1-1-2006

Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing

Yixiao Sun

Peter C.B. Phillips

Sainan Jin

OPTIMAL BANDWIDTH SELECTION
IN HETEROSKEDASTICITY-AUTHCORRELATION ROBUST TESTING

By

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

January 2006

COWLES FOUNDATION DISCUSSION PAPER NO. 1545

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

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

Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing*

Yixiao Sun
Department of Economics
University of California, San Diego

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

and

Sainan Jin
Guanghua School of Management
Peking University

October 29, 2005

*An earlier version of the paper was presented at the 9th World Congress of the Econometric Society in London. The authors thank Whitney Newey and two anonymous referees for helpful comments. Phillips acknowledges partial support from the Kelly Foundation at the School of Business, University of Auckland, and the NSF under Grant No. SES 04-142254.

ABSTRACT

In time series regressions with nonparametrically autocorrelated errors, it is now standard empirical practice to use kernel-based robust standard errors that involve some smoothing function over the sample autocorrelations. The underlying smoothing parameter b, which can be defined as the ratio of the bandwidth (or truncation lag) to the sample size, is a tuning parameter that plays a key role in determining the asymptotic properties of the standard errors and associated semiparametric tests. Small-b asymptotics involve standard limit theory such as standard normal or chi-squared limits, whereas fixed-b asymptotics typically lead to nonstandard limit distributions involving Brownian bridge functionals. The present paper shows that the nonstandard fixed-b limit distributions of such nonparametrically studentized tests provide more accurate approximations to the finite sample distributions than the standard small-b limit distribution. In particular, using asymptotic expansions of both the finite sample distribution and the nonstandard limit distribution, we confirm that the second-order corrected critical value based on the expansion of the nonstandard limiting distribution is also second-order correct under the standard small-b asymptotics. We further show that, for typical economic time series, the optimal bandwidth that minimizes a weighted average of type I and type II errors is larger by an order of magnitude than the bandwidth that minimizes the asymptotic mean squared error of the corresponding long-run variance estimator. A plug-in procedure for implementing this optimal bandwidth is suggested and simulations confirm that the new plug-in procedure works well in finite samples.

JEL Classification: C13; C14; C22; C51

Keywords: Asymptotic expansion, bandwidth choice, kernel method, long-run variance, loss function, nonstandard asymptotics, robust standard error, Type I and Type II errors

1 Introduction

In time series regressions with autocorrelation of unknown form, the standard errors of regression coefficients are usually estimated nonparametrically by kernel-based methods that involve some smoothing over the sample autocorrelations. The underlying smoothing parameter (b) may be defined as the ratio of the bandwidth to the sample size and is an important tuning parameter that determines the size and power properties of the associated test. It therefore seems sensible that the choice of b should take these properties into account. However, in conventional approaches (e.g., Andrews, 1991, and Newey and West, 1987, 1994) and most practical software, the parameter b is chosen to minimize the asymptotic mean squared error (AMSE) of the long-run variance (LRV) estimator. This approach follows what has long been standard practice in the context of spectral estimation (Grenander and Rosenblatt, 1957; Hannan, 1970) where the focus of attention is the spectrum (or, in the present context, the LRV). Such a choice of the smoothing parameter is designed to be optimal in the AMSE sense for the estimation of the relevant quantity (here, the asymptotic standard error or LRV), but is not necessarily best suited for hypothesis testing and confidence interval construction.

In contrast to the above convention, the present paper develops a new approach to choosing the smoothing parameter. We consider choosing b to optimize a loss function that involves a weighted average of the type I and type II errors, a criterion that addresses the central concerns of interest in hypothesis testing, balancing possible size distortion against possible power loss in the use of different bandwidths. This new approach to automatic bandwidth selection requires improved measurement of type I and type II errors, which are provided here by means of asymptotic expansions of both the finite sample distribution of the test statistic and the nonstandard limit distribution.

We first examine the asymptotic properties of the statistical test under different choices of b. Using a Gaussian location model, we show that the distribution of the conventionally constructed t-statistic is closer to its limit distribution derived under the fixed-b asymptotics than that derived under the small-b asymptotics. More specifically, when b is fixed, the error in the rejection probability (ERP) of the nonstandard t-test is of order $O\left(T^{-1}\right)$ while that of the standard normal test is $O(1)$. On the other hand, when b is decreasing with the sample size, the ERP of the nonstandard t-test is smaller than that of the standard normal test, although they are of the same order of magnitude. As a consequence, the nonstandard t-test has less size distortion than the standard normal test. Our theoretic findings here support earlier simulation results in Kiefer, Vogelsang and Bunzel (2000, hereafter KVB), Kiefer and Vogelsang (2002a, 2002b, hereafter KV), Vogelsang (2003) and ourselves (2005).

In view of its better size property, KV (2005) suggested using the fixed b rule and the associated nonstandard limit distribution in statistical testing. The nonstandard test is not very convenient to use in practice as the critical values of the nonstandard limit distribution have to be simulated for practical implementation. One of the contributions of the present paper is to design an easy-to-implement test based on an asymptotic expansion of the nonstandard limit distribution about a limiting chi-squared form. This expansion, which is of independent interest, leads to an associated Cornish-Fisher type expansion from which high-order corrected critical values may be computed easily. These corrected critical values provide very good approximations to the critical values of the nonstandard limit distribution. In particular, the second-order corrected critical value for the Bartlett kernel and the third-order corrected critical value for the Parzen and QS kernels are remarkably close to their respective exact ones for all values of b in $(0, 1]$.

The corrected critical values are further justified by a higher order asymptotic expansion of the finite sample distribution of the t-statistic under conventional joint limits where the sample size $T \to \infty$ and $b \to 0$ simultaneously. The higher order expansion also enables us to develop improved approximations to the type I and type II errors. More specifically, the type I error is measured by using the first correction term in the asymptotic expansion of the finite sample distribution of the test statistic about its nonstandard limit distribution. This term is of order $O\left((bT)^{-q}\right)$ where q is the Parzen characteristic exponent that measures the degree of smoothness of the kernel used. For typical economic time series, this term increases as b decreases for any given T. Similarly, the expansion under the local alternative reveals that in general the type II error decreases as b decreases. Thus, to this order in the asymptotic expansion, decreasing b reduces the type II error but also increases the type I error. Since the desirable effects on these two types of errors generally work in opposing directions, there is an opportunity to trade off these effects. Accordingly, we construct a loss function criterion by taking a weighted average of these two types of errors and show how b may be selected in such a way as to minimize the loss. This method of choosing b is consistent with the Neyman principle under which the probability of the type II error is minimized after controlling for the probability of the type I error (see, for example, Gourieroux and Monfort, 1995).

Our approach gives an optimal b which generally has a shrinking rate of at most $b = O\left(T^{-q/(q+1)}\right)$ and which can even be $O(1)$ for certain loss functions, depending on the weights that are chosen. Note that the optimal b that minimizes the asymptotic mean squared error of the corresponding LRV estimator is of order $O\left(T^{-2q/(2q+1)}\right)$ (c.f., Andrews (1991)). Thus, optimal values of b for LRV estimation are smaller as $T \to \infty$ than those which are most suited for statistical testing. The fixed b rule is obtained by attaching substantially higher weight to the type I error in the construction of the loss function. This theory therefore provides some insight into the type of loss function for which there is a decision theoretic justification for the use of fixed b rules in econometric testing.

The rest of the paper is organized as follows. Section 2 reviews the first order limit theory for the t-test as $T \to \infty$ with the parameter b fixed and as $T \to \infty$ with b approaching zero. Section 3 develops an asymptotic expansion of the nonstandard distribution under the null and local alternative hypotheses as $b \to 0$ about the usual central and noncentral chi-squared distributions. The high-order terms in this asymptotic expansion under the null delivers correction terms that can be used to adjust the critical values in the usual chi-squared test. Section 4 develops comparable expansions of the finite sample distribution of the statistic as $T \to \infty$ and $b \to 0$ at the same time. This expansion validates the use of the corrected critical values in practical work. Section 5 compares the accuracy of the nonstandard approximation with that of the standard normal approximation. Section 6 proposes a selection rule for b that is suitable for implementation in semiparametric testing. The subsequent section reports simulation evidence on the performance of the new procedure. The last section provides some concluding discussion. Proofs and additional technical results are given in the Appendix.

2 Heteroskedasticity-Autocorrelation Robust Inference

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

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

where u_t is zero mean process with a nonparametric autocorrelation structure. The nonstandard limiting distribution in this section and its asymptotic expansion in Section 3 apply to general regression models under certain conditions on the regressors, see KV (2002a, 2002b, 2005). On the other hand, the asymptotic expansion of the finite sample distribution in Section 4 applies only to the location model. Although the location model is of interest in its own right, this is a limitation of the paper and some possible extensions are discussed in Section 8.

OLS estimation of β gives $\hat{\beta} = \bar{Y} = \frac{1}{T}\sum_{t=1}^{T} y_t$, and the scaled and centred estimation error is

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

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

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

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

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

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

Under Assumption 1,

$$T^{-1/2} \hat{S}_{[Tr]} \Rightarrow \omega V(r), \quad r \in [0, 1], \tag{4}$$

where V is a standard Brownian bridge process, and

$$\sqrt{T}(\hat{\beta} - \beta) \Rightarrow \omega W(1) = N(0, \omega^2), \tag{5}$$

which provides the usual basis for robust testing about β. It is standard empirical practice to estimate ω^2 using kernel-based nonparametric estimators that involve some smoothing and possibly truncation of the autocovariances. When u_t is stationary, the long-run variance of u_t is

$$\omega^2 = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma(j), \tag{6}$$

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

$$\hat{\omega}^2(M) = \sum_{j=-T+1}^{T-1} k(\frac{j}{M})\hat{\gamma}(j),\ \hat{\gamma}(j) = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-j} \hat{u}_{t+j}\hat{u}_t & \text{for } j \geq 0 \\ \frac{1}{T}\sum_{t=-j+1}^{T} \hat{u}_{t+j}\hat{u}_t & \text{for } j < 0 \end{cases} \tag{7}$$

involving the sample covariances $\hat{\gamma}(j)$. In (7), $k(\cdot)$ is some kernel function and M is a bandwidth parameter. Consistency of $\hat{\omega}^2(M)$ requires $M \to \infty$ and $M/T \to 0$ as $T \to \infty$ (e.g. Andrews (1991), Andrews and Monahan (1992), Hansen (1992), Newey and West (1987,1994), de Jong and Davidson (2000)). Jansson (2002) provides a recent overview and weak conditions for consistency of such estimates.

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

$$t_{\hat{\omega}(M)} = T^{1/2}(\hat{\beta} - \beta_0)/\hat{\omega}(M), \tag{8}$$

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

To reduce size distortion, KVB and KV proposed the use of kernel-based estimators of ω^2 in which M is set proportional to T, i.e. $M = bT$ for some $b \in (0,1]$. In this case, the estimator $\hat{\omega}^2$ becomes

$$\hat{\omega}_b^2 = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{bT}\right)\hat{\gamma}(j), \tag{9}$$

and the associated t statistic is given by

$$t_b = T^{1/2}(\hat{\beta} - \beta_0)/\hat{\omega}_b. \tag{10}$$

When the parameter b is fixed as $T \to \infty$, KV showed that under Assumption 1 $\hat{\omega}_b^2 \Rightarrow \omega^2 \Xi_b$, where the limit Ξ_b is random and given by

$$\Xi_b = \int_0^1 \int_0^1 k_b(r-s)dV(r)dV(s), \tag{11}$$

with $k_b(\cdot) = k(\cdot/b)$ and the t_b-statistic has a nonstandard limit distribution. Under the null hypothesis

$$t_b \Rightarrow W(1)\Xi_b^{-1/2}, \tag{12}$$

whereas under the local alternative $H_1 : \beta = \beta_0 + cT^{-1/2}$,

$$t_b \Rightarrow (\delta + W(1))\,\Xi_b^{-1/2}, \tag{13}$$

where $\delta = c/\omega$. Thus, the t_b-statistic has a nonstandard limit distribution arising from the random limit of the LRV estimate $\hat{\omega}_b$ when b is fixed as $T \to \infty$. However, as b decreases, the effect of this randomness diminishes, and when $b \to 0$ the limit distributions under the null and local alternative both approach those of conventional regression tests with consistent LRV estimates.

In related work, the present authors (2005a, 2005b, hereafter PSJ$_a$, PSJ$_b$) propose using an estimator of ω^2 of the form

$$\hat{\omega}_\rho^2 = \sum_{j=-T+1}^{T-1} \left[k\left(\frac{j}{T}\right)\right]^\rho \hat{\gamma}(j), \tag{14}$$

which involves setting M equal to T and taking an arbitrary power $\rho \geq 1$ of the traditional kernel. Statistical tests based on $\hat{\omega}_\rho^2$ and $\hat{\omega}_b^2$ share many of the same properties, which is explained by the fact that ρ and b play similar roles in the construction of the estimates. The present paper focuses on $\hat{\omega}_b^2$ and tests associated with this estimate. Comparable ideas and methods to those explored in the present paper may be pursued in the context of estimates such as $\hat{\omega}_\rho^2$ and are reported in other work (PSJ$_c$ and PSJ$_d$).

3 Expansion of the Nonstandard Limit Theory

This section develops asymptotic expansions of the limit distributions given in (12) and (13) as the bandwidth parameter $b \to 0$. These expansions are taken about the relevant central and noncentral chi-squared limit distributions that apply when $b \to 0$, corresponding to the null and local alternative hypotheses.

These expansions of the nonstandard limit distributions are of some independent interest. For instance, they can be used to deliver correction terms to the limit distributions under the null, thereby providing a mechanism for adjusting the nominal critical values provided by the usual chi-squared distribution. The latter correspond to the critical values that would be used for tests based on conventional consistent HAC estimates. As we shall see, when the higher-order correction on the nominal chi-squared asymptotic critical value is implemented using this asymptotic expansion, the resulting expression provides an asymptotic justification for the polynomial approximation suggested in KV(2005) for practical testing situations.

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

Assumption 2 *(i) $k(x) : \mathbb{R} \to [0,1]$ is symmetric, piecewise smooth with $k(0) = 1$ and $\int_0^\infty k(x)xdx < \infty$.*

(ii) The Parzen characteristic exponent defined by

$$q = \max\{q_0 : q_0 \in \mathbb{Z}^+,\ g_{q_0} = \lim_{x\to 0} \frac{1-k(x)}{|x|^{q_0}} < \infty\} \tag{15}$$

is greater than or equal to 1.

(iii) $k(x)$ is positive semidefinite, i.e., for any square integrable function $f(x)$, $\int_0^\infty \int_0^\infty k(s-t)f(s)f(t)dsdt \geq 0$.

Assumption 2 imposes only mild conditions on the kernel function. All the commonly used kernels satisfy (i) and (ii). The assumption $\int_0^\infty k(x)xdx < \infty$ ensures the integrals that appear frequently in our later developments are finite. It also enables us to use the Riemann-Lebesgue lemma in our proofs. The assumption of positive semidefiniteness in (iii) ensures that the associated LRV estimator is nonnegative. Commonly used kernels that are positive semidefinite include the Bartlett kernel, the Parzen kernel and the QS kernel, which are the main focus of the present paper. For the Bartlett kernel, the Parzen characteristic exponent is 1. For the Parzen and QS kernels, the Parzen characteristic exponent is 2.

We proceed to establish the asymptotic expansion of the nonstandard limiting distribution. Let $G_\lambda = G(\cdot;\lambda^2)$ be the cdf of a non-central $\chi_1^2(\lambda^2)$ variate with noncentrality parameter λ^2. Then

$$P\left\{\left|(\delta + W(1))\,\Xi_b^{-1/2}\right| \le z\right\} = P\left\{(\delta + W(1))^2 \le z^2\Xi_b\right\} = E\left\{G_\delta(z^2\Xi_b)\right\}, \tag{16}$$

in view of the independence of $W(1)$ and Ξ_b. Set $\mu_b = E(\Xi_b)$, and a fourth-order Taylor expansion yields

$$\begin{aligned} G_\delta(z^2\Xi_b) &= G_\delta(\mu_b z^2) + \frac{1}{2}\left(G_\delta''(\mu_b z^2)z^4\right)(\Xi_b - \mu_b)^2 \\ &\quad + \frac{1}{6}\left(G_\delta'''(\mu_b z^2)z^6\right)(\Xi_b - \mu_b)^3 + \frac{1}{24}\left(G_\delta^{(4)}(\tilde{\mu}_b z^2)z^8\right)(\Xi_b - \mu_b)^4, \end{aligned} \tag{17}$$

where $\tilde{\mu}_b$ lies on the line segment between μ_b and Ξ_b. Taking expectation on both sides of the equation and using the fact that $\left|G_\delta^{(4)}(\tilde{\mu}_b z^2)z^8\right| \le C$ for some constant C, we have

$$\begin{aligned} EG_\delta(z^2\Xi_b) &= G_\delta(\mu_b z^2) + \frac{1}{2}G_\delta''(\mu_b z^2)E(\Xi_b - \mu_b)^2 z^4 \\ &\quad + \frac{1}{6}G_\delta'''(\mu_b z^2)E(\Xi_b - \mu_b)^3 z^6 + O\left(E(\Xi_b - \mu_b)^4\right), \end{aligned} \tag{18}$$

as $b \to 0$, where the $O(\cdot)$ term holds uniformly over $z \in \mathbb{R}^+$.

To characterize the asymptotic behavior of $E(\Xi_b - \mu_b)^m$ as $b \to 0$, we write

$$\Xi_b = \int_0^1\int_0^1 k_b^*(r,s)dW(r)dW(s), \tag{19}$$

where $k_b^*(r,s)$ is defined by

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

Here, $k_b^*(r,s)$ is simply the projection of $k_b(r-s)$ onto the Hilbert subspace defined by

$$\{f(\cdot,\cdot) : f \in L^2([0,1]), \int_0^1 f(r,s)dr = 0, \int_0^1 f(r,s)ds = 0, \int_0^1\int_0^1 f(r,s)drds = 0\}.$$

Since $k(r-s)$ is positive semidefinite, by Mercer's theorem (e.g., see Shorack and Wellner (1986)), $k(r-s)$ can be represented as $k(r-s) = \sum_{n=1}^\infty \lambda_n f_n(r)f_n(s)$, where $\lambda_n > 0$ are the eigenvalues of the kernel and $f_n(x)$ are the corresponding eigenfunctions, i.e.

$\lambda_n f_n(s) = \int_0^\infty k(r-s) f_n(r) dr$. Since $\lambda_n > 0$, $k_b^*(r,s)$ is also positive semidefinite because it may be written as

$$k_b^*(r,s) = \sum_{n=1}^{\infty} \lambda_n g_n(r/b) g_n(s/b) \text{ for any } (r,s) \in [0,\infty] \times [0,\infty], \tag{20}$$

where $g_n(r) = f_n(r) - \int_0^\infty f_n(\tau) d\tau$. In consequence, for any function $q(x) \in L^2(\mathbb{R}^+)$, we have

$$\int_0^\infty \int_0^\infty q(r) k_b^*(r,s) q(s) dr ds = \sum_{n=1}^{\infty} \lambda_n \left(\int_0^\infty g_n(r/b) q(r) dr \right)^2 \geq 0. \tag{21}$$

Using Mercer's theorem again, we therefore have

$$k_b^*(r,s) = \sum_{n=1}^{\infty} \lambda_n^* f_n^*(r) f_n^*(s), \tag{22}$$

where $\lambda_n^* > 0$ are the eigenvalues of the kernel and $f_n^*(r)$ are the corresponding eigenfunctions, i.e. $\lambda_n^* f_n^*(s) = \int_0^\infty k_b^*(r,s) f_n^*(r) dr$.

Using representation (22), we can write Ξ_b as $\Xi_b = \sum_{n=1}^\infty \lambda_n^* Z_n^2$, where $Z_n \sim iid$ $N(0,1)$. Therefore, the characteristic function of $\Xi_b - \mu_b$ is given by

$$\phi(t) = E\left\{ e^{it(\Xi_b - \mu_b)} \right\} = e^{-it\mu_b} \prod_{n=1}^{\infty} \left\{ 1 - 2i\lambda_n^* t \right\}^{-1/2}, \tag{23}$$

and the cumulant generating function is

$$\ln \phi(t) = \sum_{m=2}^{\infty} \left\{ 2^{m-1}(m-1)! \sum_{n=1}^{\infty} (\lambda_n^*)^m \right\} \frac{(it)^m}{m!}. \tag{24}$$

Let $\kappa_1, \kappa_2, \kappa_3, \ldots$ be the cumulants of $\Xi_b - \mu_b$. Then

$$\kappa_1 = 0 \text{ and } \kappa_m = 2^{m-1}(m-1)! \sum_{n=1}^{\infty} (\lambda_n^*)^m \text{ for } m \geq 2. \tag{25}$$

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

$$\kappa_m = 2^{m-1}(m-1)! \int_0^1 \ldots \int_0^1 \left(\prod_{j=1}^{m} k_b^*(\tau_j, \tau_{j+1}) \right) d\tau_1 \cdots d\tau_m, \tag{26}$$

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

With these preliminaries, we are able to find the dominating terms in the moments $E(\Xi_b - \mu_b)^m$, $m = 1, 2$ and develop an asymptotic expansion of $P\{\left|(\delta + W(1))\, \Xi_b^{-1/2}\right| \leq z\}$ as the bandwidth parameter $b \to 0$. In fact, a full series expansion is possible using this method, but our purpose here requires only the leading terms in the expansion.

Theorem 1 *Let* $F_\delta(z) := P\left\{\left|(\delta + W(1))\,\Xi_b^{-1/2}\right| \leq z\right\}$ *be the nonstandard limiting distribution, then under Assumption 2,*

$$F_\delta(z) = G_\delta(z^2) + p_\delta(z^2)b + q_\delta(z)b^2 + o\left(b^2\right), \tag{27}$$

where the term $o\left(b^2\right)$ *holds uniformly over* $z \in \mathbb{R}^+$ *as* $b \to 0$,

$$\begin{aligned} p_\delta(z^2) &= c_2 G_\delta''(z^2)z^4 - c_1 G_\delta'(z^2)z^2, \\ q_\delta(z^2) &= -G_\delta'(z^2)z^2 c_3 + \frac{1}{2}G_\delta''(z^2)z^4(c_4 - c_1^2) - G_\delta'''(z^2)z^6 c_1 c_2, \end{aligned} \tag{28}$$

and

$$\begin{array}{ll} c_1 = \int_{-\infty}^{\infty} k(x)dx, & c_2 = \int_{-\infty}^{\infty} k^2(x)dx, \\ c_3 = -\int_{-\infty}^{\infty} k(x)\,|x|\,dx, & c_4 = -\int_{-\infty}^{\infty} k^2(x)\,|x|\,dx. \end{array} \tag{29}$$

As is apparent from the proof of the theorem, the term $c_2 G_\delta''(z^2)z^4 b$ in $p_\delta(z^2)$ arises from the randomness of Ξ_b, whereas the term $c_1 G_\delta'(z^2)z^2 b$ in $p_\delta(z^2)$ arises from the asymptotic bias of Ξ_b. Although Ξ_b converges to 1 as $b \to 0$, we have $var(\Xi_b) = 2bc_2(1 + o(1))$ and $E(\Xi_b) = 1 - bc_1(1 + o(1))$, as is established in the proof. Ξ_b is not centered exactly at 1 because the regression errors have to be estimated. The terms in $q_\delta(z^2)$ are due to the first and second order biases of Ξ_b, the variance of Ξ_b and their interactions.

It follows from Theorem 1 that when $\delta = 0$,

$$\begin{aligned} F_0(z) = {} & D(z^2) + \left[c_2 D''(z^2)z^4 - c_1 D'(z^2)z^2\right] b \\ & \left[-D'(z^2)z^2 c_3 + \frac{1}{2}D''(z^2)z^4(c_4 - c_1^2) - D'''(z^2)z^6 c_1 c_2\right] b^2 + o\left(b^2\right), \end{aligned} \tag{30}$$

where $D(\cdot) = G_0(\cdot)$ is the cdf of χ_1^2 distribution. For any $\alpha \in (0,1)$, let $z_\alpha^2 \in \mathbb{R}^+$, $z_{\alpha,b}^2 \in \mathbb{R}^+$ such that $D(z_\alpha^2) = 1 - \alpha$ and $F_0(z_{\alpha,b}) = 1 - \alpha$. Then, using a Cornish-Fisher type expansion, we can obtain high-order corrected critical values.

Before presenting the corollary below, we introduce some terminology. We call the critical value that is correct up to the $O(b)$ order the second-order corrected critical value. We call the critical value that is correct up to the $O(b^2)$ order the third-order corrected critical value. We will use this convention throughout the rest of the paper. The following quantities appear in the corollary:

$$k_1 = \left(c_1 + \frac{1}{2}c_2\right) z_\alpha^2 + \frac{1}{2}c_2 z_\alpha^4, \tag{31}$$

$$\begin{aligned} k_2 = {} & \left(\frac{1}{2}c_1^2 + \frac{3}{2}c_1c_2 + \frac{3}{16}c_2^2 + c_3 + \frac{1}{4}c_4\right) z_\alpha^2 \\ & + \left(-\frac{1}{2}c_1 + \frac{3}{2}c_1c_2 + \frac{9}{16}c_2^2 + \frac{1}{4}c_4\right) z_\alpha^4 + \left(\frac{5}{16}c_2^2\right) z_\alpha^6 - \left(\frac{1}{16}c_2^2\right) z_\alpha^8, \end{aligned} \tag{32}$$

$$k_3 = \frac{1}{2}\left(c_1 + \frac{1}{2}c_2\right) z_\alpha + \frac{1}{4}c_2 z_\alpha^3, \tag{33}$$

$$\begin{aligned} k_4 &= \left(\frac{1}{8}c_1^2 + \frac{5}{8}c_1 c_2 + \frac{1}{16}c_2^2 + \frac{1}{2}c_3 + \frac{1}{8}c_4\right) z_\alpha \\ &+ \left(-\frac{1}{4}c_1 + \frac{5}{8}c_1 c_2 + \frac{7}{32}c_2^2 + \frac{1}{8}c_4\right) z_\alpha^3 + \frac{1}{8}c_2^2 z_\alpha^5 - \frac{1}{32}c_2^2 z_\alpha^7. \end{aligned} \tag{34}$$

Corollary 2 *For asymptotic chi-square and normal tests,*
(i) the second-order corrected critical values are

$$z_{\alpha,b}^2 = z_\alpha^2 + k_1 b + o(b),\ \ z_{\alpha,b} = z_\alpha + k_3 b + o(b), \tag{35}$$

(ii) the third-order corrected critical values are

$$z_{\alpha,b}^2 = z_\alpha^2 + k_1 b + k_2 b^2 + o(b^2),\ \ z_{\alpha,b} = z_\alpha + k_3 b + k_4 b^2 + o(b^2), \tag{36}$$

where z_α is the nominal critical value from the standard normal distribution.

Our later developments require only the second-order corrected critical values. The third-order corrected critical values are given here because the second-order correction is not enough to deliver a good approximation to the exact critical values when the Parzen and QS kernels are employed and b is large. As shown below, the third-order approximation provides a good general purpose approximation that works well over a wide range of values of b.

For the Bartlett, Parzen and QS kernels, we can compute c_1, c_2, c_3 and c_4 either analytically or numerically. They are given in Table I:

Table I. Values of c's for Different Kernels

	c_1	c_2	c_3	c_4
Bartlett	1.0000	0.6667	−0.3333	−0.1667
Parzen	0.7500	0.5393	−0.1750	−0.0920
QS	1.2500	1.0000	−0.4222	−0.3166

Using Table I, we can obtain the high-order corrected critical values in Table II.

Table II. High Order Corrected Critical Values
$z_{\alpha,b}^2 = z_\alpha^2 + k_1 b + k_2 b^2 + o(b^2)$, $z_{\alpha,b} = z_\alpha + k_3 b + k_4 b^2 + o(b^2)$

	$\alpha = 5\%$, $z_\alpha = 1.960$				$\alpha = 10\%$, $z_\alpha = 1.645$			
	k_1	k_2	k_3	k_4	k_1	k_2	k_3	k_4
Bartlett	10.0414	16.9197	2.5616	2.6423	6.0489	9.7192	1.8386	1.9267
Parzen	7.8964	9.5481	2.0144	1.4006	4.7337	5.5670	1.4388	1.0629
QS	14.1017	38.6840	3.5974	6.5671	8.3968	21.8723	2.5522	4.6682

To evaluate the accuracy of the approximate critical values given in Corollary 2, we compare them with the critical values obtained via simulations. Of course the simulation error can be made arbitrarily small, and so the simulated critical values can be regarded as exact for practical purposes. In our simulations, the Brownian motion and Brownian bridge processes are approximated by normalized partial sums of $T = 1000$ *iid* $N(0,1)$ random variables and the number of replications is 10,000. Comparing the second-order and third-order corrected critical values, we find that the second-order corrected critical value is more accurate for the Bartlett kernel while the third-order corrected critical value is more accurate for the Parzen and QS kernels. Figures 1–3 graph the approximate critical values and the exact critical values as functions of b for the Bartlett, Parzen and QS kernels. Fig. 1 shows that for the Bartlett kernel the second-order corrected critical values are remarkably close to the the exact ones for all values of b. Figs. 2 and 3 show that for the Parzen and QS kernels the third-order corrected critical values provide excellent approximations to the exact ones for all values of b.

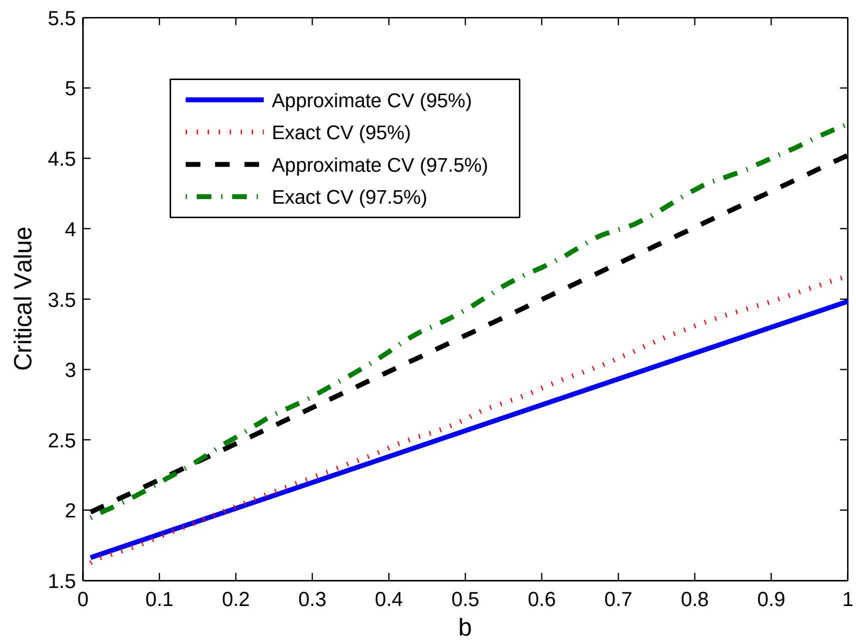

Figure 1: Comparison of the second-order corrected critical values and the exact critical values based on the Bartlett kernel

Since the limiting distributions (12) and (13) are valid for general regression models under certain conditions on the regressors (see PSJ$_a$), we may use the second-order corrected critical values for the Bartlett kernel and the third-order corrected critical values for the Parzen or QS kernel in a general regression framework.

It is important to point out that the exact critical values can be simulated and the high-order asymptotic expansion is developed here for two reasons. First, the resulting high-order corrected critical values are easy to calculate and convenient to use for practical testing situations. In addition, the asymptotic expansion provides a theoretical justifica-

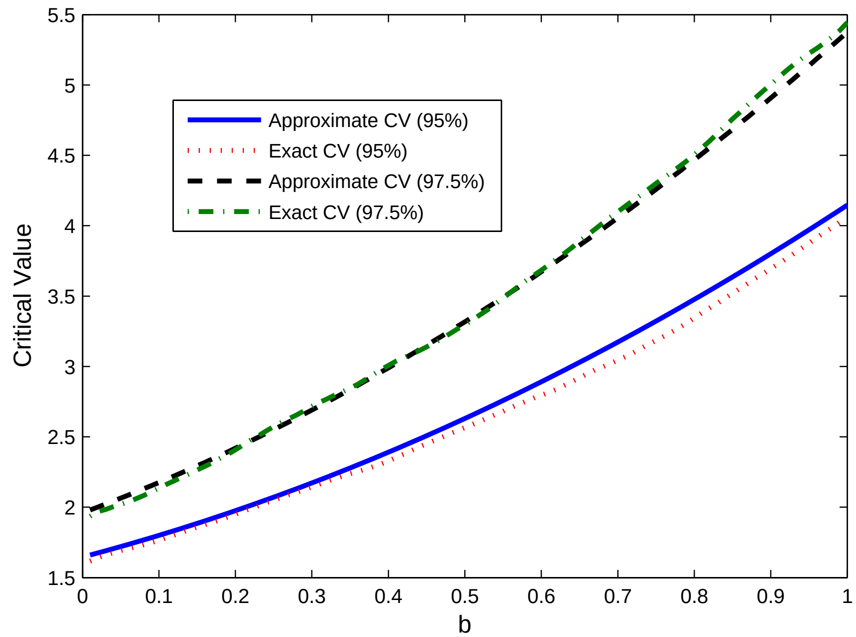

Figure 2: Comparison of the third-order corrected critical values and the exact critical values based on the Parzen kernel

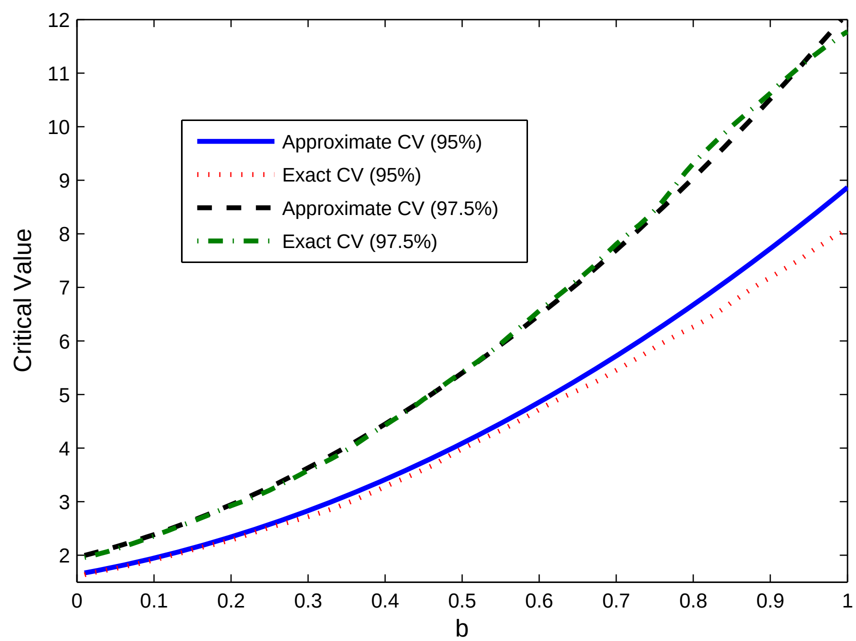

Figure 3: Comparison of the third-order corrected critical values and the exact critical values based on the QS kernel

tion for the polynomial approximation suggested in KV(2005). Second, the asymptotic expansion helps shed some new light on the accuracy of the nonstandard approximation. In the next section, we will show that the second-order corrected critical values based on the asymptotic expansion of the nonstandard distribution is also second-order correct under the conventional asymptotics. As a result, the nonstandard limiting distribution is closer to the exact distribution than the standard normal distribution even when $b \to 0$; See Section 5.

When $\delta \neq 0$ and the second-order corrected critical values are used, we can use Theorem 1 to calculate the local asymptotic power, measured by $P\{\left|(\delta + W(1))\,\Xi_b^{-1/2}\right| > z_{\alpha,b}\}$, as in the following corollary.

Corollary 3 *Let Assumption 2 hold, then the local asymptotic power satisfies*

$$P\left\{\left|(\delta + W(1))\,\Xi_b^{-1/2}\right| > z_{\alpha,b}\right\} = 1 - G_\delta(z_\alpha^2) - c_2 z_\alpha^4 K_\delta\left(z_\alpha^2\right) b + o(b), \tag{37}$$

as $b \to 0$ where

$$K_\delta\left(z\right) = \sum_{j=0}^{\infty} \frac{\left(\delta^2/2\right)^j}{j!} e^{-\delta^2/2} \frac{z^{j-1/2} e^{-z/2}}{\Gamma(j+1/2) 2^{j+1/2}} \frac{j}{z} \tag{38}$$

is positive for all z_α and δ.

According to Corollary 3, the local asymptotic test power, as measured by $P\{\left|(\delta + W(1))\,\Xi_b^{-1/2}\right| > z_{\alpha,b}\}$, decreases monotonically with b at least when b is small. Fig. 4 graphs the surface $f(z_\alpha, \delta) = z_\alpha^4 K_\delta\left(z_\alpha^2\right)$ for different values of z_α and δ. For a given critical value, $f(z_\alpha, \delta)$ achieves its maximum around $\delta = 2$, implying that the power increase resulting from the choice of a small b is greatest when the local alternative is in an intermediate neighborhood of the null hypothesis. For any given local alternative, the function is monotonically increasing in z_α. Therefore, the power improvement due to the choice of a small b increases with the confidence level $1-\alpha$. This is expected. When the confidence level is higher, the test is less powerful and the room for power improvement is greater.

4 Expansions of the Finite Sample Distribution

This section develops a finite sample expansion for the simple location model. This development, like that of Jansson (2004), relies on Gaussianity, which facilitates the derivations. The assumption could be relaxed by taking distributions based (for example) on Gram-Charlier expansions, but at the cost of much greater complexity (see, for example, Phillips (1980), Taniguchi and Puri (1996), Velasco and Robinson (2001)).

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

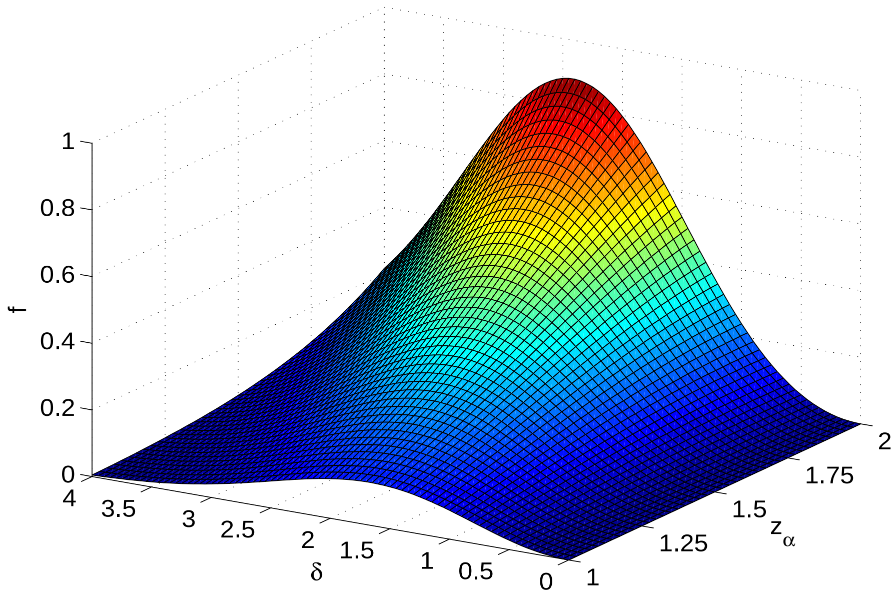

Figure 4: The graph of $f(z_\alpha, \delta) = z_\alpha^4 K_\delta \left(z_\alpha^2\right)$ as a function of z_α and δ.

Assumption 3 *u_t is a mean zero covariance-stationary Gaussian process with $\sum_{h=-\infty}^{\infty} h^2 |\gamma(h)| < \infty$, where $\gamma(h) = Eu_t u_{t-h}$.*

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

Since u_t is in general autocorrelated, $\hat{\beta}$ and $\hat{\omega}_b$ are statistically dependent, which makes it difficult to write down the finite sample distribution of the t-statistic. To tackle this difficulty, we decompose $\hat{\beta}$ and $\hat{\omega}_b$ into statistically independent components. Let $u = (\mu_1, ...u_T)'$, $y = (y_1, ..., y_T)$, $l_T = (1, ..., 1)^T$ and $\Omega_T = var(u)$. Then the GLS estimator of β is $\tilde{\beta} = \left(l_T' \Omega_T^{-1} l_T\right)^{-1} l_T' \Omega_T^{-1} y$ and

$$\hat{\beta} - \beta = \tilde{\beta} - \beta + \left(l_T' l_T\right)^{-1} l_T' \tilde{u}, \tag{39}$$

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 $\tilde{\beta} - \beta$. Since $\hat{\omega}_b^2$ can be written as a quadratic form in $\tilde{u}$, $\hat{\omega}_b^2$ is also statistically independent of $\tilde{\beta} - \beta$. Next, it is easy to see that

$$\omega_T^2 := var\left(\sqrt{T}(\hat{\beta} - \beta)\right) = T^{-1} l_T' \Omega_T l_T = \omega^2 + O\left(T^{-1}\right), \tag{40}$$

and it follows from Grenander and Rosenblatt (1957) that

$$\tilde{\omega}_T^2 := var\left(\sqrt{T}(\tilde{\beta} - \beta)\right) = T\left(l_T' \Omega_T^{-1} l_T\right)^{-1} = \omega^2 + O\left(T^{-1}\right). \tag{41}$$

Therefore $T^{-1/2}l_T'\tilde{u} = N(0, O\left(T^{-1}\right))$. Combining this result with the independence of $\tilde{\beta}$ and $(\tilde{u},\, \hat{\omega}_b)$, we have

$$\begin{aligned}
&P\left\{\sqrt{T}\left(\hat{\beta}-\beta_0\right)/\hat{\omega}_b \leq z\right\} \\
&= P\left\{\sqrt{T}\left(\tilde{\beta}-\beta\right)/\hat{\omega}_b + \sqrt{T}\left(\beta-\beta_0\right)/\hat{\omega}_b + T^{-1/2}l_T'\tilde{u}/\hat{\omega}_b \leq z\right\} \\
&= P\left\{\sqrt{T}\left(\tilde{\beta}-\beta\right)/\tilde{\omega}_T + c/\tilde{\omega}_T \leq z\hat{\omega}_b/\tilde{\omega}_T - T^{-1/2}l_T'\tilde{u}/\tilde{\omega}_T\right\} \\
&= E\Phi\left(z\hat{\omega}_b/\tilde{\omega}_T - c/\tilde{\omega}_T - T^{-1/2}l_T'\tilde{u}/\tilde{\omega}_T\right) \\
&= E\Phi\left(z\hat{\omega}_b/\tilde{\omega}_T - c/\tilde{\omega}_T\right) - T^{-1/2}E\varphi\left(z\hat{\omega}_b/\tilde{\omega}_T - c/\tilde{\omega}_T\right)l_T'\tilde{u}/\tilde{\omega}_T + O\left(T^{-1}\right) \\
&= P\left\{\sqrt{T}\left(\tilde{\beta}-\beta\right)/\tilde{\omega}_T + c/\tilde{\omega}_T \leq z\hat{\omega}_b/\tilde{\omega}_T\right\} + O\left(T^{-1}\right),
\end{aligned} \tag{42}$$

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

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

uniformly over $z \in \mathbb{R}$. Therefore

$$\begin{aligned}
F_{T,\delta}(z) &:= P\left\{\left|\sqrt{T}\left(\hat{\beta}-\beta_0\right)/\hat{\omega}_b\right| \leq z\right\} \\
&= P\left\{\left[\sqrt{T}\left(\tilde{\beta}-\beta\right)/\tilde{\omega}_T + c/\tilde{\omega}_T\right]^2 \leq z^2\hat{\omega}_b^2/\tilde{\omega}_T^2\right\} + O\left(T^{-1}\right) \\
&= E\left\{G_\delta(z^2\hat{\omega}_b^2/\tilde{\omega}_T^2)\right\} = E\left\{G_\delta(z^2\varsigma_{bT})\right\} + O\left(T^{-1}\right),
\end{aligned} \tag{43}$$

uniformly over $z \in \mathbb{R}^+$, where $\varsigma_{bT} := \left(\hat{\omega}_b/\omega_T\right)^2$ converges weakly to Ξ_b.

Setting $\mu_{bT} = E\left(\varsigma_{bT}\right)$ and following the same argument that leads to (18), we have

$$\begin{aligned}
F_{T,\delta}(z) &= G_\delta(\mu_{bT}z^2) + \frac{1}{2}G_\delta''(\mu_{bT}z^2)E\left(\varsigma_{bT}-\mu_{bT}\right)^2 z^4 \\
&\quad + \frac{1}{6}G_\delta'''(\mu_{bT}z^2)E\left(\varsigma_{bT}-\mu_{bT}\right)^3 z^6 + O\left(E\left(\varsigma_{bT}-\mu_{bT}\right)^4\right) + O\left(T^{-1}\right)
\end{aligned} \tag{44}$$

where the $O(\cdot)$ term holds uniformly over $z \in \mathbb{R}^+$. By developing asymptotic expansions of μ_{bT} and $E\left(\varsigma_{bT}-\mu_{bT}\right)^m$ for $m = 1, 2, .., 4$, we can establish a higher order expansion of the finite sample distribution for the case where $T \to \infty$ and $b \to 0$ at the same time. This expansion validates for finite samples the use of the second-order corrected critical values given in the previous section which were derived there on the basis of an expansion of the (nonstandard) limit distribution.

Theorem 4 *Let Assumptions 2 and 3 hold. If $bT \to \infty$ as $T \to \infty$ and $b \to 0$, then*

$$\begin{aligned}
F_{T,\delta}(z) &= G_\delta(z^2) + \left[c_2G_\delta''(\mu_b z^2)z^4 - c_1G_\delta'(z^2)z^2\right]b \\
&\quad - g_q d_{qT}G_\delta'(z^2)z^2(bT)^{-q} + o\left\{b + (bT)^{-q}\right\} + O\left(T^{-1}\right)
\end{aligned} \tag{45}$$

where $d_{qT} = \omega_T^{-2}\sum_{h=-\infty}^{\infty}|h|^q\gamma(h)$, $\omega_T^2 = T^{-1}l_T'\Omega_T l_T$ and the $o(\cdot)$ and $O(\cdot)$ terms hold uniformly over $z \in \mathbb{R}^+$.

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

$$F_{T,0}(z) = D(z^2) + \left[c_2 D''(z^2)z^4 - c_1 D'(z^2)z^2\right] b \\ - g_q d_{qT} D'(z^2) z^2 (bT)^{-q} + o\left\{b + (bT)^{-q}\right\} + O\left(T^{-1}\right). \tag{46}$$

The leading two terms (up to order $O(b)$) in this expansion are the same as those in the corresponding expansion of the limit distribution $F_0(z)$ given in (30) above. Thus, use of the second-order corrected critical values given in (35), which take account of terms up to order $O(b)$, should lead to size improvements when $T^q b^{q+1} \to 0$.

The third term in the expansion (46) is $O(T^{-q})$ when b is fixed. When b decreases with T, this term provides an asymptotic measure of the size distortion in tests based on the use of the first two terms of (46), or equivalently those based on the nonstandard limit theory, at least to order $O(b)$. Thus, the third term of (46) approximately measures how satisfactory the second-order corrected critical values given in (35) are for any given values of b and T.

If critical values from the standard normal are used, then the ERP is approximated by the sum of the $O(b)$ and $O((bT)^{-q})$ terms in 46, viz.,

$$e_T(b) = \left[c_2 D''(z^2)z^4 - c_1 D'(z^2)z^2\right] b - g_q d_{qT} D'(z^2) z^2 (bT)^{-q}, \tag{47}$$

which is a simple rational function in b. In (47) the $O(b)$ term is negative, and the $O((bT)^{-q})$ term is also negative for time series with $d_{qT} > 0$, as is typical for economic data. In this case, the ERP function $e_T(b)$ is negative, has a negative asymptote at $b = 0$, and a simple maximum at

$$b = \left(-\frac{q g_q d_{qT} D'(z^2)}{c_2 D''(z^2) z^2 - c_1 D'(z^2)}\right)^{\frac{1}{q+1}} T^{-\frac{q}{q+1}}. \tag{48}$$

For this choice of b, $e_T(b)$ is closest to zero, and the resulting best rate of convergence of the ERP to zero is $O\left(T^{-q/(q+1)}\right)$. This rate increases with q and can be arbitrarily close to $O\left(T^{-1}\right)$ if q is large and the autocovariance decays at an exponential rate. Examples of kernels with $q > 2$ include the familiar truncated kernel and the flat top kernel proposed by Politis and Romano (1995, 1998). These kernels are not positive semidefinite. We consider only the commonly-used positive semidefinite kernels with $q \leq 2$ in this paper and leave the analysis of higher order kernels for future research.

If we construct a two-sided confidence interval based on the asymptotic normality, then the ERP is also the coverage error. Therefore, the optimal b that achieves the greatest coverage accuracy is of order $O\left(T^{-q/(q+1)}\right)$ for typical economic time series. As we discuss below, this rate is different from $O\left(T^{-2q/(2q+1)}\right)$, the rate that is appropriate for point estimation of the standard error. So, some undersmoothing is required to achieve improved coverage accuracy in confidence intervals.

In the case where $d_{qT} < 0$, the ERP-optimal b is different from that given in (48). When $d_{qT} < 0$, the ERP function $e_T(b)$ has a positive asymptote at $b = 0$ and is monotonically decreasing and passes through the origin as b increases, so the optimal b that achieves the greatest coverage accuracy is the value for which $e_T(b) = 0$, i.e.,

$$b = \left(\frac{g_q d_{qT} D'(z^2)}{c_2 D''(z^2) z^2 - c_1 D'(z^2)}\right)^{\frac{1}{q+1}} T^{-\frac{q}{q+1}}. \tag{49}$$

This choice of b completely removes the $O(b)$ and $O((bT)^{-q})$ terms in (46), leading to a coverage error of a smaller order, i.e. $o\left(T^{-q/(q+1)}\right)$ instead of $O\left(T^{-q/(q+1)}\right)$.

The optimal b given in (48) and (49) is not feasible as it involves the unknown parameter d_{qT}. In practice, we can estimate d_{qT} using a nonparametric approach. A minimal requirement for this estimator is its consistency. Based on the sign of this estimator, we can decide which of the two formulae to use in practical situations. We leave investigations along these lines to future research.

Velasco and Robinson (2001) established an Edgeworth expansion of $P\{t_b \leq z\}$ for the Gaussian local model. Note that $P\{t_b \leq z\} = \frac{1}{2}(1 + F_{T,0}(z))$ for $z \geq 0$ under the null hypothesis, and we have

$$
\begin{aligned}
P\{t_b \leq z\} &= 1 + \frac{1}{2}D(z^2) + \frac{1}{2}\left[c_2 D''(z^2)z^4 - c_1 D'(z^2)z^2\right] b \\
&- \frac{1}{2} g_q d_{qT} D'(z^2) z^2 (bT)^{-q} + o\left\{b + (bT)^{-q}\right\} + O\left(T^{-1}\right). \qquad (50)
\end{aligned}
$$

Using the identity

$$
D'(z^2) = \frac{\phi(z)}{z}, \; D''(z^2) = -\frac{\phi(z)}{2z} - \frac{\phi(z)}{2z^3}, \qquad (51)
$$

we deduce for $z \geq 0$

$$
\begin{aligned}
P\{t_b \leq z\} &= \Phi(z) - \frac{1}{2}\left[\frac{c_2}{2}\left(z^3 + z\right) + c_1 z\right]\phi(z) b \\
&- \frac{1}{2} g_q d_{qT} z \phi(z) (bT)^{-q} + o\left\{b + (bT)^{-q}\right\} + O\left(T^{-1}\right). \qquad (52)
\end{aligned}
$$

Combining this with $P\{t_b \leq -z\} = 1 - P\{t_b \leq z\}$, we can show that the preceding expansion also holds for $z \leq 0$. The expansion (52) is identical to equation (11) in Velasco and Robinson (2001) after notational changes and a small correction[1].

Under the local alternative hypothesis, the power of the test based on the second-order corrected critical values is $1 - F_{T,\delta}(z_{\alpha,b})$. Theorem 4 shows that $F_{T,\delta}(z_{\alpha,b})$ can be approximated by

$$
G_\delta(z_{\alpha,b}^2) + \left[c_2 G_\delta''(z_{\alpha,b}^2) z_{\alpha,b}^4 - c_1 G_\delta'(z_{\alpha,b}^2) z_{\alpha,b}^4\right] b - g_q d_{qT} G_\delta'(z^2) z^2 (bT)^{-q},
$$

with an approximation error of order $o\left((bT)^{-q} + b\right) + O\left(T^{-1}\right)$.

These results on size distortion and local power are formalized in the following corollary.

[1] There is a mistake in Theorem 5 of Velasco and Robinson (2001). Given their Theorem 4 and equation (11), the second order corrected critical value should be

$$
\begin{aligned}
w_\alpha &= z_\alpha - \frac{\int_{-\infty}^{z_\alpha} r_N(x)\phi(x)dx}{\phi(z_\alpha)} \frac{M}{N} \\
&= z_\alpha + \frac{1}{2}\left(z_\alpha^3 - 3z_\alpha\right)\pi \|K\|_2^2 \frac{M}{N} - \frac{1}{2} z_\alpha \left[b_1 N M^{-d-1} - 4\pi \|K\|_2^2 - 2\pi K(0)\right] \frac{M}{N} \\
&= z_\alpha + \frac{1}{2}\left(z_\alpha^3 + z_\alpha\right)\pi \|K\|_2^2 \frac{M}{N} + \frac{1}{2} z_\alpha \{2\pi K(0)\} \frac{M}{N} - \frac{1}{2} z_\alpha b_1 \frac{1}{M^d}
\end{aligned}
$$

Since $2\pi K(0) = c_1$ and $\|K\|_2^2 = c_2$, the correction terms of order $O(M/N)$ are the same as that given in (35) in this paper.

Corollary 5 *Let Assumptions 2 and 3 hold. If $bT \to \infty$ as $T \to \infty$ and $b \to 0$, then:*

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

$$(1 - F_{T,0}(z_{\alpha,b})) - \alpha = g_q d_{qT} D'(z_\alpha^2) z_\alpha^2 (bT)^{-q} + o\left((bT)^{-q} + b\right) + O\left(T^{-1}\right). \tag{53}$$

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

$$\begin{aligned} 1 - F_{T,\delta}(z_{\alpha,b}) &= 1 - G_\delta(z_\alpha^2) - c_2 z_\alpha^4 K_\delta\left(z_\alpha^2\right) b \\ &\quad + g_q d_{qT} G'_\delta(z_\alpha^2) z_\alpha^2 (bT)^{-q} + o\left((bT)^{-q} + b\right) + O\left(T^{-1}\right). \end{aligned} \tag{54}$$

5 Accuracy of the Nonstandard Approximation

In the previous section, we have shown that when b goes to zero at a certain rate, the ERP of the standard normal or chi-squared test is at least of order $O(T^{-q/(q+1)})$ for typical economic time series. This section establishes a related result for the nonstandard test when b is fixed and then compares the ERP of these two tests under the same asymptotic specification, i.e. either b is fixed or $b \to 0$. As in the previous section, we focus on the Gaussian location model.

In view of (43), the error of the nonstandard approximation is given by

$$\begin{aligned} F_{T,0}(z) - F_0(z) &:= P\left\{\left|\sqrt{T}\left(\hat{\beta} - \beta\right)/\hat{\omega}_b\right| \le z\right\} - P\left\{\left|W(1)\Xi_b^{-1/2}\right| \le z\right\} \\ &= ED(z^2\varsigma_{bT}) - ED(z^2\Xi_b) + O\left(T^{-1}\right). \end{aligned} \tag{55}$$

To evaluate the difference $ED(z^2\varsigma_{bT}) - ED(z^2\Xi_b)$, we proceed to compute the cumulants of both $\varsigma_{bT} - \mu_{bT}$ and $\Xi_b - \mu_b$. Since $\hat{\omega}_b^2 = T^{-1}\hat{u}'W_b\hat{u} = T^{-1}u'A_TW_bA_Tu$, where W_b is $T \times T$ with (j,s)-th element $k_b((j-s)/T)$ and $A_T = I_T - l_Tl_T'/T$, ς_{bT} is a quadratic form in a Gaussian vector. It is easy to show that the characteristic function of $\varsigma_{bT} - \mu_{bT}$ is given by

$$\phi_{bT}(t) = \left|I - 2it\frac{\Omega_T A_T W_b A_T}{T\omega_T^2}\right|^{-1/2} \exp\left\{-it\mu_{bT}\right\}, \tag{56}$$

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

$$\ln\left(\phi_{bT}(t)\right) = -\frac{1}{2}\log\det\left(I - 2it\frac{\Omega_T A_T W_b A_T}{T\omega_T^2}\right) - it\mu_{bT} := \sum_{m=1}^{\infty}\kappa_{m,T}\frac{(it)^m}{m!}, \tag{57}$$

where the $\kappa_{m,T}$ are the cumulants of $\varsigma_{bT} - \mu_{bT}$. It follows from (57) that $\kappa_{1,T} = 0$ and

$$\kappa_{m,T} = 2^{m-1}(m-1)!T^{-m}\left(\omega_T^2\right)^{-m}\text{Trace}\left[(\Omega_T A_T W_b A_T)^m\right] \text{ for } m \ge 2. \tag{58}$$

By proving $\kappa_{m,T}$ is close to κ_m in the precise sense given in Lemma 2 in the Appendix, we can establish the following theorem, which gives the order of magnitude of the error in the nonstandard limit distribution of t_b as $T \to \infty$ with fixed b. The requirement $b < 1/(16 \int_{-\infty}^{\infty} |k(x)| dx)$ on b that appears in the statement of the result is a technical condition in the proof that facilitates the use of a power series expansion. The requirement can be relaxed but at the cost of more extensive and tedious calculations.

Theorem 6 *Let Assumptions 2 and 3 hold. If* $b < 1/(16 \int_{-\infty}^{\infty} |k(x)| dx)$, *then*

$$F_{T,0}(z) = F_0(z) + O\left(T^{-1}\right), \tag{59}$$

uniformly over $z \in \mathbb{R}^+$ *when* $T \to \infty$ *with fixed* b.

Under the null hypothesis $H_0 : \beta = \beta_0$, we have $\delta = 0$. In this case, Theorem 6 indicates that the ERP for tests with b fixed and using critical values obtained from the nonstandard limit distribution of $W(1)\Xi_b^{-1/2}$ is $O\left(T^{-1}\right)$. The theorem is an extension to the result of Jansson (2004) who considered only the Bartlett-type kernel with $b = 1$ and proved that the ERP is of order $O\left(T^{-1}\log(T)\right)$. It is an open question in Jansson (2004) whether the $\log(T)$ factor can be omitted. Theorem 6 provides a positive answer to this question.

In the previous section, we showed that when $b \to 0, T \to \infty$ such that $bT \to \infty$,

$$F_{T,0}(z) = D(z^2) + O\left(T^{-q/(q+1)}\right) \tag{60}$$

for typical economic time series. Comparing (59) with (60), one may conclude that the error of the nonstandard approximation is smaller than that of the standard normal approximation by an order of magnitude. However, the two $O(\cdot)$ terms are obtained under different asymptotic specifications. The $O(\cdot)$ term in (59) holds for fixed b while the $O(\cdot)$ term in (60) holds for diminishing b. Since the $O(\cdot)$ term in (59) does not hold uniformly over $b \in (0, 1]$, the two $O(\cdot)$ terms can not be directly compared, although they are obviously suggestive of the relative quality of the two approximations when b is small, as it typically will be in practical applications.

Indeed, $F_0(z)$ and $D(z^2)$ are just different approximations to the same quantity $F_{T,0}(z)$. To compare the two approximations more formally, we need to evaluate $F_{T,0}(z) - F_0(z)$ and $F_{T,0}(z) - D(z^2)$ under the same asymptotic specification, i.e. either b is fixed or $b \to 0$.

First, when b is fixed, we have

$$F_{T,0}(z) - D(z^2) = F_{T,0}(z) - F_0(z) + F_0(z) - D(z^2) = O(1). \tag{61}$$

This is because $F_{T,0}(z) - F_0(z) = O(1/T)$ as shown in Theorem 6 and $F_0(z) - D(z^2) = O(1)$. Comparing (61) with (59), we conclude that when b is fixed, the error of the nonstandard approximation is smaller than that of the standard approximation by an order of magnitude.

Second, when $b = O(T^{-q/(q+1)})$, we have

$$\begin{aligned} F_{T,0}(z) - F_0(z) &= \left[F_{T,0}(z) - D(z^2)\right] - \left[F_0(z) - D(z^2)\right] \\ &= \left[c_2 D''(z^2)z^4 - c_1 D'(z^2)z^2\right] b \\ &- g_q d_{qT} D'(z^2)z^2(bT)^{-q} + o\left\{b + (bT)^{-q}\right\} + O\left(T^{-1}\right) \\ &- \left[c_2 D''(z^2)z^4 - c_1 D'(z^2)z^2\right] b + o(b) \\ &= -g_q d_{qT} D'(z^2)z^2(bT)^{-q} + o\left\{b + (bT)^{-q}\right\} + O\left(T^{-1}\right), \end{aligned} \tag{62}$$

where we have used Theorems 1 and 4. Therefore, when $d_{qT} > 0$, which is typical for economic time series, the error of the nonstandard approximation is smaller than that of the standard normal approximation, although they are of the same order of magnitude for this choice of b.

We can conclude from the above analysis that the nonstandard distribution provides a more accurate approximation to the finite sample distribution regardless of the asymptotic specification employed. There are two reasons for the better performance: the nonstandard distribution mimics the randomness of the denominator of the t-statistic, and it accounts for the bias of the LRV estimator resulting from the unobservability of the regressor errors. As a result, the critical values from the nonstandard limiting distribution provide a higher order correction on the critical values from the standard normal distribution. However, just as in the standard limiting theory, the nonstandard limiting theory does not deal with another source of bias, i.e. the usual bias that arises in spectral density estimation even when a time series is known to be mean zero and observed. This second source of bias manifests itself in the error of approximation given in (62).

6 Optimal Bandwidth Choice

It is well known that the optimal choice of b that minimizes the asymptotic mean squared error in LRV estimation has the form $b = O\left(T^{-2q/(2q+1)}\right)$. However, there is no reason to expect that such a choice is the most appropriate in statistical testing using nonparametrically studentized statistics. Developing an optimal choice of b for semiparametric testing is not straightforward and involves some conceptual as well as technical challenges. In what follows we provide one possible approach to constructing an optimizing criterion that is based on balancing the type I and type II errors.

In view of the asymptotic expansion (53), we know that the type I error for a nominal size α test can be expressed as

$$1 - F_{T,0}(z_{\alpha,b}) = \alpha + g_q d_{qT} D'(z_\alpha^2)z_\alpha^2(bT)^{-q} + o\left((bT)^{-q} + b\right) + O\left(T^{-1}\right), \tag{63}$$

Similarly, from (54), the type II error has the form

$$G_\delta(z_\alpha^2) + c_2 z_\alpha^4 K_\delta(z_\alpha^2)b - g_q d_{qT} G'_\delta(z_\alpha^2)z_\alpha^2(bT)^{-q} + o\left((bT)^{-q} + b\right) + O\left(T^{-1}\right). \tag{64}$$

A loss function for the test may be constructed based on the following three factors: (i) The magnitude of the type I error, as measured by the second term of (63); (ii) The magnitude of the type II error, as measured by the $O(b)$ and $O((bT)^{-q})$ terms in (64); and (iii) The relative importance of the type I and type II errors.

For most economic time series we can expect that $d_{qT} > 0$ and then both $g_q d_{qT} D'(z_\alpha^2) z_\alpha^2 > 0$ and $g_q d_{qT} G'_\delta(z_\alpha^2) z_\alpha^2 > 0$. Hence, the type I error increases as b decreases. On the other hand, the $(bT)^{-q}$ term in (64) indicates that there is a corresponding decrease in the type II error as b decreases. Indeed, for $\delta > 0$ the decrease in the type II error will generally exceed the increase in the type I error because $G'_\delta(z_\alpha^2) > D'(z_\alpha^2)$ for $\delta \in (0, 7.5)$ and $z_\alpha = 1.645, 1,960$ or 2.580. Fig. 5 graphs the ratio $G'_\delta(z^2)/D'(z^2)$ against δ for different values of z, illustrating the relative magnitude of $G'_\delta(z_\alpha^2)$ and $D'(z_\alpha^2)$. The situation is further complicated by the fact that there is an additional $O(b)$ term in the type II error. As we have seen earlier, $K_\delta(z_\alpha^2) > 0$ so that the second term of (64) leads to a reduction in the type II error as b decreases. Thus, the type II error generally decreases with b for two reasons — one from the nonstandard limit theory and the other from the (typical) downward bias in estimating the long-run variance.

The case of $d_{qT} < 0$ usually arises where there is negative serial correlation in the errors and so tends to be less typical for economic time series. In such a case, (63) shows that the type I error decreases with b while the type II error may increase or decrease with b depending on which of the two terms in (64) dominates.

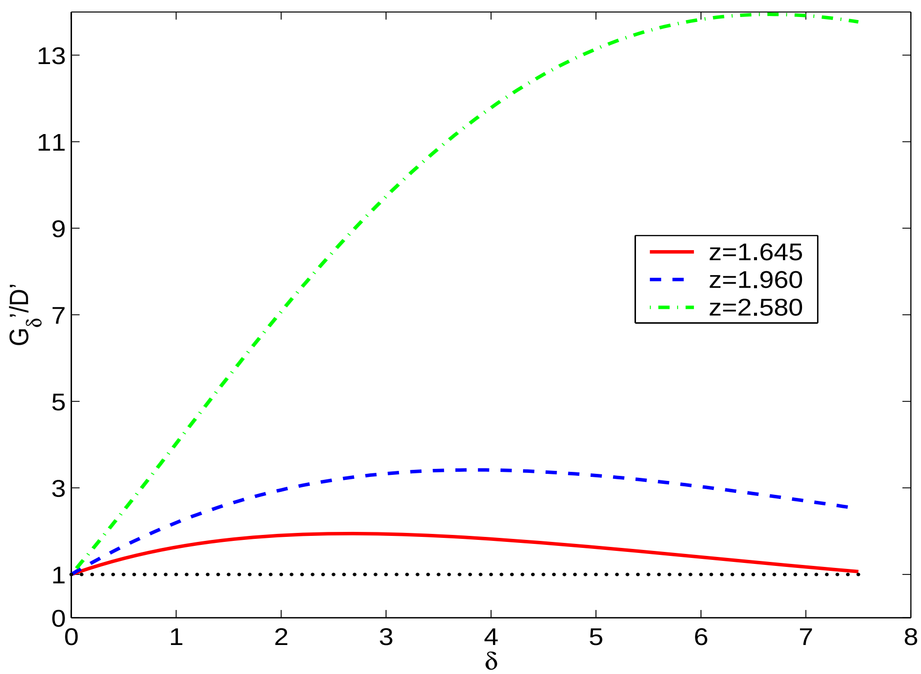

Figure 5: The graph of $G'_\delta(z^2)/D'_\delta(z^2)$ as a function of δ for different values of z

These considerations suggest that a loss function may be constructed by taking a suitable weighted average of the type I and type II errors given in (63) and (64). Setting

$$\begin{aligned} e_T^I &= \alpha + g_q d_{qT} D'(z^2) z^2 (bT)^{-q}, \\ e_T^{II} &= G_\delta(z_\alpha^2) + c_2 z_\alpha^4 K_\delta(z_\alpha^2) b - g_q d_{qT} G'_\delta(z_\alpha^2) z_\alpha^2 (bT)^{-q}, \end{aligned} \tag{65}$$

we define the loss function to be

$$L(b; \delta, T, z_\alpha) = \frac{w_T(\delta)}{1 + w_T(\delta)} e_T^I + \frac{1}{1 + w_T(\delta)} e_T^{II}, \tag{66}$$

where $w_T(\delta)$ is a function that determines the relative weight on the type I and II errors and this function is allowed to depend on the sample size T and δ. Obviously, the loss $L(b; \delta, T, z_\alpha)$ is here specified for a particular value of δ and this function could be adjusted in a simple way so that the type II error is averaged over a range of values of δ with respect to some (prior) distribution over alternatives.

We focus on the case of a fixed local alternative below, in which case we can suppress the dependence of $w_T(\delta)$ on δ and write $w_T = w_T(\delta)$. To sum up, the loss function we consider is of the form:

$$L(b; \delta, T, z_\alpha) = \left[g_q d_{qT} \left\{ w_T D'(z_\alpha^2) - G'_\delta(z_\alpha^2) \right\} z_\alpha^2 (bT)^{-q} + c_2 z_\alpha^4 K_\delta(z_\alpha^2) b \right] \frac{1}{1 + w_T} + C_T, \tag{67}$$

where $C_T = \left[w_T \alpha + G_\delta(z_\alpha^2) \right] / [1 + w_T]$, which does not depend on b. In the rest of this section, we consider the case $w_T D'(z_\alpha^2) - G'_\delta(z_\alpha^2) > 0$, which holds if the relative weight w_T is large enough.

It turns out that the optimal choice of b depends on whether $d_{qT} > 0$ or $d_{qT} < 0$. We consider these two cases in turn. When $d_{qT} > 0$, the loss function $L(b; \delta, T, z_\alpha)$ is minimized for the following choice of b :

$$b_{\text{opt}} = \left\{ \frac{q g_q d_{qT} \left[w_T D'(z_\alpha^2) - G'_\delta(z_\alpha^2) \right]}{c_2 z_\alpha^2 K_\delta(z_\alpha^2)} \right\}^{1/(q+1)} T^{-q/(q+1)}. \tag{68}$$

Therefore, the optimal shrinkage rate for b is of order $O(T^{-q/(q+1)})$ when w_T is a fixed constant. If $w_T \to \infty$ as $T \to \infty$, we then have

$$b_{\text{opt}} = \left\{ \frac{q g_q d_{qT} D'(z_\alpha^2)}{z_\alpha^2 c_2 K_\delta(z_\alpha^2)} \right\}^{1/(q+1)} (w_T / T^q)^{1/(q+1)}. \tag{69}$$

Fixed b rules may then be interpreted as assigning relative weight $w_T = O(T^q)$ in the loss function so that the emphasis in tests based on such rules is a small type I error, at least when we expect the type I error to be larger than the nominal size of the test. This gives us an interpretation of fixed b rules in terms of the loss perceived by the econometrician using such a rule. Within the more general framework given by (68), b may be fixed or shrink with T up to an $O\left(T^{-q/(q+1)}\right)$ rate corresponding to the relative importance that is placed in the loss function on the type I and type II errors.

Observe that when $b = O((w_T/T^q)^{1/(q+1)})$, size distortion is $O(w_T T)^{-q/(q+1)}$ rather than $O(T^{-1})$, as it is when b is fixed. Thus, the use of $b = b_{\text{opt}}$ for a finite w_T involves some compromise by allowing the error order in the rejection probability to be somewhat larger in order to achieve higher power. Such compromise is an inevitable consequence of balancing the two elements in the loss function (67). Note that even in this case, the order of ERP is smaller than $O(T^{-q/(2q+1)})$, which is the order of the ERP for the conventional procedure in which standard normal critical values are used and b is set to be $O(T^{-2q/(2q+1)})$.

For Parzen and QS kernels the optimal rate of b is $T^{-2/3}$, whereas for the Bartlett kernel the optimal rate is $T^{-1/2}$. Therefore, in large samples, a smaller b should be used with the Parzen and QS kernels than with the Bartlett kernel. In the former cases, the ERP is at most of order $T^{-2/3}$ and in the latter case the ERP is at most of order $T^{-1/2}$. The $O(T^{-2/3})$ rate of the ERP for the quadratic kernels represents an improvement on the Bartlett kernel. Note that for the optimal b, the rate of the ERP is also the rate for which the loss function $L(b;\delta,T,z_\alpha)$ approaches C_T from above. Therefore, the loss $L(b;\delta,T,z_\alpha)$ is expected to be smaller for quadratic kernels than for the Bartlett kernel in large samples. Finite sample performance may not necessarily follow this ordering, however, and will depend on the sample size and the shape of the spectral density of $\{u_t\}$ at the origin.

The formula for b_{opt} involves the unknown parameter d_{qT}, which could be estimated nonparametrically (e.g. Newey and West (1994)) or by a standard plug-in procedure based on a simple model like AR(1) (e.g. Andrews (1991)). Both methods achieve a valid order of magnitude and the procedure is obviously analogous to conventional data-driven methods for HAC estimation.

When $w_T D'(z_\alpha^2) - G'_\delta(z_\alpha^2) > 0$ and $d_{qT} < 0$, $L(b;\delta,T,z_\alpha)$ is an increasing function of b. To minimize loss in this case, we can choose b to be as small as possible. Since the loss function is constructed under the assumption that $b \to 0$ and $bT \to \infty$, the choice of b is required to be compatible with these two rate conditions. These considerations lead to a choice of b of the form $b = J_T/T$ for some J_T that goes to infinity but at a slowly varying rate relative to T. In the simulation study below, we set $J_T = \log(T)$ so that $b = (\log T)/T$.

To sum up, for typical economic time series, the value of b which minimizes the weighted type I and type II errors has a shrinkage rate of $b = O(T^{-q/(q+1)})$. This rate may be compared with the optimal rate of $b = O(T^{-2q/(2q+1)})$ that applies when minimizing the mean squared error of estimation of the corresponding HAC estimate, $\hat{\omega}_b^2$, itself (Andrews (1991)). Thus, the AMSE optimal values of b for HAC estimation are smaller as $T \to \infty$ than those which are most suited for statistical testing. In effect, optimal HAC estimation tolerates more bias in order to reduce variance in estimation. In contrast, optimal b selection in HAC testing undersmooths the long-run variance estimate to reduce bias and allows for greater variance in long-run variance estimation through higher order adjustments to the nominal asymptotic critical values or by direct use of the nonstandard limit distribution.

7 Simulation Evidence

This section provides some simulation evidence on the finite sample performance of the t-test based on the plug-in procedure that optimizes the loss function constructed in the previous section.

We consider the simple local model with Gaussian ARMA(1,1) errors:

$$y_t = \beta + c/\sqrt{T} + u_t \tag{70}$$

where $c = 0$ or 2 and

$$u_t = \phi u_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}, \varepsilon_t \sim iidN(0,1). \tag{71}$$

Under the null $c = 0$, and under the local alternative $c = 2$. The latter value of c is chosen such that when $\phi = 0.3$ and $\theta = 0$, the local asymptotic power of the t-test is 60%. In other words, $c = 2$ solves $P\{|(c(1-\phi) + W(1))| \geq 1.645\} = 60\%$ for $\phi = 0.3$. The qualitative results are similar for other nonzero values of c. We consider three sample sizes $T = 50, 100$ and 200.

For the ARMA(1,1) process,

$$d_1 = \frac{2(1+\phi\theta)(\phi+\theta)}{(1-\phi^2)(1+\theta)^2},\ d_2 = \frac{2(1+\phi\theta)(\phi+\theta)}{(1-\phi)^2(1+\theta)^2}. \tag{72}$$

To estimate d_1 and d_2, we employ the AR(1) plug-in procedure. More specifically, let

$$\hat{\rho} = \frac{\sum_{t=2}^{T}(u_t - \bar{u})(u_{t-1} - \bar{u})}{\sum_{t=2}^{T}(u_{t-1} - \bar{u})^2} \tag{73}$$

be the OLS estimate of the AR parameter, then we estimate d_1 and d_2 by

$$\hat{d}_1 = 2\hat{\rho}/(1-\hat{\rho}^2),\ \hat{d}_2 = 2\hat{\rho}/(1-\hat{\rho})^2. \tag{74}$$

It is easy to see that

$$\hat{\rho} \to \frac{(\theta+\phi)(1+\theta\phi)}{\theta^2 + 2\theta\phi + 1},\ \text{as } T \to \infty. \tag{75}$$

Therefore

$$\begin{aligned} d_1/\hat{d}_1 &\to \frac{(1-\phi)\left(\theta^2 - (1-\phi)\theta + 1\right)}{\theta^2 + 2\theta\phi + 1}\left\{\left(1-\phi^2\right)(1+\theta)^2\right\}^{-1} > 0 \\ d_2/\hat{d}_2 &\to \left(1 - \frac{(\theta+\phi)(1+\theta\phi)}{\theta^2 + 2\theta\phi + 1}\right)^2 \left\{(1-\phi)^2(1+\theta)^2\right\}^{-1} > 0. \end{aligned} \tag{76}$$

Since both limits are positive when $|\phi| < 1$, $\hat{d}_q$ and d_q have the same sign with probability approaching one as $T \to \infty$. We can thus choose b based on the sign of $\hat{d}_q$. More specifically,

$$\hat{b}_{\text{opt}} = \begin{cases} \left[\frac{q g_q \hat{d}_q\left\{w_T D'(z_\alpha^2) - G'_\delta(z_\alpha^2)\right\}}{c_2 z_\alpha^2 K_\delta(z_\alpha^2)}\right]^{1/(q+1)} T^{-q/(q+1)}, & \hat{d}_q > 0 \\ T^{-1}\log T, & \hat{d}_q < 0 \end{cases} \tag{77}$$

We consider the following relative weights: $w_T = 10, 20, 30$, and 40. We set the significance level to be $\alpha = 10\%$ and the corresponding nominal critical value for the two-sided test is $z_\alpha = 1.645$. For all the DGP's considered, we let $\delta = 2$ in computing the optimal bandwidth parameter. For each choice of w_T, we obtain $\hat{b}_{\text{opt}}$ and use it to construct the LRV estimate and corresponding $t_{\hat{b}}$-statistic. We reject the null hypothesis if $\left|t_{\hat{b}}\right|$ is larger than the corrected critical values given in (35) and (36). For the Bartlett kernel, the second-order corrected critical value is used. For the Parzen and QS kernels, the third-order corrected critical value is used. Using 50,000 replications, we compute the empirical type I error (when $c = 0$) and type II error (when $c = 2$). We construct the empirical loss by taking a weighted average of the type I and type II errors. The weights associated with the type I and II errors are $w_T/(1 + w_T)$ and $1/(1 + w_T)$, respectively.

For comparative purposes, we also compute the empirical loss function when the bandwidth is the 'optimal' one that minimizes the asymptotic mean squared errors of the LRV estimate. This bandwidth rule is given in Andrews (1991) and the AR(1) plug-in version is

$$\begin{aligned}\hat{b}_{MSE}^{BT} &= 1.1447\left(\frac{4\hat{\rho}^2}{(1-\hat{\rho})^2(1+\hat{\rho})^2}\right)^{1/3} T^{-2/3},\\ \hat{b}_{MSE}^{PR} &= 2.6614\left(\frac{4\hat{\rho}^2}{(1-\hat{\rho})^4}\right)^{1/5} T^{-4/5}, \text{ or}\\ \hat{b}_{MSE}^{QS} &= 1.3221\left(\frac{4\hat{\rho}^2}{(1-\hat{\rho})^4}\right)^{1/5} T^{-4/5}.\end{aligned} \tag{78}$$

Tables III-V report the empirical loss only for the sample size $T = 100$, as it is representative of other sample sizes.

First, we compare the new plug-in procedure with the conventional plug-in procedure for a given kernel. It is clear that the new plug-in procedure incurs significantly smaller loss than the conventional plug-in procedure when $d_q > 0$, which is typical for economic time series. This is true for all values of w_T and parameter combinations considered. When $d_q \leq 0$, the performance of the new plug-in procedure is better than that of the conventional plug-in procedure for both the Bartlett and Parzen kernels. However, for the QS kernels, the former is slightly outperformed by the latter for some DGP's.

Second, we compare the performance of the new plug-in procedure based on different kernels. For all values of w_T and the DGP's, the Parzen and QS kernels lead to more or less the same loss. This is not surprising as both Parzen and QS exhibit quadratic behavior around the origin. Compared with the Bartlett kernel, the Parzen and QS kernels incur smaller losses for almost all the cases. For the few exceptional cases, the Bartlett kernel outperforms the Parzen and QS kernels by only a small margin. This result is consistent with the faster rate of convergence of the loss function for the Parzen and QS kernels as compared with the Bartlett kernel.

To sum up, our simulation results reveal that the new plug-in procedure works well in terms of incurring a smaller loss than the conventional plug-in procedure. Compared with the Bartlett kernel, the Parzen and QS kernels both seem to offer the prospect of smaller loss.

8 Concluding Discussion

Automatic bandwidth choice is a long standing problem in time series models when the autocorrelation is of unknown form. Existing automatic methods are all based on minimizing the asymptotic mean square error of the standard error estimator, a criterion that is not directed at statistical testing. In hypothesis testing, the focus of attention is the type I and type II errors that arise in testing and it is these errors that give rise to loss. Consequently, it is desirable to make the errors of incorrectly rejecting a true null hypothesis and failing to reject a false null hypothesis as small as possible. While these two types of errors may not be simultaneously reduced, it is possible to design bandwidth choice to control the loss from these errors. This paper develops for the first time a theory of optimal bandwidth choice that achieves this end by minimizing a weighted average of the type I and type II errors.

The type I and type II errors are measured by the first dominating terms in the asymptotic expansions of the distribution of the test under the null and alternative hypotheses. The rule for selecting the optimal bandwidth (M) according to the above principle has an expansion rate of $O\left(T^{1/(q+1)}\right)$ for typical economic time series. This rate is slower than the rate $O\left(T^{1/(2q+1)}\right)$ for optimizing the asymptotic mean squared error in HAC estimation. Thus, optimal bandwidth selection for semiparametric testing is different from optimal bandwidth selection for point estimation of the long-run variance. Semiparametric testing along these lines actually undersmooths the long-run variance estimate to reduce bias and allows for greater variance in long-run variance estimation as it is manifested in the test statistic by means of higher order adjustments to the nominal asymptotic critical values or by direct use of the nonstandard limit distribution.

The asymptotic expansions of the finite sample distribution of $\hat{\beta}$ could be extended to the regression model of the form: $y_t = \beta + x_t'\gamma + u_t$ where x_t is a strongly exogenous and stationary mean zero vector process. In this case, the OLS and GLS estimators of β satisfy $var\left(\sqrt{T}(\hat{\beta}-\beta) - \sqrt{T}(\tilde{\beta}-\beta)\right) = O(1/T)$ and $\sqrt{T}(\hat{\beta}-\beta) - \sqrt{T}(\tilde{\beta}-\beta)$ is independent of $\sqrt{T}(\tilde{\beta}-\beta)$. These properties ensure that $F_{T,\delta}(z) = E\left\{G_\delta(z^2\varsigma_{\rho T})\right\} + O(1/T)$, a crucial step in establishing the asymptotic expansions. Replacing u by $u^* = (I - X(X'X)^{-1}X')u$ for $X = (x_1, ..., x_T)'$ in Assumption 2 and using the same proofs, we can establish the asymptotic expansions in Section 4 conditioning on X.

The results in this paper suggest areas of future research. Using asymptotic expansion formulae, we have derived the optimal bandwidth that achieves the highest coverage accuracy. It is desirable to investigate the finite sample performance of this procedure when a consistent nonparametric estimate of d_{qT} is plugged into the optimal bandwidth formula. In this paper, we have focused on the Gaussian location model. The basic ideas and methods explored here can be used to tackle the bandwidth choice problem for nonparametric studentized testing and confidence interval construction in a general regression setting.

Table III: Empirical Loss Using Different Plug-in b's for AR(1) Process with AR Parameter ϕ (T =100, Number of replication =50000)

	$w_T = 10$		$w_T = 20$		$w_T = 30$		$w_T = 40$	
	b_{opt}	b_{MSE}	b_{opt}	b_{MSE}	b_{opt}	b_{MSE}	b_{opt}	b_{MSE}
				Bartlett kernel				
0.9	0.2707	0.3682	0.2478	0.3547	0.2393	0.3500	0.2345	0.3475
0.6	0.1941	0.2395	0.1685	0.2186	0.1606	0.2112	0.1577	0.2074
0.3	0.1621	0.1878	0.1429	0.1722	0.1373	0.1666	0.1342	0.1637
−0.3	0.0912	0.1006	0.0851	0.0960	0.0830	0.0944	0.0819	0.0935
−0.6	0.0564	0.0902	0.0531	0.0905	0.0520	0.0906	0.0514	0.0907
−0.9	0.0294	0.0914	0.0181	0.0940	0.0141	0.0949	0.0121	0.0953
				Parzen kernel				
0.9	0.2122	0.3621	0.1756	0.3480	0.1624	0.3430	0.1556	0.3404
0.6	0.1737	0.2265	0.1417	0.2040	0.1309	0.1960	0.1244	0.1919
0.3	0.1530	0.1786	0.1301	0.1617	0.1220	0.1556	0.1175	0.1525
−0.3	0.0956	0.1038	0.0904	0.0998	0.0886	0.0984	0.0876	0.0977
−0.6	0.0777	0.0904	0.0772	0.0911	0.0770	0.0913	0.0769	0.0914
−0.9	0.0658	0.0709	0.0670	0.0725	0.0674	0.0730	0.0676	0.0733
				QS kernel				
0.9	0.2067	0.3516	0.1689	0.3364	0.1545	0.3310	0.1457	0.3282
0.6	0.1724	0.2242	0.1398	0.2014	0.1280	0.1933	0.1218	0.1892
0.3	0.1521	0.1783	0.1292	0.1613	0.1209	0.1552	0.1163	0.1521
−0.3	0.1053	0.1021	0.1008	0.0979	0.0992	0.0964	0.0984	0.0956
−0.6	0.0938	0.0904	0.0944	0.0911	0.0946	0.0914	0.0947	0.0915
−0.9	0.0865	0.0858	0.0890	0.0884	0.0899	0.0893	0.0903	0.0898

Table IV: Empirical Loss Using Different Plug-in b's for an MA(1) Process with MA parameter θ (T = 100, Number of replications = 50,000)

θ	$w_T = 10$		$w_T = 20$		$w_T = 30$		$w_T = 40$	
	b_{opt}	b_{MSE}	b_{opt}	b_{MSE}	b_{opt}	b_{MSE}	b_{opt}	b_{MSE}
				Bartlett kernel				
0.9	0.1722	0.1911	0.1483	0.1687	0.1422	0.1607	0.1395	0.1567
0.6	0.1654	0.1821	0.1447	0.1624	0.1387	0.1555	0.1363	0.1519
0.3	0.1526	0.1664	0.1361	0.1509	0.1317	0.1455	0.1290	0.1427
−0.3	0.0736	0.0789	0.0688	0.0747	0.0670	0.0733	0.0661	0.0725
−0.6	0.0095	0.0185	0.0090	0.0189	0.0088	0.0190	0.0087	0.0191
−0.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
				Parzen kernel				
0.9	0.1620	0.1885	0.1335	0.1657	0.1235	0.1575	0.1185	0.1534
0.6	0.1557	0.1789	0.1307	0.1587	0.1219	0.1515	0.1168	0.1478
0.3	0.1463	0.1612	0.1258	0.1449	0.1183	0.1391	0.1147	0.1361
−0.3	0.0741	0.0802	0.0693	0.0762	0.0676	0.0748	0.0668	0.0741
−0.6	0.0099	0.0136	0.0094	0.0135	0.0093	0.0135	0.0092	0.0135
−0.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
				QS kernel				
0.9	0.1608	0.1849	0.1327	0.1616	0.1219	0.1533	0.1165	0.1491
0.6	0.1548	0.1753	0.1302	0.1546	0.1208	0.1473	0.1155	0.1435
0.3	0.1456	0.1595	0.1250	0.1430	0.1178	0.1371	0.1139	0.1341
−0.3	0.0930	0.0769	0.0903	0.0724	0.0893	0.0708	0.0888	0.0700
−0.6	0.0433	0.0119	0.0452	0.0115	0.0459	0.0114	0.0462	0.0113
−0.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table V: Empirical Loss Using Different Plug-in b's for an ARMA(1,1) Process with AR parameter ϕ and MA parameter θ (T =100, Number of replications =50,000)

	$w_T = 10$		$w_T = 20$		$w_T = 30$		$w_T = 40$	
(ϕ, θ)	b_{opt}	b_{MSE}	b_{opt}	b_{MSE}	b_{opt}	b_{MSE}	b_{opt}	b_{MSE}
				Bartlett kernel				
$(-0.6, 0.3)$	0.0996	0.1126	0.0926	0.1075	0.0901	0.1056	0.0888	0.1047
$(0.3, -0.6)$	0.0392	0.0392	0.0356	0.0350	0.0344	0.0335	0.0337	0.0327
$(0.3, 0.3)$	0.1754	0.2019	0.1520	0.1814	0.1461	0.1741	0.1436	0.1704
$(0.0, 0.0)$	0.1261	0.1312	0.1152	0.1208	0.1119	0.1171	0.1096	0.1152
$(0.6, -0.3)$	0.1815	0.2362	0.1591	0.2214	0.1508	0.2161	0.1468	0.2134
$(-0.3, 0.6)$	0.1473	0.1551	0.1321	0.1400	0.1278	0.1346	0.1256	0.1318
				Parzen kernel				
$(-0.6, 0.3)$	0.1064	0.1168	0.1005	0.1124	0.0984	0.1109	0.0974	0.1101
$(0.3, -0.6)$	0.0355	0.0371	0.0310	0.0329	0.0294	0.0313	0.0286	0.0306
$(0.3, 0.3)$	0.1627	0.1957	0.1343	0.1742	0.1245	0.1665	0.1195	0.1626
$(0.0, 0.0)$	0.1244	0.1319	0.1123	0.1215	0.1083	0.1178	0.1061	0.1159
$(0.6, -0.3)$	0.1700	0.2208	0.1420	0.2039	0.1307	0.1979	0.1257	0.1948
$(-0.3, 0.6)$	0.1422	0.1522	0.1234	0.1365	0.1165	0.1309	0.1131	0.1281
				QS kernel				
$(-0.6, 0.3)$	0.1111	0.1155	0.1053	0.1110	0.1032	0.1093	0.1022	0.1085
$(0.3, -0.6)$	0.0596	0.0347	0.0587	0.0296	0.0583	0.0277	0.0581	0.0268
$(0.3, 0.3)$	0.1619	0.1920	0.1333	0.1700	0.1224	0.1622	0.1173	0.1582
$(0.0, 0.0)$	0.1267	0.1308	0.1145	0.1202	0.1099	0.1165	0.1080	0.1146
$(0.6, -0.3)$	0.1695	0.2237	0.1402	0.2072	0.1293	0.2013	0.1242	0.1983
$(-0.3, 0.6)$	0.1416	0.1500	0.1234	0.1341	0.1165	0.1284	0.1128	0.1255

9 Appendix

A.1 Technical Lemmas and Supplements

Lemma 1 *Let* $c_1^* = 4\int_{-\infty}^{\infty} |k(v)|\, dv$, *then under Assumption 2 the cumulants of* $\Xi_b - \mu_b$ *satisfy*

$$|\kappa_m| \le 2^m (m-1)!\, (c_1^* b)^{m-1} \text{ for } m \ge 1 \tag{A.1}$$

and the moments $\alpha_m = E\,(\Xi_b - \mu_b)^m$ *satisfy*

$$|\alpha_m| \le 2^{2m} m!\, (c_1^* b)^{m-1} \text{ for } m \ge 1. \tag{A.2}$$

Proof of Lemma 1. Note that

$$\begin{aligned}
&\left| \int_0^1 \dots \int_0^1 \left(\prod_{j=1}^m k_b^*(\tau_j, \tau_{j+1}) \right) d\tau_1 \cdots d\tau_m \right| \\
&\le \int_0^1 \dots \int_0^1 |k_b^*(\tau_1,\tau_2) k_b^*(\tau_2,\tau_3) \cdots k_b^*(\tau_{m-1},\tau_m)|\, |k_b^*(\tau_m,\tau_1)|\, d\tau_1 \cdots d\tau_m \\
&\le 2 \int_0^1 \dots \int_0^1 |k_b^*(\tau_1,\tau_2) k_b^*(\tau_2,\tau_3) \cdots k_b^*(\tau_{m-1},\tau_m)|\, d\tau_1 \cdots d\tau_m \\
&\le 2 \sup_{\tau_2} \int_0^1 |k_b^*(\tau_1,\tau_2)|\, d\tau_1 \int_0^1 |k_b^*(\tau_2,\tau_3) k_b^*(\tau_3,\tau_4) \cdots k_b^*(\tau_{m-1},\tau_m)|\, d\tau_2 \cdots d\tau_m \\
&\le 2 \sup_{\tau_2} \int_0^1 |k_b^*(\tau_1,\tau_2)|\, d\tau_1 \sup_{\tau_3} \int_0^1 |k_b^*(\tau_2,\tau_3)|\, d\tau_2 \dots \sup_{\tau_m} \int_0^1 [k_b^*(\tau_{m-1},\tau_m)]\, d\tau_{m-1} \\
&= 2 \left(\sup_s \int_0^1 |k_b^*(r,s)|\, dr \right)^{m-1}.
\end{aligned} \tag{A.3}$$

In view of the definition

$$k_b^*(r,s) = k_b(r-s) - \int_0^1 k_b(r-p)dp - \int_0^1 k_b(s-q)dq + \int_0^1 \int_0^1 k_b(p-q)dpdq, \tag{A.4}$$

we have

$$\begin{aligned}
&\sup_s \int_0^1 |k_b^*(r,s)|\, dr \le 4 \sup_s \int_0^1 |k_b(r-s)|\, dr \\
&= 4 \sup_{s\in[0,1]} \left(\int_{-s}^{1-s} |k_b(v)|\, dv \right) \le 4 \int_{-\infty}^{\infty} |k_b(v)|\, dv \\
&= b c_1^*.
\end{aligned} \tag{A.5}$$

As a result

$$\left| \int_0^1 \dots \int_0^1 \left(\prod_{j=1}^m k_b^*(\tau_j, \tau_{j+1}) \right) d\tau_1 \cdots d\tau_m \right| \le 2\, (c_1^* b)^{m-1}, \tag{A.6}$$

and

$$|\kappa_m| \leq 2^m (m-1)!\,(c_1^* b)^{m-1}\,. \tag{A.7}$$

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

$$\alpha_m = \sum_{\pi} \frac{m!}{(j_1!)^{m_1}(j_2!)^{m_2}\cdots(j_\ell!)^{m_\ell}} \frac{1}{m_1!m_2!\cdots m_\ell!} \prod_{j\in\pi} \kappa_j, \tag{A.8}$$

where the sum is taken over the elements

$$\pi = [\underbrace{j_1,\cdots j_1}_{m_1 \text{ times}}, \underbrace{j_2,\cdots j_2}_{m_2 \text{ times}}, \cdots \underbrace{j_\ell,\cdots j_\ell}_{m_\ell \text{ times}}] \tag{A.9}$$

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

Combining the preceding formula with (A.7) gives

$$\begin{aligned} |\alpha_m| &< 2^m m!\,(c_1^* b)^{m-1} \sum_{\pi} \frac{(j_1)^{-m_1}(j_2)^{-m_2}\cdots(j_\ell)^{-m_\ell}}{m_1!m_2!\cdots m_\ell!} \\ &\leq 2^{2m} m!\,(c_1^* b)^{m-1}\,, \end{aligned} \tag{A.10}$$

where the last line follows because

$$\sum_{\pi} \frac{(j_1)^{-m_1}(j_2)^{-m_2}\cdots(j_\ell)^{-m_\ell}}{m_1!m_2!\cdots m_\ell!} \leq \sum_{\pi} \frac{1}{m_1!m_2!\cdots m_\ell!} < 2^m. \tag{A.11}$$

■

Lemma 2 *Let Assumptions 2 and 3 hold. When $T \to \infty$ for a fixed b, we have:*

(a)

$$\mu_{bT} = \mu_b + O\left(\frac{1}{T}\right); \tag{A.12}$$

(b)

$$\kappa_{m,T} = \kappa_m + O\left\{\frac{m!2^m}{T^2}(c_1^* b)^{m-2}\right\}, \tag{A.13}$$

uniformly over $m \geq 1$;.

(c)

$$\alpha_{m,T} = E(\varsigma_{bT} - \mu_{bT})^m = \alpha_m + O\left\{\frac{m!2^{2m}}{T^2}(c_1^* b)^{m-2}\right\}, \tag{A.14}$$

uniformly over $m \geq 1$.

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

$$\tilde{k}_b\left(\frac{i}{T},\frac{j}{T}\right) = k_b\left(\frac{i-j}{T}\right) - \frac{1}{T}\sum_{p=1}^{T} k_b\left(\frac{i-p}{T}\right) - \frac{1}{T}\sum_{q=1}^{T} k_b\left(\frac{q-j}{T}\right) + \frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T} k_b\left(\frac{p-q}{T}\right). \tag{A.15}$$

So

$$\begin{aligned}
&\text{Trace}\left(\Omega_T A_T W_b A_T\right) = \text{Trace}\left(\Omega_T W_b^*\right) \\
&= \sum_{1\le r_1,r_2\le T}\left\{\gamma(r_1-r_2)\tilde{k}_b\left(\frac{r_1}{T},\frac{r_2}{T}\right)\right\} = \sum_{r_2=1}^{T}\sum_{h_1=1-r_2}^{T-r_2}\gamma(h_1)\tilde{k}_b\left(\frac{r_2+h_1}{T},\frac{r_2}{T}\right) \\
&= \left(\sum_{h_1=1}^{T-1}\sum_{r_2=1}^{T-h_1} + \sum_{h_1=1-T}^{0}\sum_{r_2=1-h_1}^{T}\right)\gamma(h_1)\tilde{k}_b\left(\frac{r_2+h_1}{T},\frac{r_2}{T}\right).
\end{aligned} \tag{A.16}$$

But

$$\begin{aligned}
&\sum_{r_2=1}^{T-h_1}\tilde{k}_b\left(\frac{r_2+h_1}{T},\frac{r_2}{T}\right) = \sum_{r_2=1}^{T-h_1} k_b\left(\frac{h_1}{T}\right) - \frac{1}{T}\sum_{r_1=1+h_1}^{T}\sum_{p=1}^{T} k_b\left(\frac{r_1-p}{T}\right) \\
&- \frac{1}{T}\sum_{r_2=1}^{T-h_1}\sum_{q=1}^{T} k_b\left(\frac{q-r_2}{T}\right) + \sum_{r_2=1}^{T-h_1}\frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T} k_b\left(\frac{p-q}{T}\right) \\
&= -\frac{1}{T}\sum_{r_1=1}^{T}\sum_{p=1}^{T} k_b\left(\frac{r_1-p}{T}\right) - \frac{1}{T}\sum_{r_2=1}^{T}\sum_{q=1}^{T} k_b\left(\frac{q-r_2}{T}\right) \\
&+ \sum_{r_2=1}^{T}\frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T} k_b\left(\frac{p-q}{T}\right) + Tk_b\left(\frac{h_1}{T}\right) + C(h_1) \\
&= -\frac{1}{T}\sum_{r=1}^{T}\sum_{s=1}^{T} k_b\left(\frac{r-s}{T}\right) + Tk_b\left(\frac{h_1}{T}\right) + C(h_1) \\
&= \sum_{r_2=1}^{T}\tilde{k}_b\left(\frac{r_2}{T},\frac{r_2}{T}\right) + T\left\{k_b\left(\frac{h_1}{T}\right) - k_b\left(0\right)\right\} + C(h_1),
\end{aligned} \tag{A.17}$$

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

$$\sum_{r_2=1-h_1}^{T}\tilde{k}_b\left(\frac{r_2+h_1}{T},\frac{r_2}{T}\right) = \sum_{r_2=1}^{T}\tilde{k}_b\left(\frac{r_2}{T},\frac{r_2}{T}\right) + T\left\{k_b\left(\frac{h_1}{T}\right) - k_b\left(0\right)\right\} + C(h_1). \tag{A.18}$$

Therefore, Trace$(\Omega_T A_T W_b A_T)$ is equal to

$$\sum_{h=-T+1}^{T-1}\gamma(h)\sum_{r_2=1}^{T}\tilde{k}_b\left(\frac{r_2}{T},\frac{r_2}{T}\right)+T\sum_{h=-T+1}^{T-1}\gamma(h)\left\{k_b\left(\frac{h}{T}\right)-k_b\left(0\right)\right\}+O\left(1\right)$$
$$=\sum_{h=-T+1}^{T-1}\gamma(h)\sum_{r_2=1}^{T}\tilde{k}_b\left(\frac{r_2}{T},\frac{r_2}{T}\right)+T(bT)^{-q}\sum_{h=-T+1}^{T-1}|h|^q\gamma(h)\left\{\frac{k\left(h/(bT)\right)-k\left(0\right)}{|h/(bT)|^q}\right\}+O\left(1\right)$$
$$=\sum_{h=-T+1}^{T-1}\gamma(h)\sum_{r_2=1}^{T}\tilde{k}_b\left(\frac{r_2}{T},\frac{r_2}{T}\right)+T(bT)^{-q}g_q\sum_{h=-\infty}^{\infty}|h|^q\gamma(h)(1+o(1))+O(1). \qquad (A.19)$$

Using

$$\sum_{h=-T+1}^{T-1}\gamma(h)=\omega_T^2(1+O(\frac{1}{T})), \qquad (A.20)$$

and

$$\frac{1}{T}\sum_{r_2=1}^{T}\tilde{k}_b\left(\frac{r_2}{T},\frac{r_2}{T}\right)=\int_0^1 k_b^*(r,r)dr+O(\frac{1}{T}), \qquad (A.21)$$

we now have

$$\mu_{bT}=\int_0^1 k_b^*(r,r)dr-(bT)^{-q}g_q\left(\omega_T^{-2}\sum_{h=-\infty}^{\infty}|h|^q\gamma(h)\right)(1+o(1))+O\left(\frac{1}{T}\right). \qquad (A.22)$$

By definition, $\mu_b=E\Xi_b=\int_0^1 k_b^*(r,r)dr$ and thus $\mu_{bT}=\mu_b+O\left(T^{-1}\right)$ as desired.

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

$$Trace\left[(\Omega_T A_T W_b A_T)^m\right]$$
$$=\sum_{r_1,r_2,\ldots,r_{2m+1}=1}^{T}\prod_{j=1}^{m}\gamma(r_{2j-1}-r_{2j})\tilde{k}_b\left(\frac{r_{2j}}{T},\frac{r_{2j+1}}{T}\right)$$
$$=\sum_{r_2,r_4,\ldots,r_{2m}=1}^{T}\sum_{h_1=1-r_2}^{T-r_2}\sum_{h_2=1-r_4}^{T-r_4}\cdots\sum_{h_m=1-r_{2m}}^{T-r_{2m}}\prod_{j=1}^{m}\gamma(h_j)\tilde{k}_b\left(\frac{r_{2j}}{T},\frac{r_{2j+2}+h_{j+1}}{T}\right)$$
$$=\left(\sum_{h_1=1}^{T-1}\sum_{r_2=1}^{T-h_1}+\sum_{h_1=1-T}^{0}\sum_{r_2=1-h_1}^{T}\right)\cdots\left(\sum_{h_m=1}^{T-1}\sum_{r_{2m}=1}^{T-h_m}+\sum_{h_m=1-T}^{0}\sum_{r_{2m}=1-h_m}^{T}\right)$$
$$\prod_{j=1}^{m}\gamma(h_j)\tilde{k}_b\left(\frac{r_{2j}}{T},\frac{r_{2j+2}+h_{j+1}}{T}\right)$$
$$=I+II, \qquad (A.23)$$

where

$$I = \left(\sum_{h_1=1}^{T-1} \sum_{r_2=1}^{T-h_1} + \sum_{h_1=1-T}^{0} \sum_{r_2=1-h_1}^{T} \right) \cdots \left(\sum_{h_m=1}^{T-1} \sum_{r_{2m}=1}^{T-h_m} + \sum_{h_m=1-T}^{0} \sum_{r_{2m}=1-h_m}^{T} \right)$$
$$\prod_{j=1}^{m} \gamma(h_j) \tilde{k}_b \left(\frac{r_{2j}}{T}, \frac{r_{2j+2}}{T} \right), \tag{A.24}$$

and

$$II = O\left\{ \left(\sum_{h_1=1}^{T-1} \sum_{r_2=1}^{T-h_1} + \sum_{h_1=1-T}^{0} \sum_{r_2=1-h_1}^{T} \right) \cdots \left(\sum_{h_m=1}^{T-1} \sum_{r_{2m}=1}^{T-h_m} + \sum_{h_m=1-T}^{0} \sum_{r_{2m}=1-h_m}^{T} \right) \right.$$
$$\left. \prod_{j=1}^{m} |\gamma(h_j)| \left(\frac{|h_{j+1}|}{bT} \right) \right\}. \tag{A.25}$$

Here we have used

$$\left| \tilde{k}_b \left(\frac{r_{2j}}{T}, \frac{r_{2j+2} + h_{j+1}}{T} \right) - \tilde{k}_b \left(\frac{r_{2j}}{T}, \frac{r_{2j+2}}{T} \right) \right| = O\left(\frac{|h_{j+1}|}{bT} \right). \tag{A.26}$$

To show this, note that

$$\frac{1}{T} \sum_{p=1}^{T} k_b \left(\frac{p - r_{2j+2} - h_{j+1}}{T} \right) = \frac{1}{T} \sum_{p=1-h_{j+1}}^{T-h_{j+1}} k_b \left(\frac{p - r_{2j+2}}{T} \right)$$
$$= \frac{1}{T} \sum_{p=1}^{T} k_b \left(\frac{p - r_{2j+2}}{T} \right) + O\left(\frac{|h_{j+1}|}{T} \right), \tag{A.27}$$

and

$$\left| k_b \left(\frac{r_{2j} - r_{2j+2} - h_{j+1}}{T} \right) - k_b \left(\frac{r_{2j} - r_{2j+2}}{T} \right) \right| = O\left(\frac{|h_{j+1}|}{bT} \right), \tag{A.28}$$

so that

$$\tilde{k}_b \left(\frac{r_{2j}}{T}, \frac{r_{2j+2} + h_{j+1}}{T} \right)$$
$$= \tilde{k}_b \left(\frac{r_{2j}}{T}, \frac{r_{2j+2}}{T} \right) + k_b \left(\frac{r_{2j} - r_{2j+2} - h_{j+1}}{T} \right) - k_b \left(\frac{r_{2j} - r_{2j+2}}{T} \right) + O(\frac{|h_{j+1}|}{T})$$
$$= \tilde{k}_b \left(\frac{r_{2j}}{T}, \frac{r_{2j+2}}{T} \right) + O\left(\frac{|h_{j+1}|}{bT} \right). \tag{A.29}$$

The first term (I) can be written as

$$
\begin{aligned}
I =& \left(\sum_{h_1=1-T}^{T-1} \sum_{r_2=1}^{T} - \sum_{h_1=1}^{T-1} \sum_{r_2=T-h_1+1}^{T} - \sum_{h_1=1-T}^{0} \sum_{r_2=1}^{-h_1} \right) \cdots \\
& \left(\sum_{h_m=1-T}^{T-1} \sum_{r_{2m}=1}^{T} - \sum_{h_m=1}^{T-1} \sum_{r_{2m}=T-h_m+1}^{T} - \sum_{h_m=1-T}^{0} \sum_{r_{2m}=1}^{-h_m} \right) \prod_{j=1}^{m} \gamma(h_j) \left\{ \tilde{k}_b \left(\frac{r_{2j}}{T}, \frac{r_{2j+2}}{T} \right) \right\} \\
=& \sum_{\pi} \sum_{h_1,r_2} \cdots \sum_{h_m,r_{2m}} \prod_{j=1}^{m} \gamma(h_j) \left\{ \tilde{k}_b \left(\frac{r_{2j}}{T}, \frac{r_{2j+2}}{T} \right) \right\}, \qquad \text{(A.30)}
\end{aligned}
$$

where $\sum_{h_j,r_{2j}}$ is one of the three choices $\sum_{h_j=1-T}^{T-1} \sum_{r_{2j}=1}^{T}$, $-\sum_{h_j=1}^{T-1} \sum_{r_{2j}=T-h_j+1}^{T}$, $-\sum_{h_j=1-T}^{0} \sum_{r_{2j}=1}^{-h_j}$ and $\sum_{\pi}$ is the summation over all possible combinations of $\left(\sum_{h_1,r_2} \cdots \sum_{h_m,r_{2m}}\right)$. The 3^m summands in (A.30) can be divided into two groups with the first group consisting of the summands all of whose r indices run from 1 to T and the second group consisting of the rest. It is obvious that the first group can be written as

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

The dominating terms (in terms of the order of magnitude) in the second group are of the forms

$$
\sum_{h_1=1-T}^{T-1} \sum_{r_2=1}^{T} \cdots \sum_{h_p=1-T}^{T-1} \sum_{r_{2p}=T-h_p+1}^{T} \cdots \sum_{h_m=1-T}^{T-1} \sum_{r_{2m}=1}^{T} \prod_{j=1}^{m} \gamma(h_j) \left\{ \tilde{k}_b \left(\frac{r_{2j}}{T}, \frac{r_{2j+2}}{T} \right) \right\},
$$

or

$$
\sum_{h_1=1-T}^{T-1} \sum_{r_2=1}^{T} \cdots \sum_{h_p=1-T}^{T-1} \sum_{r_{2p}=1}^{-h_p} \cdots \sum_{h_m=1-T}^{T-1} \sum_{r_{2m}=1}^{T} \prod_{j=1}^{m} \gamma(h_j) \left\{ \tilde{k}_b \left(\frac{r_{2j}}{T}, \frac{r_{2j+2}}{T} \right) \right\}.
$$

These are the summands with only one r index not running from 1 to T. Both of the above terms are bounded by

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

using the same approach as in (A.3). Approximating the sum by an integral and noting that the second group contains $(m-1)$ terms, all of which are of the same order of magnitude as the above typical dominating terms, we conclude that the second group is of order $O\left[2mT^{m-2}\left(c_1^*b\right)^{m-2}\right]$ uniformly over m. As a consequence,

$$I=\left(\sum_h\prod_{j=1}^{m}\gamma(h_j)\right)\sum_r\left\{\tilde{k}_b\left(\frac{r_{2j}}{T},\frac{r_{2j+2}}{T}\right)\right\}+O\left\{2mT^{m-2}\left(c_1^*b\right)^{m-2}\right\} \tag{A.31}$$

uniformly over m.

The second term (II) is easily shown to be of order $o\left(2mT^{m-2}\left(c_1^*b\right)^{m-2}\right)$ uniformly over m. Therefore

$$\begin{aligned}&Trace\left[(\Omega_TA_TW_bA_T)^m\right]\\&=\left(\sum_h\gamma(h)\right)^m\sum_r\left\{\tilde{k}_b\left(\frac{r_{2j}}{T},\frac{r_{2j+2}}{T}\right)\right\}+O\left\{2mT^{m-2}\left(c_1^*b\right)^{m-2}\right\}\end{aligned} \tag{A.32}$$

and

$$\begin{aligned}\kappa_{m,T}&=2^{m-1}(m-1)!T^{-m}\left(\omega_T^2\right)^{-m}\text{Trace}\left[(\Omega_TA_TW_bA_T)^m\right]\\&=2^{m-1}(m-1)!\left\{T^{-m}\sum_r\tilde{k}_b\left(\frac{r_{2j}}{T},\frac{r_{2j+2}}{T}\right)+O\left[\frac{2m}{T^2}\left(c_1^*b\right)^{m-2}\right]\right\}\\&=2^{m-1}(m-1)!\left\{\int\prod_{j=1}^{m}\int_0^1k_b^*(\tau_j,\tau_{j+1})d\tau_jd\tau_{j+1}+O\left[\frac{2m}{T^2}\left(c_1^*b\right)^{m-2}\right]\right\}\\&=\kappa_m+O\left\{\frac{m!2^m}{T^2}\left(c_1^*b\right)^{m-2}\right\},\end{aligned} \tag{A.33}$$

uniformly over m.

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

$$\alpha_{m,T}=\sum_\pi\frac{m!}{(j_1!)^{m_1}(j_2!)^{m_2}\cdots(j_k!)^{m_k}}\frac{1}{m_1!m_2!\cdots m_k!}\prod_{j\in\pi}\kappa_{j,T} \tag{A.34}$$

where the summation $\sum_\pi$ is defined in (A.8). Combining the preceding formula with part (b) gives

$$\begin{aligned}\alpha_{m,T}&=\alpha_m+O\left\{\frac{2^m}{T^2}\left(c_1^*b\right)^{m-2}\sum_\pi\frac{m!}{m_1!m_2!\cdots m_k!}\right\}\\&=\alpha_m+O\left\{\frac{m!2^{2m}}{T^2}\left(c_1^*b\right)^{m-2}\right\},\end{aligned} \tag{A.35}$$

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

Lemma 3 *Let Assumptions 2 and 3 hold. If $b \to 0$ and $T \to \infty$ such that $bT \to \infty$, then:*
(a)

$$\mu_{bT} = \int_0^1 k_b^*(r,r)dr - (bT)^{-q} g_q \left(\omega_T^{-2} \sum_{h=-\infty}^{\infty} |h|^q \gamma(h) \right) (1 + o(1)) + O\left(\frac{1}{T}\right); \tag{A.36}$$

(b)

$$\kappa_{2,T} = 2 \int_0^1 \int_0^1 \left(k_b^*(r,s)\right)^2 drds(1 + o(1)) + O\left(\frac{1}{T}\right); \tag{A.37}$$

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

$$\kappa_{m,T} = O\left(b^{m-1}\right) + O\left(\frac{1}{T}\right). \tag{A.38}$$

Proof of Lemma 3. We have proved (A.36) in the proof of Lemma 2 as equation (A.22) holds for both fixed b and decreasing b. It remains to consider $\kappa_{m,T}$ for $m = 2$, 3, and 4. We first consider $\kappa_{2,T} = 2T^{-2}\left(\omega_T^{-4}\right)\text{Trace}\left[(\Omega_T A_T W_b A_T)^2\right]$. As a first step, we have

$$\begin{aligned}
&\text{Trace}\left[(\Omega_T A_T W_b A_T)^2\right] \\
&= \sum_{r_1,r_2,r_3,r_4} \left\{ \tilde{k}_b\left(\frac{r_2}{T}, \frac{r_3}{T}\right) \tilde{k}_b\left(\frac{r_4}{T}, \frac{r_1}{T}\right) \right\} \gamma(r_1 - r_2)\gamma(r_3 - r_4) \\
&= \sum_{r_2=1}^{T} \sum_{h_1=1-r_2}^{T-r_2} \sum_{r_4=1}^{T} \sum_{h_2=1-r_4}^{T-r_4} \left\{ \tilde{k}_b\left(\frac{r_2}{T}, \frac{r_4+h_2}{T}\right) \tilde{k}_b\left(\frac{r_4}{T}, \frac{r_2+h_1}{T}\right) \right\} \gamma(h_1)\gamma(h_2) \\
&= \left(\sum_{h_1=1}^{T-1} \sum_{r_2=1}^{T-h_1} + \sum_{h_1=1-T}^{0} \sum_{r_2=1-h_1}^{T} \right) \left(\sum_{h_2=1}^{T-1} \sum_{r_4=1}^{T-h_2} + \sum_{h_2=1-T}^{0} \sum_{r_4=1-h_2}^{T} \right) \\
&\left\{ \tilde{k}_b\left(\frac{r_2}{T}, \frac{r_4+h_2}{T}\right) \tilde{k}_b\left(\frac{r_4}{T}, \frac{r_2+h_1}{T}\right) \right\} \gamma(h_1)\gamma(h_2) \\
&:= I_1 + I_2 + I_3 + I_4,
\end{aligned} \tag{A.39}$$

where

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

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

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

and

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

We now consider each term in turn. Using equation (A.29), we have

$$\tilde{k}_b \left(\frac{r_2}{T}, \frac{r_4+h_2}{T} \right) = \tilde{k}_b \left(\frac{r_2}{T}, \frac{r_4}{T} \right) + O\left(\frac{b\,|h_2|}{T} \right), \tag{A.40}$$

and

$$\tilde{k}_b \left(\frac{r_4}{T}, \frac{r_2+h_1}{T} \right) = \tilde{k}_b \left(\frac{r_4}{T}, \frac{r_2}{T} \right) + O\left(\frac{b\,|h_1|}{T} \right). \tag{A.41}$$

It follows from (A.40) and (A.41) that

$$\begin{aligned}
I_1 &= \sum_{h_1=1}^{T-1} \sum_{h_2=1}^{T-1} \sum_{r_2=1}^{T-1} \sum_{r_4=1}^{T-1} \left\{ \tilde{k}_b \left(\frac{r_2}{T}, \frac{r_4+h_2}{T} \right) \tilde{k}_b \left(\frac{r_4}{T}, \frac{r_2+h_1}{T} \right) \right\} \gamma(h_1)\gamma(h_2) \\
&\quad + O\left(\sum_{h_1=1}^{T-1} \sum_{h_2=1}^{T-1} \left[T\left(|h_1|+|h_2|\right) + |h_1 h_2| \right] |\gamma(h_1)\gamma(h_2)| \right) \\
&= \sum_{h_1=1}^{T-1} \sum_{h_2=1}^{T-1} \sum_{r_2=1}^{T-1} \sum_{r_4=1}^{T-1} \left\{ \tilde{k}_b \left(\frac{r_2}{T}, \frac{r_4+h_2}{T} \right) \tilde{k}_b \left(\frac{r_4}{T}, \frac{r_2+h_1}{T} \right) \right\} \gamma(h_1)\gamma(h_2) + O(T) \\
&= \sum_{h_1=1}^{T-1} \sum_{h_2=1}^{T-1} \sum_{r_2=1}^{T-1} \sum_{r_4=1}^{T-1} \left\{ \tilde{k}_b \left(\frac{r_2}{T}, \frac{r_4}{T} \right) \tilde{k}_b \left(\frac{r_4}{T}, \frac{r_2}{T} \right) \right\} \gamma(h_1)\gamma(h_2) + O(T) \\
&\quad + O\left\{ \sum_{h_1=1}^{T-1} \sum_{h_2=1}^{T-1} \sum_{r_2=1}^{T-1} \sum_{r_4=1}^{T-1} \left| \tilde{k}_b \left(\frac{r_2}{T}, \frac{r_4}{T} \right) \right| \left(\frac{b\left(|h_1|+|h_2|\right)}{T} \right) |\gamma(h_1)\gamma(h_2)| \right\} \\
&= \sum_{h_1=1}^{T-1} \sum_{h_2=1}^{T-1} \sum_{r_2=1}^{T-1} \sum_{r_4=1}^{T-1} \left\{ \tilde{k}_b \left(\frac{r_2}{T}, \frac{r_4}{T} \right) \tilde{k}_b \left(\frac{r_4}{T}, \frac{r_2}{T} \right) \right\} \gamma(h_1)\gamma(h_2) + O(T).
\end{aligned} \tag{A.42}$$

Following the same procedure, we can show that

$$I_2 = \sum_{h_1=1}^{T-1} \sum_{r_2=1}^{T} \sum_{h_2=1-T}^{0} \sum_{r_4=1}^{T} \left\{ \tilde{k}_b \left(\frac{r_2}{T}, \frac{r_4}{T} \right) \tilde{k}_b \left(\frac{r_4}{T}, \frac{r_2}{T} \right) \right\} \gamma(h_1)\gamma(h_2) + O(T), \tag{A.43}$$

$$I_3 = \sum_{h_1=1-T}^{0} \sum_{r_2=1}^{T} \sum_{r_2=1}^{T} \sum_{r_4=1}^{T} \left\{ \tilde{k}_b \left(\frac{r_2}{T}, \frac{r_4}{T} \right) \tilde{k}_b \left(\frac{r_4}{T}, \frac{r_2}{T} \right) \right\} \gamma(h_1)\gamma(h_2) + O(T), \tag{A.44}$$

and

$$I_4 = \sum_{h_1=1-T}^{0} \sum_{r_2=1}^{T} \sum_{h_2=1-T}^{0} \sum_{r_4=1}^{T} \left\{ \tilde{k}_b \left(\frac{r_2}{T}, \frac{r_4}{T} \right) \tilde{k}_b \left(\frac{r_4}{T}, \frac{r_2}{T} \right) \right\} \gamma(h_1)\gamma(h_2) + O(T). \tag{A.45}$$

As a consequence,

$$\begin{aligned}&\text{Trace}\left[(\Omega_T A_T W_b A_T)^2\right]\\&=\sum_{r_2,r_4}\left\{\tilde{k}_b\left(\frac{r_2}{T},\frac{r_4}{T}\right)\right\}^2\left(\sum_{h=1-T}^{T-1}\gamma(h_1)\right)^2+O\left(T\right),\end{aligned}\tag{A.46}$$

and

$$\begin{aligned}\kappa_{2,T}&=2T^{-2}\left(\omega_T^{-4}\right)Trace\left[(\Omega_T A_T W_b A_T)^2\right]\\&=2\int_0^1\int_0^1\left(k_b^*(r,s)\right)^2 drds(1+o(1))+O\left(\frac{1}{T}\right).\end{aligned}\tag{A.47}$$

The proof for $\kappa_{m,T}$ for $m=3$ and 4 is essentially the same except that we use Lemma 1 to obtain the first term $O(b^{m-1})$. The details are omitted. ■

A.2 Proofs of the Main Results

Proof of Theorem 1. It follows from Lemma 1 that

$$|\alpha_4|=O(b^3),\ \ |\alpha_3|\le|\alpha_4|^{3/4}=O(b^{9/4})=o(b^2).\tag{A.48}$$

As a consequence,

$$\begin{aligned}F_\delta(z)&=P\left\{\left|(W(1)+\delta)\Xi_b^{-1/2}\right|<z\right\}\\&=G_\delta(\mu_b z^2)+\frac{1}{2}\left(G_\delta''(\mu_b z^2)z^4\right)\alpha_2+o(b^2),\end{aligned}\tag{A.49}$$

uniformly over $z\in z\in\mathbb{R}^+$ where

$$\mu_b=E\Xi_b=\int_0^1 k_b^*(r,r)dr=1-\int_0^1\int_0^1 k_b(r-s)drds,\tag{A.50}$$

and

$$\begin{aligned}\alpha_2&=2\left(\int_0^1\int_0^1 k_b(r-s)drds\right)^2+2\int_0^1\int_0^1 k_b^2(r-s)drds\\&-4\int_0^1\int_0^1\int_0^1 k_b(r-p)k_b(r-q)drdpdq.\end{aligned}\tag{A.51}$$

We first develop an asymptotic expansion of μ_b and α_2 as $b\to 0$. Let

$$\mathcal{K}_1\left(\lambda\right)=\frac{1}{2\pi}\int_{-\infty}^{\infty}k(x)\exp(-i\lambda x)dx,\ \mathcal{K}_2\left(\lambda\right)=\frac{1}{2\pi}\int_{-\infty}^{\infty}k^2(x)\exp(-i\lambda x)dx,\tag{A.52}$$

then

$$k(x)=\int_{-\infty}^{\infty}\mathcal{K}_1\left(\lambda\right)\exp(i\lambda x)d\lambda,\ k^2(x)=\int_{-\infty}^{\infty}\mathcal{K}_2\left(\lambda\right)\exp(i\lambda x)d\lambda.\tag{A.53}$$

For the integral that appears in both μ_b and α_2, we have

$$\begin{aligned}
&\int_0^1 \int_0^1 k_b(r-s)drds \\
&= \int_{-\infty}^{\infty} \mathcal{K}_1(\lambda) \int_0^1 \int_0^1 \exp(\frac{i\lambda(r-s)}{b})drdsd\lambda \\
&= \int_{-\infty}^{\infty} \mathcal{K}_1(\lambda) \left(\int_0^1 \exp(\frac{i\lambda r}{b})dr\right)\left(\int_0^1 \exp(-\frac{i\lambda s}{b})ds\right) \\
&= \int_{-\infty}^{\infty} \mathcal{K}_1(\lambda) \left(\frac{b^2}{\lambda^2}\left(\left(1-\cos\left(\frac{\lambda}{b}\right)\right)^2 + \left(\sin\left(\frac{\lambda}{b}\right)\right)^2\right)\right) d\lambda \\
&= b\int_{-\infty}^{\infty} \mathcal{K}_1(\lambda)\, b \left(\frac{\sin\frac{\lambda}{2b}}{\frac{\lambda}{2}}\right)^2 d\lambda \\
&= 2\pi b\mathcal{K}_1(0) + 4b^2 \int_{-\infty}^{\infty} \frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\left(\sin\frac{\lambda}{2b}\right)^2 d\lambda, \qquad \text{(A.54)}
\end{aligned}$$

where the last equality hold because

$$\int_{-\infty}^{\infty} \left(\frac{\lambda}{2b}\right)^{-2}\left(\sin\frac{\lambda}{2b}\right)^2 d\lambda = 2\int_{-\infty}^{\infty} x^{-2}\sin^2 x dx = 2\pi. \qquad \text{(A.55)}$$

Now,

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\left(\sin\frac{\lambda}{2b}\right)^2 d\lambda \\
&= \int_{-\infty}^{\infty} \frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\left(\left(\sin\frac{\lambda}{2b}\right)^2 - \frac{1}{2}\right) d\lambda + \frac{1}{2}\int_{-\infty}^{\infty} \frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2} d\lambda \\
&= -\frac{1}{2}\int_{-\infty}^{\infty} \left(\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\right)\left(\cos\frac{1}{b}\lambda\right) d\lambda + \frac{1}{2}\int_{-\infty}^{\infty} \frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2} d\lambda \\
&= \frac{1}{2}\int_{-\infty}^{\infty} \left(\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\right) d\lambda + o(1) \qquad \text{(A.56)}
\end{aligned}$$

as $b \to 0$, where we have used the Riemann-Lebesgue lemma. In view of the symmetry of $k(x)$, $\mathcal{K}_1(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} k(x)\cos(\lambda x)dx$, and, therefore, (A.54) and (A.56) lead to

$$\begin{aligned}
&\int_0^1 \int_0^1 k_b(r-s)drds \\
&= 2\pi b\mathcal{K}_1(0) + 2b^2 \int_{-\infty}^{\infty} \left(\frac{\mathcal{K}_1(\lambda)-\mathcal{K}_1(0)}{\lambda^2}\right) d\lambda \\
&= 2\pi b\mathcal{K}_1(0) + b^2\frac{1}{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k(x)\frac{\cos\lambda x - 1}{\lambda^2} dxd\lambda \\
&= 2\pi b\mathcal{K}_1(0) - b^2\int_{-\infty}^{\infty} k(x)\left|x\right| dx. \\
&= bc_1 + b^2 c_3 + o(b^2). \qquad \text{(A.57)}
\end{aligned}$$

Similarly,

$$\int_0^1 \int_0^1 k_b^2(r-s)drds = bc_2 + b^2c_4 + o(b^2). \tag{A.58}$$

Next,

$$\begin{aligned}
&\int_0^1 k_b(r-s)ds \\
&= \frac{1}{2}\int_{-\infty}^{\infty} \mathcal{K}(\lambda) \int_0^1 \left\{ \exp(\frac{i\lambda(r-s)}{b}) + \exp(-\frac{i\lambda(r-s)}{b}) \right\} dsd\lambda \\
&= \int_{-\infty}^{\infty} \mathcal{K}(\lambda) \int_0^1 \cos\left(\frac{\lambda(r-s)}{b}\right) dsd\lambda \qquad \text{(A.59)} \\
&= b\int_{-\infty}^{\infty} \mathcal{K}(\lambda) \frac{1}{\lambda} \left(\left\{ \sin\left(\frac{\lambda(r-1)}{b}\right) - \sin\left(\frac{\lambda r}{b}\right) \right\} \right) d\lambda \\
&= b\int_{-\infty}^{\infty} \mathcal{K}(xb) \frac{1}{x} \left(\left\{ \sin(x(r-1)) - \sin(xr) \right\} \right) dx,
\end{aligned}$$

so

$$\begin{aligned}
&\int_0^1 \int_0^1 \int_0^1 k_b(r-p)k_b(r-q)drdpdq \\
&= b^2 \int_0^1 \left[\int_{-\infty}^{\infty} \mathcal{K}(xb) \frac{1}{x} \left(\left\{ \sin(x(r-1)) - \sin(xr) \right\} \right) dx \right]^2 dr \\
&= b^2\mathcal{K}^2(0) \int_0^1 \left(\int_{-\infty}^{\infty} \frac{1}{x} \sin(x(r-1))\, dx - \int_{-\infty}^{\infty} \frac{1}{x} \sin(xr)\, dx \right)^2 dr \\
&= b^2\mathcal{K}^2(0) \int_0^1 \left(\int_{-\infty}^{\infty} \frac{\sin(x(r-1))}{x(r-1)} d(x(r-1)) - \int_{-\infty}^{\infty} \frac{1}{xr} \sin(xr)\, dxr \right)^2 dr \\
&= b^2\mathcal{K}^2(0) \int_0^1 \left(2\int_{-\infty}^{\infty} \frac{1}{y} \sin(y)\, dy \right)^2 dr = c_1^2 b^2. \qquad \text{(A.60)}
\end{aligned}$$

Combining (A.57), (A.58), and (A.60) yields

$$\mu_b = 1 - bc_1 - b^2c_3 + o(b^2), \tag{A.61}$$

and

$$\alpha_2 = 2bc_2 + b^2(c_4 - 2c_1^2) + o(b^2). \tag{A.62}$$

Now

$$
\begin{aligned}
F_\delta(z) &= G_\delta(\mu_b z^2) + \frac{1}{2}\left(G''_\delta(\mu_b z^2) z^4\right)\alpha_2 + o(b^2) \\
&= G_\delta(z^2) - G'_\delta(z^2) z^2 b c_1 - G'_\delta(z^2) z^2 b^2 c_3 + \frac{1}{2} G''_\delta(z^2) z^4 c_1^2 b^2 \\
&+ G''_\delta(z^2) z^4 b c_2 + \frac{1}{2} G''_\delta(z^2) z^4 b^2 (c_4 - 2c_1^2) \\
&+ \frac{1}{2} G'''_\delta(z^2) z^6 \left(-b c_1\right)\left(2 b c_2\right) + o(b^2) \\
&= G_\delta(z^2) + \left[c_2 G''_\delta(z^2) z^4 - c_1 G'_\delta(z^2) z^2\right] b \\
&- \left(G'_\delta(z^2) z^2 c_3 - \frac{1}{2} G''_\delta(z^2) z^4 (c_4 - c_1^2) + G'''_\delta(z^2) z^6 c_1 c_2\right) b^2 + o(b^2)
\end{aligned}
\tag{A.63}
$$

uniformly over $z \in \mathbb{R}^+$, where the uniformity holds because $\left|G'_\delta(z^2) z^2\right| \le C$, $\left|G''_\delta(z^2) z^4\right| < C$ and $\left|G''_\delta(z^2) z^6\right| < C$ for some constant C. This completes the proof of the theorem. ■

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

$$
\begin{aligned}
F_0(z_{\alpha,b}) &= D(z_{\alpha,b}^2) + \left[c_2 D''(z_{\alpha,b}^2) z_{\alpha,b}^4 - c_1 D'(z_{\alpha,b}^2) z_{\alpha,b}^2\right] b \\
&- \left(D'(z_{\alpha,b}^2) z_{\alpha,b}^2 c_3 - \frac{1}{2} D''(z_{\alpha,b}^2) z_{\alpha,b}^4 (c_4 - c_1^2) + D'''(z_{\alpha,b}^2) z_{\alpha,b}^6 c_1 c_2\right) b^2 + o\left(b^2\right) \\
&= D(z_\alpha^2) + D'(z_\alpha^2)\left(z_{\alpha,b}^2 - z_\alpha^2\right) + \frac{1}{2} D''(z_\alpha^2)\left(z_{\alpha,b}^2 - z_\alpha^2\right)^2 \\
&+ \left[c_2 D''(z_\alpha^2) z_\alpha^4 - c_1 D'(z_\alpha^2) z_\alpha^2\right] b \\
&+ \left[c_2 D'''(z_\alpha^2) z_\alpha^4 + 2 c_2 D''(z_\alpha^2) z_\alpha^2 - c_1 D''(z_\alpha^2) z_\alpha^2 - c_1 D'(z_\alpha^2)\right]\left(z_{\alpha,b}^2 - z_\alpha^2\right) b \\
&- \left(D'(z_\alpha^2) z_\alpha^2 c_3 - \frac{1}{2} D''(z_\alpha^2) z_\alpha^4 (c_4 - c_1^2) + D'''(z_\alpha^2) z_\alpha^6 c_1 c_2\right) b^2 + o(b^2),
\end{aligned}
\tag{A.64}
$$

i.e.

$$
\begin{aligned}
& D'(z_\alpha^2)\left(z_{\alpha,b}^2 - z_\alpha^2\right) + \frac{1}{2} D''(z_\alpha^2)\left(z_{\alpha,b}^2 - z_\alpha^2\right)^2 \\
&+ \left[c_2 D''(z_\alpha^2) z_\alpha^4 - c_1 D'(z_\alpha^2) z_\alpha^2\right] b \\
&+ \left[c_2 D'''(z_\alpha^2) z_\alpha^4 + 2 c_2 D''(z_\alpha^2) z_\alpha^2 - c_1 D''(z_\alpha^2) z_\alpha^2 - c_1 D'(z_\alpha^2)\right]\left[\left(z_{\alpha,b}^2 - z_\alpha^2\right)\right] b \\
&- \left(D'(z_\alpha^2) z_\alpha^2 c_3 - \frac{1}{2} D''(z_\alpha^2) z_\alpha^4 (c_4 - c_1^2) + D'''(z_\alpha^2) z_\alpha^6 c_1 c_2\right) b^2 + o(b^2) = 0.
\end{aligned}
\tag{A.65}
$$

Let

$$
z_{\alpha,b}^2 = z_\alpha^2 + k_1 b + k_2 b^2 + o(b^2),
\tag{A.66}
$$

then

$$
\begin{aligned}
&\left[c_2 D''(z_\alpha^2) z_\alpha^4 - c_1 D'(z_\alpha^2) z_\alpha^2\right] b + D'(z_\alpha^2) k_1 b \\
&- \left(D'(z_\alpha^2) z_\alpha^2 c_3 - \frac{1}{2} D''(z_\alpha^2) z_\alpha^4 (c_4 - c_1^2) + D'''(z_\alpha^2) z_\alpha^6 c_1 c_2 \right) b^2 \\
&+ \left[c_2 D'''(z_\alpha^2) z_\alpha^4 + 2 c_2 D''(z_\alpha^2) z_\alpha^2 - c_1 D''(z_\alpha^2) z_\alpha^2 - c_1 D'(z_\alpha^2)\right] k_1 b^2 \\
&+ \frac{1}{2} D''(z_\alpha^2) k_1^2 b^2 + D'(z_\alpha^2) k_1 b + o(b^2) = 0.
\end{aligned} \tag{A.67}
$$

This implies that

$$
k_1 = -\frac{1}{D'(z_\alpha^2)} \left[c_2 D''(z_\alpha^2) z_\alpha^4 - c_1 D'(z_\alpha^2) z_\alpha^2\right], \tag{A.68}
$$

and

$$
\begin{aligned}
k_2 = -\tfrac{1}{D'(z_\alpha^2)} \Big[&- \left(D'(z_\alpha^2) z_\alpha^2 c_3 - \tfrac{1}{2} D''(z_\alpha^2) z_\alpha^4 (c_4 - c_1^2) + D'''(z_\alpha^2) z_\alpha^6 c_1 c_2 \right) \\
&+ \left(c_2 D'''(z_\alpha^2) z_\alpha^4 + 2 c_2 D''(z_\alpha^2) z_\alpha^2 - c_1 D''(z_\alpha^2) z_\alpha^2 - c_1 D'(z_\alpha^2) \right) k_1 \\
&+ \tfrac{1}{2} D''(z_\alpha^2) k_1^2 \Big].
\end{aligned} \tag{A.69}
$$

Now

$$
D'(z) = \frac{z^{-1/2} e^{-z/2}}{\Gamma(1/2)\sqrt{2}}, \; D''(z) = \frac{1}{4\sqrt{\pi} z^2} \left(-\sqrt{2z} e^{-\frac{1}{2}z} - z^{\frac{3}{2}} \sqrt{2} e^{-\frac{1}{2}z} \right), \tag{A.70}
$$

$$
D'''(z) = \frac{1}{8\sqrt{\pi} z^{\frac{7}{2}}} \left(3z\sqrt{2} e^{-\frac{1}{2}z} + 2z^2 \sqrt{2} e^{-\frac{1}{2}z} + z^3 \sqrt{2} e^{-\frac{1}{2}z} \right) \tag{A.71}
$$

and thus

$$
\frac{D''(z^2)}{D'(z^2)} = \frac{1}{4z^3} \left(-2z - 2z^3 \right), \; \frac{D'''(z^2)}{D'(z^2)} = \frac{1}{4z^4} \left(2z^2 + z^4 + 3 \right). \tag{A.72}
$$

Hence

$$
k_1 = \left(c_1 + \frac{1}{2} c_2 \right) z_\alpha^2 + \frac{1}{2} c_2 z_\alpha^4, \tag{A.73}
$$

and

$$
\begin{aligned}
k_2 = &\left(\frac{1}{2} c_1^2 + \frac{3}{2} c_1 c_2 + \frac{3}{16} c_2^2 + c_3 + \frac{1}{4} c_4 \right) z_\alpha^2 \\
&+ \left(-\frac{1}{2} c_1 + \frac{3}{2} c_1 c_2 + \frac{9}{16} c_2^2 + \frac{1}{4} c_4 \right) z_\alpha^4 + \left(\frac{5}{16} c_2^2 \right) z_\alpha^6 - \left(\frac{1}{16} c_2^2 \right) z_\alpha^8
\end{aligned} \tag{A.74}
$$

as desired.

It follows from $z_{\alpha,b}^2 = z_\alpha^2 + k_1 b + k_2 b^2 + o(b^2)$ that

$$\begin{aligned} z_{\alpha,b} &= z_\alpha \left(1 + \frac{1}{2}\frac{k_1 b + k_2 b^2}{z_\alpha^2} - \frac{1}{8}\frac{k_1^2 b^2}{z_\alpha^4}\right) + o(b^2) \\ &= z_\alpha + \frac{1}{2}\frac{k_1}{z_\alpha} b + \left(\frac{1}{2}\frac{k_2}{z_\alpha} - \frac{1}{8}\frac{k_1^2}{z_\alpha^3}\right) b^2 + o(b^2) \\ &= z_\alpha + k_3 b + k_4 b^2 + o(b^2) \qquad \text{(A.75)} \end{aligned}$$

where

$$k_3 = \frac{1}{2}\left[\left(c_1 + \frac{1}{2}c_2\right) z_\alpha + \frac{1}{2} c_2 z_\alpha^3\right] \qquad \text{(A.76)}$$

and

$$\begin{aligned} k_4 = &\left(\frac{1}{2}c_3 + \frac{1}{8}c_4 + \frac{5}{8}c_1 c_2 + \frac{1}{8}c_1^2 + \frac{1}{16}c_2^2\right) z_\alpha \\ &+ \left(-\frac{1}{4}c_1 + \frac{1}{8}c_4 + \frac{5}{8}c_1 c_2 + \frac{7}{32}c_2^2\right) z_\alpha^3 + \frac{1}{8}c_2^2 z_\alpha^5 - \frac{1}{32}c_2^2 z_\alpha^7 \qquad \text{(A.77)} \end{aligned}$$

■

Proof of Corollary 3. For notational convenience, let

$$p^{(1)}(z_\alpha^2) = \left(c_1 + \frac{c_2}{2}\right) z_\alpha^2 + \left(\frac{c_2}{2}\right) z_\alpha^4, \qquad \text{(A.78)}$$

and then $z_{\alpha,b}^2 = z_\alpha^2 + p^{(1)}(z_\alpha^2) b + o(b)$. We have

$$\begin{aligned} &1 - EG_\delta(z_{\alpha,b}^2 \Xi_b) \\ &= 1 - G_\delta\left(z_\alpha^2 + p^{(1)}(z_\alpha^2)b\right) - \left\{c_2 G''_\delta\left(z_\alpha^2 + p^{(1)}(z_\alpha^2)b\right)\left[z_\alpha^2 + bp^{(1)}(z_\alpha^2)\right]^2\right\} b \\ &+ c_1\left\{G'_\delta\left(z_\alpha^2 + p^{(1)}(z_\alpha^2)b\right)\left(z_\alpha^2 + p^{(1)}(z_\alpha^2)b\right)\right\} b + o(b) \\ &= 1 - G_\delta(z_\alpha^2) - G'_\delta(z_\alpha^2)p^{(1)}(z_\alpha^2)b - \left[c_2 G''_\delta(z_\alpha^2)z_\alpha^4 - c_1 G'_\delta(z_\alpha^2)z_\alpha^2\right] b + o(b) \\ &= 1 - G_\delta(z_\alpha^2) - G'_\delta(z_\alpha^2)\left[\left(c_1 + \frac{c_2}{2}\right) z_\alpha^2 + \left(\frac{c_2}{2}\right) z_\alpha^4\right] b - \left[c_2 G''_\delta(z_\alpha^2)z_\alpha^4 - c_1 G'_\delta(z_\alpha^2)z_\alpha^2\right] b + o(b) \\ &= 1 - G_\delta(z_\alpha^2) - G'_\delta(z_\alpha^2)\left[\left(c_1 + \frac{c_2}{2}\right) z_\alpha^2 + \left(\frac{c_2}{2}\right) z_\alpha^4\right] b - \left[c_2 G''_\delta(z_\alpha^2)z_\alpha^4 - c_1 G'_\delta(z_\alpha^2)z_\alpha^2\right] b + o(b) \\ &= 1 - G_\delta(z_\alpha^2) - \left(\frac{c_2}{2}G'_\delta(z_\alpha^2)z_\alpha^4 + c_2 G''_\delta(z_\alpha^2)z_\alpha^4 + \frac{c_2}{2}G'_\delta(z_\alpha^2)z_\alpha^2\right) b + o(b) \\ &= 1 - G_\delta(z_\alpha^2) - c_2\left(\frac{1}{2}G'_\delta(z_\alpha^2)z_\alpha^4 + G''_\delta(z_\alpha^2)z_\alpha^4 + \frac{1}{2}G'_\delta(z_\alpha^2)z_\alpha^2\right) b + o(b). \qquad \text{(A.79)} \end{aligned}$$

Note that

$$G'_\delta(z) = \sum_{j=0}^{\infty} \frac{\left(\delta^2/2\right)^j}{j!} e^{-\delta^2/2} \frac{z^{j-\frac{1}{2}} e^{-z/2}}{\Gamma(j+1/2)2^{j+1/2}}, \qquad \text{(A.80)}$$

and

$$G''_{\delta}(z) = \sum_{j=0}^{\infty} \frac{\left(\delta^2/2\right)^j}{j!} e^{-\delta^2/2} \left((j - \frac{1}{2})\frac{1}{z} \frac{z^{j-\frac{1}{2}} e^{-z/2}}{\Gamma(j+1/2) 2^{j+1/2}} - \frac{1}{2} \frac{z^{j-\frac{1}{2}} e^{-z/2}}{\Gamma(j+1/2) 2^{j+1/2}} \right)$$
$$= \left(-\frac{1}{2z} - \frac{1}{2} \right) G'_{\delta}(z) + \sum_{j=0}^{\infty} \frac{\left(\delta^2/2\right)^j}{j!} e^{-\delta^2/2} \frac{z^{j-1/2} e^{-z/2}}{\Gamma(j+1/2) 2^{j+1/2}} \frac{j}{z}$$
$$= -\frac{1}{2} G'_{\delta}(z) \left(\frac{1}{z} + 1 \right) + K_{\delta}(z), \tag{A.81}$$

so that

$$\frac{1}{2} G'_{\delta}(z_\alpha^2) z_\alpha^4 + G''_{\delta}(z_\alpha^2) z_\alpha^4 + \frac{1}{2} G'_{\delta}(z_\alpha^2) z_\alpha^2 = z_\alpha^4 K_{\delta}\left(z_\alpha^2\right), \tag{A.82}$$

and

$$1 - EG_{\delta}(z_{\alpha,b}^2 \Xi_b) = 1 - G_{\delta}(z_\alpha^2) - c_2 z_\alpha^4 K_{\delta}\left(z_\alpha^2\right) b + o(b), \tag{A.83}$$

completing the proof of the corollary. ■

Proof of Theorem 4. It follows from Lemma 3 that when $b \to 0$,

$$\begin{aligned}
\alpha_{2,T} &= \kappa_{2,T} = 2bc_2(1 + o(1)) + O\left(T^{-1}\right), \\
\alpha_{3,T} &= \kappa_{3,T} = O(b^2) + O\left(T^{-1}\right), \\
\alpha_{4,T} &= \kappa_{4,T} + 3\kappa_{2,T}^2 = O(b^2) + O\left(T^{-1}\right),
\end{aligned} \tag{A.84}$$

and

$$\mu_{bT} = \mu_b - (bT)^{-q} g_q \left(\omega_T^{-2} \sum_{h=-\infty}^{\infty} |h|^q \gamma(h) \right) (1 + o(1)) + O\left(T^{-1}\right), \tag{A.85}$$

Thus, as $b \to 0$

$$\begin{aligned}
F_{T,\delta}(z) &= P\left\{ \left| \sqrt{T} \left(\hat{\beta} - \beta_0 \right) / \hat{\omega}_b \right| \le z \right\} = E\left\{ G_{\delta}(z^2 \varsigma_{bT}) \right\} + O\left(T^{-1}\right) \\
&= G_{\delta}(\mu_{bT} z^2) + \frac{1}{2} G''_{\delta}(\mu_{bT} z^2) z^4 \alpha_{2,T} + o(b) \\
&= G_{\delta}(\mu_{bT} z^2) + \frac{1}{2} G''_{\delta}(\mu_{bT} z^2) z^4 (2bc_2) + o(b) + O\left(T^{-1}\right) \\
&= G_{\delta}(\mu_b z^2) + G'_{\delta}(\mu_b z^2) z^2 (\mu_{bT} - \mu_b) + bc_2 G''_{\delta}(\mu_b z^2) z^4 \\
&\quad + o(b) + O\left(T^{-1}\right),
\end{aligned} \tag{A.86}$$

uniformly over $z \in \mathbb{R}^+$, using (A.84) and (A.85). But

$$\begin{aligned}
G_{\delta}(\mu_b z^2) &= G_{\delta}(z^2) + G'_{\delta}(z^2) z^2 (\mu_b - 1) + o(b) \\
&= G_{\delta}(z^2) - bc_1 G'_{\delta}(z^2) z^2 + o(b),
\end{aligned} \tag{A.87}$$

uniformly over $z \in \mathbb{R}^+$, and

$$
\begin{aligned}
& G'_\delta(\mu_b z^2) z^2 \left(\mu_{bT} - \mu_b\right) \\
&= \left(G'_\delta(z^2) + O\left(b\right)\right) z^2 \left\{ -g_q \left(\omega_T^{-2} \sum_{h=-\infty}^{\infty} |h|^q \gamma(h) \right) (bT)^{-q} (1 + o(1)) + O\left(T^{-1}\right) \right\} \\
&= -g_q \left(\omega_T^{-2} \sum_{h=-\infty}^{\infty} |h|^q \gamma(h) \right) G'_\delta(z^2) z^2 (bT)^{-q} (1 + o(1)) + o(b) + O\left(T^{-1}\right), \qquad \text{(A.88)}
\end{aligned}
$$

uniformly over $z \in \mathbb{R}^+$. So

$$
\begin{aligned}
F_{T,\delta}(z) = {} & G_\delta(z^2) + \left(c_2 G''_\delta(\mu_b z^2) z^4 - c_1 G'_\delta(z^2) z^2\right) b - g_q d_{qT} G'_\delta(z^2) z^2 (bT)^{-q} \\
& + o(b + (bT)^{-q}) + O\left(T^{-1}\right), \qquad \text{(A.89)}
\end{aligned}
$$

uniformly over $z \in \mathbb{R}^+$, as desired. ■

Proof of Corollary 5. Part (a) Using Theorem 4, we have, as $b + 1/T + 1/(bT) \to 0$

$$
\begin{aligned}
F_{T,0}(z_{\alpha,b}) &= D(z_{\alpha,b}^2) + \left[c_2 D''(z_{\alpha,b}^2) z_{\alpha,b}^4 - c_1 D'(z_{\alpha,b}^2) z_{\alpha,b}^2\right] b \\
&\quad - g_q d_{qT} D'(z_{\alpha,b}^2) z_{\alpha,b}^2 (bT)^{-q} + O\left(T^{-1}\right) + o\left(b + (bT)^{-q}\right) \\
&= F_0(z_{\alpha,b}) - g_q d_{qT} D'(z_{\alpha,b}^2) z_{\alpha,b}^2 (bT)^{-q} + O\left(T^{-1}\right) + o\left(b + (bT)^{-q}\right) \qquad \text{(A.90)} \\
&= 1 - \alpha - g_q d_{qT} D'(z_\alpha^2) z_\alpha^2 (bT)^{-q} + O\left(T^{-1}\right) + o\left(b + (bT)^{-q}\right).
\end{aligned}
$$

So

$$
1 - F_{T,0}(z_{\alpha,b}) - \alpha = g_q d_{qT} D'(z_\alpha^2) z_\alpha^2 (bT)^{-q} + O\left(T^{-1}\right) + o\left(b + (bT)^{-q}\right). \qquad \text{(A.91)}
$$

Part (b) Plugging $z_{\alpha,b}^2$ into (45) yields

$$
\begin{aligned}
& P\left(\left| \frac{\sqrt{T}(\hat{\beta} - \beta_0)}{\hat{\omega}_b} \right|^2 \geq z_{\alpha,b}^2 \right) \\
&= 1 - G_\delta(z_{\alpha,b}^2) - \left[c_2 G''_\delta(z_{\alpha,b}^2) z_{\alpha,b}^4 - c_1 G'_\delta(z_{\alpha,b}^2) z_{\alpha,b}^2\right] b \\
&+ g_q d_{qT} G'_\delta(z_{\alpha,b}^2) z_{\alpha,b}^2 (bT)^{-q} + O\left(T^{-1}\right) + o\left(b + (bT)^{-q}\right) \qquad \text{(A.92)} \\
&= 1 - G_\delta(z_\alpha^2) - c_2 z_\alpha^4 K_\delta\left(z_\alpha^2\right) b \\
&+ g_q d_{qT} G'_\delta(z_\alpha^2) z_\alpha^2 (bT)^{-q} + O\left(T^{-1}\right) + o\left(b + (bT)^{-q}\right),
\end{aligned}
$$

where the last equality follows as in the proof of Corollary 3. ■

Proof of Theorem 6. First, since $D(\cdot)$ is a bounded function, we can rewrite (18) as

$$
\begin{aligned}
P\left\{ \left| W(1) \Xi_b^{-1/2} \right| \leq z \right\} &= \lim_{B \to \infty} E D(z^2 \Xi_b) 1\left\{ |\Xi_b - \mu_b| \leq B \right\} \\
&= \lim_{B \to \infty} E \sum_{m=1}^{\infty} \frac{1}{m!} D^{(m)}(\mu_b z^2) \left(\Xi_b - \mu_b\right)^m z^{2m} 1\left\{ |\Xi_b - \mu_b| \leq B \right\} \\
&= \lim_{B \to \infty} \sum_{m=1}^{\infty} \frac{1}{m!} D^{(m)}(\mu_b z^2) \alpha_m z^{2m} 1\left\{ |\Xi_b - \mu_b| \leq B \right\}, \qquad \text{(A.93)}
\end{aligned}
$$

where the last line follows because the infinite sum $\sum_{m=1}^{\infty}\frac{1}{m!}D^{(m)}(\mu_b z^2)\alpha_m z^{2m}$ converges uniformly to $D(z^2\Xi_b)$ when $|\Xi_b-\mu_b|\leq B$. Uniformity holds because $D(\cdot)$ is infinitely differentiable with bounded derivatives.

Since $D(z^2)$ decays exponentially as $z^2\to\infty$, there exists a constant C such that $\left|D^{(m)}(\mu_b z^2)z^{2m}\right|<C$ for all m. Using this and Lemma 1, we have

$$\begin{aligned}&\left|\sum_{m=1}^{\infty}\frac{1}{m!}D^{(m)}(\mu_b z^2)\alpha_m z^{2m}\right|\\&\leq C\sum_{m=1}^{\infty}\frac{1}{m!}\left|\alpha_m\right|\leq C\sum_{m=1}^{\infty}\frac{1}{m!}2^{2m}m!\,(c_1^*b)^{m-1}\\&=C\,(c_1^*b)^{-1}\sum_{m=1}^{\infty}(4c_1^*b)^m<\infty,\end{aligned}\tag{A.94}$$

provided that $b<1/(4c_1^*)$. As a consequence, the operation $\lim_{B\to\infty}$ can be moved inside the summation sign in (A.93), giving

$$P\left\{\left|W(1)\Xi_b^{-1/2}\right|\leq z\right\}=\sum_{m=1}^{\infty}\frac{1}{m!}D^{(m)}(\mu_b z^2)\alpha_m z^{2m},\tag{A.95}$$

when $b<1/(4c_1^*)$.

Second, it follows from (43) that

$$P\left\{\left|\sqrt{T}\left(\hat{\beta}-\beta_0\right)/\hat{\omega}_b\right|\leq z\right\}=E\left\{D(z^2\varsigma_{bT})\right\}+O\left(T^{-1}\right).\tag{A.96}$$

But

$$E\left\{D(z^2\varsigma_{bT})\right\}=\sum_{m=1}^{\infty}\frac{1}{m!}D^{(m)}(\mu_{bT}z^2)\alpha_{m,T}z^{2m},\tag{A.97}$$

where the right hand side converges to $E\left\{D(z^2\varsigma_{bT})\right\}$ uniformly over T because

$$\alpha_{m,T}=\alpha_m+O\left\{\frac{2^{2m}m!}{T^2}\,(c_1^*b)^{m-2}\right\},$$

uniformly over m by Lemma 2, $\left|D^{(m)}(\mu_{bT}z^2)z^{2m}\right|<C$ for some constant C, and thus

$$\left|\sum_{m=1}^{\infty}\frac{1}{m!}D^{(m)}(\mu_{bT}z^2)\alpha_{m,T}z^{2m}\right|\leq C\sum_{m=1}^{\infty}\frac{1}{m!}\left|\alpha_m\right|+\frac{C}{T^2}\sum_{m=1}^{\infty}2^{2m}\,(c_1^*b)^{m-2}<\infty,$$

when $b<1/(4c_1^*)$. Therefore

$$P\left\{\left|\sqrt{T}\left(\hat{\beta}-\beta_0\right)/\hat{\omega}_b\right|\leq z\right\}=\sum_{m=1}^{\infty}\frac{1}{m!}D^{(m)}(\mu_{bT}z^2)\alpha_{m,T}z^{2m}+O(\frac{1}{T}),\tag{A.98}$$

uniformly over $z \in \mathbb{R}^+$.

It follows from (A.95) and (A.98) that

$$\begin{aligned}
&|F_{T,0}(z) - F_0(z)| \\
&= \left| P\left\{ \left| \sqrt{T}\left(\hat{\beta} - \beta\right)/\hat{\omega}_b \right| \leq z \right\} - P\left\{ \left| W(1)\Xi_b^{-1/2} \right| < z \right\} \right| \\
&= \left| \sum_{m=1}^{\infty} \frac{1}{m!} D^{(m)}(\mu_{bT} z^2)\alpha_{m,T} z^{2m} - \sum_{m=1}^{\infty} \frac{1}{m!} D^{(m)}(\mu_b z^2)\alpha_m z^{2m} \right| + O(\frac{1}{T}) \qquad \text{(A.99)} \\
&= \left| \sum_{m=1}^{\infty} \frac{1}{m!} D^{(m)}(\mu_b z^2)\alpha_{m,T} z^{2m} - \sum_{m=1}^{\infty} \frac{1}{m!} D^{(m)}(\mu_b z^2)\alpha_m z^{2m} \right| + O(\frac{1}{T})
\end{aligned}$$

uniformly over $z \in \mathbb{R}$, where the second equality holds because

$$D^{(m)}(\mu_{bT} z^2) = D^{(m)}(\mu_b z^2) + O\left(D^{(m+1)}(\mu_b z^2) z^2/T\right),$$

uniformly over $z \in \mathbb{R}$ and

$$\left| \sum_{m=1}^{\infty} \frac{1}{m!} D^{(m+1)}(\mu_b z^2)\alpha_{m,T} z^{2m+2} \right| < \infty.$$

Therefore,

$$\begin{aligned}
|F_{T,0}(z) - F_0(z)| &= \left| \sum_{m=1}^{\infty} \frac{1}{m!} D^{(m)}(\mu_b z^2)\left(\alpha_{m,T} - \alpha_m\right) z^{2m} \right| + O(\frac{1}{T}) \\
&= O\left\{ \frac{1}{T^2} \sum_{m=1}^{\infty} 2^{2m}\left(c_1^* b)\right)^{m-2} \right\} + O\left(\frac{1}{T}\right) \\
&= O\left(\frac{1}{T}\right), \qquad \text{(A.100)}
\end{aligned}$$

uniformly over $z \in \mathbb{R}^+$ as desired. ■

References

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

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

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

[4] Gourieroux, C. and A. Monfort (1995): *Statistics and Econometric Models*, Vol 2, Cambridge University Press.

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

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

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

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

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

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

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

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

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

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

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

[16] Parzen, E. (1957): "On the Consistent Estimates of the Spectrum of a Stationary Time Series," *Annals of Mathematical Statistics*, 28, 329–348.

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

[18] Phillips, P. C. B., Y. Sun and S. Jin, (2005a): "long-run Variance Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation," forthcoming, *Journal of Statistical Planning and Inference.*

[19] ———, (2005b): "Spectral Density Estimation and Robust Hypothesis Testing using Steep Origin Kernels without Truncation," forthcoming, *International Economic Review.*

[20] ———, (2005c): "Improved HAR Inference Using Power Kernels without Truncation," Yale University, mimeographed.

[21] ———, (2005d), "Balancing Size and Power in Non Parametric Studentized Testing with Quadratic Power Kernels without Truncation," Department of Economics, UCSD.

[22] Politis, D. N. and J. P. Romano (1995): "Bias Corrected Nonparametric Spectral Density Estimator," *Journal of Time Series Analysis,* 16, 67–103.

[23] Politis, D. N. and J. P. Romano (1998): "Multivariate Density Estimation with General Flat-top Kernels of Infinite Order," *Journal of Multivariate Analysis*, 68, 1–25.

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

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

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