho\left(\frac{r_2-r_4}{T}\right)\right| = O\left(\frac{\rho|h_2|}{T}\right),$$

we have

$$\begin{aligned}
&\tilde{k}_\rho\left(\frac{r_2}{T}, \frac{r_4+h_2}{T}\right) \\
&\quad= \tilde{k}_\rho\left(\frac{r_2}{T}, \frac{r_4}{T}\right) + k_\rho\left(\frac{r_2-r_4-h_2}{T}\right) - k_\rho\left(\frac{r_2-r_4}{T}\right) + O\left(\frac{|h_2|}{T}\right) \\
&\quad= \tilde{k}_\rho\left(\frac{r_2}{T}, \frac{r_4}{T}\right) + O\left(\frac{\rho|h_2|}{T}\right).
\end{aligned} \tag{A.15}$$

Similarly

$$\tilde{k}_\rho\left(\frac{r_4}{T}, \frac{r_2+h_1}{T}\right) = \tilde{k}_\rho\left(\frac{r_4}{T}, \frac{r_2}{T}\right) + O\left(\frac{\rho|h_1|}{T}\right). \tag{A.16}$$

It follows from (A.15) and (A.16) that

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

Following the same procedure, we can show that

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

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

and

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

In consequence,

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

and

$$\kappa_{2,T}=2T^{-2}\left(\omega_T^{-4}\right)\text{Trace}\left[\left(\Omega_T A_T W_\rho A_T\right)^2\right]=2\int_0^1\int_0^1\left(k_\rho^*(r,s)\right)^2 dr\,ds+O\left(\frac{1}{T}\right).$$

So for the first-order power kernels,

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

and for the second-order power kernels,

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

where the last line uses equation (A.17) below.

The proof for $\kappa_{m,T}$ for $m=3,4$ is essentially the same except that we use Lemma A.1(a) or $\int_0^1 (1-x)^\rho\,dx = O(1/\rho)$ and Lemma A.2 to obtain the first term $O\left(\left(\rho^{-1/q}\right)^{m-1}\right)$. Details are omitted. ■

A.2. Proofs of the Main Results.

Proof of Theorem 1. First we consider the second-order power kernels. Combining Lemma A.1(a) with Lemma A.2, we have

$$|\alpha_m| = O\left(\rho^{-(m-1)/q}\right).$$

As a consequence,

$$\begin{aligned} F_\delta(z) &= P\left\{\left|(W(1)+\delta)\Xi_\rho^{-1/2}\right| < z\right\} \\ &= G_\delta(\mu_\rho z^2) + \frac{1}{2}G''_\delta(\mu_\rho z^2)z^4\alpha_2 + O(\rho^{-2/q}), \end{aligned}$$

where

$$\mu_\rho = \mathrm{E}\,\Xi_\rho = \int_0^1 k^*_\rho(r,r)dr = 1 - \int_0^1\int_0^1 k_\rho(r-s)drds,$$

and

$$\alpha_2 = 2\left(\int_0^1\int_0^1 k_\rho(r-s)drds\right)^2 - 4\int k_\rho(r-s)k_\rho(r-q) + 2\int_0^1\int_0^1 k_\rho^2(r-s)drds.$$

We proceed to approximate μ_ρ and α_2 as $\rho\to\infty$. First, we have

$$\begin{aligned} &\int_0^1\int_0^1 k_\rho(r-s)drds \\ &\quad = 2\int_0^1\left(\int_0^r k_\rho(r-s)ds\right)dr = 2\int_0^1\left(\int_0^r k_\rho(\tau)d\tau\right)dr \\ &\quad = 2\int_0^1\left(\int_\tau^1 k_\rho(\tau)dr\right)d\tau = 2\int_0^1(1-\tau)k_\rho(\tau)d\tau \\ &\quad = \left(\frac{\pi}{\rho g}\right)^{1/2} + O(1/\rho), \end{aligned}$$

by Lemma A.1(b). Second, it follows from Lemma A.1(a) that

$$\begin{aligned} &\int_0^1\left(\int_0^1\int_0^1 k_\rho(r-s)k_\rho(r-q)dsdq\right)dr \\ &\quad = \int_0^1\left(\int_{r-1}^r k_\rho(s)ds\right)^2 dr \le \left(\int_{-1}^1 k_\rho(s)ds\right)^2 = O(1/\rho). \end{aligned}$$

As a consequence,

$$\alpha_2 = 2\int_0^1 \int_0^1 k_\rho^2(r-s)drds + O\left(\frac{1}{\rho}\right) \tag{A.17}$$
$$= 2\left(\frac{\pi}{2\rho g}\right)^{1/2} + O\left(\frac{1}{\rho}\right).$$

Then

$$\begin{aligned} F_\delta(z) &= G_\delta(z^2) + G'_\delta(z^2)\left(\mu_\rho - 1\right)z^2 + \frac{1}{2}G''_\delta(z^2)z^4\alpha_2 + O\left(\frac{1}{\rho}\right) \\ &= G_\delta(z^2) - G'_\delta(z^2)z^2\left(\frac{\pi}{\rho g}\right)^{1/2} + G''_\delta(z^2)z^4\left(\frac{\pi}{2\rho g}\right)^{1/2} + O\left(\frac{1}{\rho}\right) \\ &= G_\delta(z^2) + \left(\frac{\pi}{2\rho g}\right)^{1/2}\left(G''_\delta(z^2)z^4 - \sqrt{2}G'_\delta(z^2)z^2\right) + O\left(\frac{1}{\rho}\right) \\ &= G_\delta(z^2) + c_q\rho^{-1/q}\mathcal{L}_q(G_\delta, z) + O\left(\rho^{-2/q}\right), \end{aligned}$$

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

For the case with the first-order power kernels, we have, after some brute force calculations,

$$\mu_\rho = 1 - \int_0^1 \int_0^1 k_\rho(r-s)drds = \frac{\rho}{\rho+2}$$
$$\alpha_2 = \frac{2}{\rho+1} + O\left(\frac{1}{\rho^2}\right).$$

So

$$\begin{aligned} F(z) &= G_\delta(\mu_\rho z^2) + G''_\delta(\mu_\rho z^2)z^4\frac{1}{\rho+1} + O\left(1/\rho^2\right) \\ &= G_\delta(z^2) - G''_\delta(z^2)z^2\frac{2}{\rho+2} + G''_\delta(z^2)z^4\frac{1}{\rho+1} + O\left(1/\rho^2\right) \\ &= G_\delta(z^2) + \left[G''_\delta(z^2)z^4 - 2G'_\delta(z^2)z^2\right]\frac{1}{\rho} + O\left(1/\rho^2\right) \\ &= G_\delta(z^2) + c_q\rho^{-1/q}\mathcal{L}_q(G_\delta, z) + O\left(\rho^{-2/q}\right), \end{aligned}$$

as desired. ■

Proof of Corollary 1. We focus on the case with the second-order power kernels, as the proof for the first-order power kernels is similar. Using a Taylor series expansion, we have

$$\begin{aligned} &F(z_{\alpha,\rho}) \\ &\quad= D(z_{\alpha,\rho}^2) + \left(\frac{\pi}{\rho g}\right)^{1/2}\left(\frac{1}{\sqrt{2}}D''(z_{\alpha,\rho}^2)z_{\alpha,\rho}^4 - D'(z_{\alpha,\rho}^2)z_{\alpha,\rho}^2\right) + O\left(\frac{1}{\rho}\right) \\ &\quad= D(z_\alpha^2) + \left(\frac{\pi}{\rho g}\right)^{1/2}\left(\frac{1}{\sqrt{2}}D''(z_\alpha^2)z_\alpha^4 - D'(z_\alpha^2)z_\alpha^2\right) + D'(z_\alpha^2)\left(z_{\alpha,\rho}^2 - z_\alpha^2\right) + O\left(\frac{1}{\rho}\right); \end{aligned}$$

that is,

$$0 = \left(\frac{\pi}{\rho g}\right)^{1/2} \left[\frac{1}{\sqrt{2}} D''(z_\alpha^2) z_\alpha^4 - D'(z_\alpha^2) z_\alpha^2\right] + D'(z_\alpha^2)\left(z_{\alpha,\rho}^2 - z_\alpha^2\right) + O\left(\frac{1}{\rho}\right).$$

So

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

Now

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

and thus

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

Hence

$$\begin{aligned} z_{\alpha,\rho}^2 &= z_\alpha^2 + \left(\frac{\pi}{\rho g}\right)^{1/2}\left(z_\alpha^2 - z_\alpha^4 \frac{1}{4z_\alpha^3}\left(\frac{-2z_\alpha - 2z_\alpha^3}{\sqrt{2}}\right)\right) + O\left(\frac{1}{\rho}\right) \\ &= z_\alpha^2 + \left(\frac{\pi}{\rho g}\right)^{1/2}\left\{\left(1 + \frac{\sqrt{2}}{4}\right) z_\alpha^2 + \frac{\sqrt{2}}{4} z_\alpha^4\right\} + O\left(\frac{1}{\rho}\right), \end{aligned}$$

from which we get

$$\begin{aligned} z_{\alpha,\rho} &= \left\{z_\alpha^2 + \left(\frac{\pi}{\rho g}\right)^{1/2}\left[\left(1 + \frac{\sqrt{2}}{4}\right) z_\alpha^2 + \frac{\sqrt{2}}{4} z_\alpha^4\right] + O\left(\frac{1}{\rho}\right)\right\}^{1/2} \\ &= z_\alpha\left\{1 + \left(\frac{\pi}{\rho g}\right)^{1/2}\left[\left(1 + \frac{\sqrt{2}}{4}\right) + \frac{\sqrt{2}}{4} z_\alpha^2\right] + O\left(\frac{1}{\rho}\right)\right\}^{1/2} \\ &= z_\alpha\left\{1 + \frac{1}{2}\left(\frac{\pi}{\rho g}\right)^{1/2}\left[\left(1 + \frac{\sqrt{2}}{4}\right) + \frac{\sqrt{2}}{4} z_\alpha^2\right]\right\} + O\left(\frac{1}{\rho}\right) \\ &= z_\alpha + \frac{1}{2}\left(\frac{\pi}{\rho g}\right)^{1/2}\left\{\left(1 + \frac{\sqrt{2}}{4}\right) z_\alpha + \frac{\sqrt{2}}{4} z_\alpha^3\right\} + O\left(\frac{1}{\rho}\right), \end{aligned}$$

as stated. ■

Proof of Corollary 2. Again, we focus on the case with the second-order power kernels, as the proof for the first-order power kernels is similar. For notational convenience, let

$$f(z_\alpha^2) = \left(\frac{\pi}{g}\right)^{1/2}\left\{\left(1 + \frac{\sqrt{2}}{4}\right) z_\alpha^2 + \frac{\sqrt{2}}{4} z_\alpha^4\right\},$$

and then

$$z_{\alpha,\rho}^2 = z_\alpha^2 + \frac{f(z_\alpha^2)}{\sqrt{\rho}} + O\left(\frac{1}{\rho}\right).$$

We have

$$\begin{aligned}
&1 - \mathrm{E}G_\delta(z_{\alpha,\rho}^2 \Xi_\rho) \\
&\quad = 1 - G_\delta\left(z_\alpha^2 + \frac{1}{\sqrt{\rho}} f(z_\alpha^2)\right) - \left(\frac{\pi}{\rho g}\right)^{1/2} \left\{\frac{1}{\sqrt{2}} G_\delta''\left(z_\alpha^2 + \frac{f(z_\alpha^2)}{\sqrt{\rho}}\right)\left(z_\alpha^2 + \frac{1}{\sqrt{\rho}} f(z_\alpha^2)\right)^2\right\} \\
&\qquad + \left(\frac{\pi}{\rho g}\right)^{1/2} \left\{G_\delta'\left(z_\alpha^2 + \frac{f(z_\alpha^2)}{\sqrt{\rho}}\right)\left(z_\alpha^2 + \frac{f(z_\alpha^2)}{\sqrt{\rho}}\right)\right\} + O\left(\frac{1}{\rho}\right) \\
&\quad = 1 - G_\delta(z_\alpha^2) - G_\delta'(z_\alpha^2)\frac{1}{\sqrt{\rho}} f(z_\alpha^2) - \left(\frac{\pi}{\rho g}\right)^{1/2}\left[\frac{1}{\sqrt{2}} G_\delta''(z_\alpha^2) z_\alpha^4 - G_\delta'(z_\alpha^2) z_\alpha^2\right] + O\left(\frac{1}{\rho}\right) \\
&\quad = 1 - G_\delta(z_\alpha^2) - \left(\frac{\pi}{\rho g}\right)^{1/2}\left\{G_\delta'(z_\alpha^2)\left[\left(1 + \frac{\sqrt{2}}{4}\right) z_\alpha^2 + \frac{\sqrt{2}}{4} z_\alpha^4\right]\right. \\
&\qquad \left. + \frac{1}{\sqrt{2}} G_\delta''(z_\alpha^2) z_\alpha^4 - G_\delta'(z_\alpha^2) z_\alpha^2\right\} + O\left(\frac{1}{\rho}\right) \\
&\quad = 1 - G_\delta(z_\alpha^2) - \left(\frac{\pi}{\rho g}\right)^{1/2}\left(\frac{\sqrt{2}}{4} G_\delta'(z_\alpha^2) z_\alpha^4 + \frac{\sqrt{2}}{2} G_\delta''(z_\alpha^2) z_\alpha^4 + \frac{\sqrt{2}}{4} G_\delta'(z_\alpha^2) z_\alpha^2\right) + O\left(\frac{1}{\rho}\right) \\
&\quad = 1 - G_\delta(z_\alpha^2) - \left(\frac{\pi}{\rho g}\right)^{1/2}\frac{\sqrt{2}}{2}\left(\frac{1}{2} G_\delta'(z_\alpha^2) z_\alpha^4 + G_\delta''(z_\alpha^2) z_\alpha^4 + \frac{1}{2} G_\delta'(z_\alpha^2) z_\alpha^2\right) + O\left(\frac{1}{\rho}\right).
\end{aligned}$$

Note that

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

and

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

so

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

and

$$1 - \mathrm{E}G_\delta(z_{\alpha,\rho}^2 \Xi_\rho) = 1 - G_\delta(z_\alpha^2) - c_q z_\alpha^4 K_\delta\left(z_\alpha^2\right)\frac{1}{\sqrt{\rho}} + O\left(\frac{1}{\rho}\right).$$

To derive the alternative expression for $K_\delta(\cdot)$, we note that

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

$$= \sum_{k=0}^{\infty}\frac{\left(\delta^2/2\right)^{k+1}}{k!}e^{-\delta^2/2}\frac{z^{k+1/2}e^{-z/2}}{\Gamma(k+3/2)2^{j+3/2}}\frac{1}{z}$$

$$= \frac{\delta^2}{2z}\sum_{k=0}^{\infty}\frac{\left(\delta^2/2\right)^{k}}{k!}e^{-\delta^2/2}\frac{z^{k+1/2}e^{-z/2}}{\Gamma(k+3/2)2^{k+3/2}} = \frac{\delta^2}{2z}G'_\delta(z,3).$$

It is also easy to show that $K_\delta(z) = G''_\delta(z) - \left[D''(z)/D'(z)\right]G'_\delta(z)$. ■

Proof of Theorem 2. First, since $G_\delta(\cdot)$ is a bounded function, we can rewrite (18) as

$$P\left\{\left|(W(1)+\delta)\,\Xi_\rho^{-1/2}\right| \le z\right\}$$
$$= \lim_{B\to\infty}\mathrm{E}G_\delta(\Xi_\rho z^2)1\left\{\left|\Xi_\rho - \mu_\rho\right| \le B\right\}$$
$$= \lim_{B\to\infty}\mathrm{E}\sum_{m=1}^{\infty}\frac{1}{m!}G_\delta^{(m)}(\mu_\rho z^2)\left(\Xi_\rho - \mu_\rho\right)^m z^{2m}1\left\{\left|\Xi_\rho - \mu_\rho\right| \le B\right\}$$
$$= \lim_{B\to\infty}\sum_{m=1}^{\infty}\frac{1}{m!}G_\delta^{(m)}(\mu_\rho z^2)\alpha_m z^{2m}1\left\{\left|\Xi_\rho - \mu_\rho\right| \le B\right\}, \quad \textbf{(A.18)}$$

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 $G_\delta(\Xi_\rho z^2)$ when $\left|\Xi_\rho - \mu_\rho\right| \le B$. The uniformity holds because $G_\delta(\cdot)$ is infinitely differentiable with bounded derivatives. Using Lemma A.2, we have, for some constant $C > 0$,

$$\left|\sum_{m=1}^{\infty}\frac{1}{m!}G_\delta^{(m)}(\mu_\rho z^2)\alpha_m z^{2m}\right|$$
$$\le C\sum_{m=1}^{\infty}\frac{1}{m!}z^{2m}\left|\alpha_m\right| \le C\sum_{m=1}^{\infty}\frac{1}{m!}z^{2m}2^{4m-4}(m-1)!\left(\int_0^1 k_\rho(v)dv\right)^{m-1}$$
$$= \frac{C}{16}\sum_{m=1}^{\infty}\frac{1}{m}\left(16z^2\right)^m\left(\int_0^1 k_\rho(v)dv\right)^{m-1} < \infty, \quad \textbf{(A.19)}$$

provided that $\int_0^1 k_\rho(v)dv < 1/\left(16z^2\right)$. As a consequence, the limit $\lim_{B\to\infty}$ can be moved inside the summation sign in (A.18), giving

$$P\left\{\left|(W(1)+\delta)\,\Xi_\rho^{-1/2}\right| \le z\right\} = \sum_{m=1}^{\infty}\frac{1}{m!}G_\delta^{(m)}(\mu_\rho z^2)\alpha_m z^{2m}, \quad \textbf{(A.20)}$$

when $\int_0^1 k_\rho(v)dv < 1/\left(16z^2\right)$.

Second, it follows from (29) that

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

But

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

where the right-hand side converges to $\mathrm{E}\left\{G_\delta(z^2\varsigma_{\rho T})\right\}$ uniformly over T because (i)

$$\alpha_{m,T} = \alpha_m + O\left\{\frac{2^{2m-1}m!}{T^2}\left(4\int_0^1 k_\rho(v)dv\right)^{m-2}\right\}$$

uniformly over m by Lemma A.3, (ii) $G_\delta^{(m)}(\cdot)$ is a bounded function, and (iii)

$$\begin{aligned}
&\sum_{m=1}^{\infty}\frac{1}{m!}z^{2m}\frac{2^{2m-1}m!}{T^2}\left(4\int_0^1 k_\rho(v)dv\right)^{m-2}, \\
&\quad = \frac{1}{32T^2}\sum_{m=1}^{\infty}\left(16z^2\right)^m\left(\int_0^1 k_\rho(v)dv\right)^{m-2} < \infty
\end{aligned} \tag{A.21}$$

when $\int_0^1 k_\rho(v)dv < 1/\left(16z^2\right)$. Therefore,

$$P\left\{\left|\sqrt{T}\left(\hat{\beta}-\beta_0\right)/\hat{\omega}\right| \le z\right\} = \sum_{m=1}^{\infty}\frac{1}{m!}G_\delta^{(m)}(\mu_{\rho T}z^2)\alpha_{m,T}z^{2m} + O(\frac{1}{T}). \tag{A.22}$$

It follows from (A.20) and (A.22) that

$$\begin{aligned}
&\left|P\left\{\left|\sqrt{T}\left(\hat{\beta}-\beta_0\right)/\hat{\omega}\right| \le z\right\} - P\left\{\left|(W(1)+\delta)\,\Xi_\rho^{-1/2}\right| < z\right\}\right| \\
&\quad = \left|\sum_{m=1}^{\infty}\frac{1}{m!}G_\delta^{(m)}(\mu_{\rho T}z^2)\alpha_{m,T}z^{2m} - \sum_{m=1}^{\infty}\frac{1}{m!}G_\delta^{(m)}(\mu_\rho z^2)\alpha_m z^{2m}\right| + O\left(\frac{1}{T}\right) \\
&\quad = \left|\sum_{m=1}^{\infty}\frac{1}{m!}G_\delta^{(m)}(\mu_\rho z^2)\alpha_{m,T}z^{2m} - \sum_{m=1}^{\infty}\frac{1}{m!}G_\delta^{(m)}(\mu_\rho z^2)\alpha_m z^{2m}\right| + O\left(\frac{1}{T}\right) \\
&\quad = \left|\sum_{m=1}^{\infty}\frac{1}{m!}G_\delta^{(m)}(\mu_\rho z^2)\left(\alpha_{m,T}-\alpha_m\right)z^{2m}\right| + O\left(\frac{1}{T}\right) \\
&\quad = O\left\{\frac{1}{T^2}\sum_{m=1}^{\infty}2^{2m-1}\left(4\int_0^1 k_\rho(v)dv\right)^{m-2}z^{2m}\right\} + O\left(\frac{1}{T}\right) \\
&\quad = O\left(\frac{1}{T}\right),
\end{aligned} \tag{A.23}$$

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 A.3 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 (A.21). This completes the proof of Theorem 2. ■

Proof of Theorem 3. It follows from Lemma A.4 that when $\rho \to \infty$,

$$\alpha_{2,T} = \kappa_{2,T} = 2c_q \rho^{-1/q} + O\left(\rho^{-2/q} + T^{-1}\right), \tag{A.24}$$

$$\alpha_{3,T} = \kappa_{3,T} = O\left(\rho^{-2/q}\right) + O\left(T^{-1}\right), \tag{A.25}$$

$$\alpha_{4,T} = \kappa_{4,T} + 3\kappa_{2,T}^2 = O\left(\rho^{-2/q}\right) + O\left(T^{-1}\right), \tag{A.26}$$

and

$$\mu_{\rho T} = \mu_\rho - g\left(\rho T^{-q}\right)\omega_T^{-2} \sum_{h=-T+1}^{T-1} h\Gamma(h)(1+o(1)) + O\left(T^{-1}\right). \tag{A.27}$$

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

$$\begin{aligned}
F_{T,\delta}(z) &= P\left\{\left|\sqrt{T}\left(\hat{\beta} - \beta_0\right)/\hat{\omega}\right| \le z\right\} = \mathrm{E}\left\{G_\delta(z^2 \varsigma_{\rho T})\right\} + O(T^{-1}) \\
&= G_\delta(\mu_{\rho T} z^2) + c_q G_\delta''(\mu_{\rho T} z^2) z^4 \rho^{-1/q} + O\left(\rho^{-2/q}\right) + O\left(T^{-1}\right) \\
&= G_\delta(\mu_\rho z^2) + G_\delta'(\mu_\rho z^2) z^2\left(\mu_{\rho T} - \mu_\rho\right) + c_q G_\delta''(\mu_{\rho T} z^2) z^4 \rho^{-1/q} \\
&\quad + O\left(\rho^{-2/q}\right) + O\left(T^{-1}\right),
\end{aligned}$$

using (A.24) to (A.27). But

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

and

$$\begin{aligned}
&G_\delta'(\mu_\rho z^2) z^2\left(\mu_{\rho T} - \mu_\rho\right) \\
&\quad = \left[G_\delta'(z^2) + O\left(\rho^{-1/q}\right)\right] z^2 \left(-g\left(\rho T^{-q}\right)\omega_T^{-2} \sum_{h=1-T}^{T-1} h^q \Gamma(h)(1+o(1)) + O\left(T^{-1}\right)\right) \\
&\quad = -g\left(\rho T^{-q}\right)\omega_T^{-2} \sum_{h=1-T}^{T-1} h^q \Gamma(h) G_\delta'(z^2) z^2 (1+o(1)) + O\left(\rho^{-2/q}\right) + O\left(T^{-1}\right),
\end{aligned}$$

so that

$$\begin{aligned}
F_{T,\delta}(z) &= G_\delta(z^2) + \rho^{-1/q} c_q\left(G_\delta''(\mu_\rho z^2) z^4 - 2^{1/q} G_\delta'(z^2) z^2\right) - g d_{\gamma T} G_\delta'(z^2) z^2\left(\rho T^{-q}\right) \\
&\quad + O\left(\rho^{-2/q}\right) + O\left(T^{-1}\right) + o\left(\rho T^{-q}\right) \\
&= G_\delta(z^2) + c_q \mathcal{L}_q(G_\delta, z)\rho^{-1/q} - g d_{\gamma T} G_\delta'(z^2) z^2\left(\rho T^{-q}\right) \\
&\quad + O\left(\rho^{-2/q}\right) + O\left(T^{-1}\right) + o\left(\rho T^{-q}\right),
\end{aligned}$$

as desired. ■

Proof of Corollary 3. Part (a). Using Theorem 3, we have, as $1/\rho + 1/T + \rho/T^q \to 0$.

$$\begin{aligned}
F_{T,0}(z_{\alpha,\rho}) &= D(z_{\alpha,\rho}^2) + \left[D''(z_{\alpha,\rho}^2)z_{\alpha,\rho}^4 - 2^{1/q}D'(z_{\alpha,\rho}^2)z_{\alpha,\rho}^2\right]c_q\rho^{-1/q} \\
&\quad - d_{\gamma T}D'(z_{\alpha,\rho}^2)z_{\alpha,\rho}^2\left(\rho T^{-q}\right) + O\left(T^{-1} + \rho^{-2/q}\right) + o\left(\left(\rho T^{-q}\right)\right) \\
&= F(z_{\alpha,\rho}) - d_{\gamma T}D'(z_{\alpha,\rho}^2)z_{\alpha,\rho}^2\left(\rho T^{-q}\right) + O\left(T^{-1} + \rho^{-2/q}\right) + o\left(\rho T^{-q}\right) \\
&= 1 - \alpha - d_{\gamma T}D'(z_{\alpha,\rho}^2)z_{\alpha,\rho}^2\left(\rho T^{-q}\right) + O\left(T^{-1} + \rho^{-2/q}\right) + o\left(\rho T^{-q}\right).
\end{aligned}$$

Rearranging the above equation gives

$$1 - F_{T,0}(z_{\alpha,\rho}) - \alpha = d_{\gamma T}D'(z^2)z^2\left(\rho T^{-q}\right) + O\left(T^{-1} + \rho^{-2/q}\right) + o\left(\rho T^{-q}\right).$$

Part (b). Plugging $z_{\alpha,\rho}^2$ into (32) yields

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

where the last equality follows from the same proof as Corollary 2. ■