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

Sainan Jin
Guanghua School of Management
Peking University

January 5, 2011

Abstract

The paper considers studentized tests in time series regressions with nonparametrically autocorrelated errors. The studentization is based on robust standard errors with truncation lag $M = bT$ for some constant $b \in (0, 1]$ and sample size T. It is shown 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. 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 (not reported here) confirm that the new plug-in procedure works well in finite samples.

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 autocovariances. 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); 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)) 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.

In contrast to the above convention, the present paper develops a new approach to choosing the smoothing parameter. We focus on two-sided tests and 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 $O(1)$. The result is related to that of Jansson (2004) who shows that the ERP of the nonstandard test based on the Bartlett kernel with $b = 1$ is of order $O(\log T/T)$. Our result strengthens and generalizes Jansson's result in two aspects. First, we show that the $\log(T)$ factor can be dropped. Second, while Jansson's result applies only to the Bartlett kernel with $b = 1$, our result applies to more general kernels with both $b = 1$ and $b < 1$. 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 analytical findings here support earlier simulation results in Kiefer, Vogelsang and Bunzel (2000, hereafter KVB), Kiefer and Vogelsang (2002a, 2002b, hereafter KV), Gonçalves and Vogelsang (2005), and ourselves (Phillips, Sun and Jin (2006, 2007), hereafter PSJ).

The theoretical development is based on two high order asymptotic expansions. The first is the expansion of the nonstandard limiting distribution around its limiting chi-squared form. The second is the 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. These higher order expansions enable us to develop improved approximations to the type I and type II errors. For typical economic time series, the dominant term in the type I error increases as b decreases while the dominant term in the type II error decreases as b decreases. 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.

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 $O\left(1\right)$ 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.

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

1. Specify the null hypothesis $H_0 : \beta = \beta_0$ and an alternative hypothesis $H_1 : |\beta - \beta_0| = c_0 > 0$, where c_0 may reflect a value of scientific interest or economic significance if such a value is available. (In the absence of such a value, we recommend that the user may set the default discrepancy parameter value $\delta = 2$ in (1) below).

2. Specify the significance level α of the test and the associated two-sided standard normal critical value z_α satisfying $\Phi(z_\alpha) = 1 - \alpha/2$.

3. Specify a weighted average loss function (see (45) below) capturing the relative importance of the type I and II errors, and reflecting the relative cost of false acceptance and false rejection. Normalize the weight for the type II error to be to unity and let w be the weight for the type I error.

4. Estimate the model by OLS and construct the residuals $\hat{u}_t$.

5. Fit an AR(1) model to the estimated residuals and compute

$$\hat{d} = \frac{2\hat{\rho}}{(1 - \hat{\rho})^2}, \ \hat{\sigma}^2 = \frac{1}{T^2} \sum_{t=1}^{T} (\hat{u}_t - \hat{\rho}\hat{u}_{t-1})^2 ,$$
$$\delta = \sqrt{T}c_0 \left(1 - \hat{\rho}\right)/\hat{\sigma}, \tag{1}$$

where $\hat{\rho}$ is the OLS estimator of the AR coefficient. Set $\delta = 2$ as a default value if the user is unsure about the alternative hypothesis.

3

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

7. Compute the automatic bandwidth:

$$\hat{b} = \begin{cases} \left(\frac{qg\hat{d}[wG_0'(z_\alpha^2) - G_\delta'(z_\alpha^2)]}{cz_\alpha^2 K_\delta(z_\alpha^2)} \right)^{1/3} T^{-2/3}, & \hat{d}\left[wG_0'(z_\alpha^2) - G_\delta'(z_\alpha^2)\right] > 0 \\ (\log T)/T, & \hat{d}\left[wG_0'(z_\alpha^2) - G_\delta'(z_\alpha^2)\right] \le 0 \end{cases} \qquad (2)$$

where

$$g = \begin{cases} 6.000, & \text{Parzen kernel} \\ 1.421, & \text{QS kernel} \end{cases} \qquad c = \begin{cases} 0.539, & \text{Parzen kernel} \\ 1.000, & \text{QS kernel} \end{cases} \qquad (3)$$

$G_\delta'(\cdot)$ is the cdf of a (non)central chi-squared variate with one degree of freedom and noncentrality parameter δ^2, and $K_\delta(\cdot)$ is defined in equation (26).

8. Compute the HAC standard error using bandwidth $\hat{M} = \hat{b}T$ and construct the usual t-statistic $t_{\hat{b}}$.

9. Let $z_{\alpha,\hat{b}} = z_\alpha + k_3\hat{b} + k_4\hat{b}^2$ where k_3 and k_4 are given in Table I. Reject the null hypothesis if $|t_{\hat{b}}| \ge z_{\alpha,\hat{b}}$.

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. 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. 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 last section provides some concluding discussion. Proofs and relevant technical results are available in Sun, Phillips and Jin (2006, hereafter SPJ).

2 Heteroskedasticity-Autocorrelation Robust Inference

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

$$y_t = \beta + u_t, \quad t = 1, 2, ..., T, \qquad (4)$$

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 7.

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

$$\sqrt{T}(\hat{\beta} - \beta) = \frac{1}{\sqrt{T}} S_T, \tag{5}$$

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*

$$T^{-1/2} S_{[Tr]} \Rightarrow \omega W(r), \quad r \in [0, 1] \tag{6}$$

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

Under Assumption 1,

$$T^{-1/2} \hat{S}_{[Tr]} \Rightarrow \omega V(r), \quad r \in [0, 1], \tag{7}$$

where V is a standard Brownian bridge process, and

$$\sqrt{T}(\hat{\beta} - \beta) \Rightarrow \omega W(1) = N(0, \omega^2), \tag{8}$$

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

$$\omega^2 = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma(j), \tag{9}$$

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

$$\hat{\omega}^2(M) = \sum_{j=-T+1}^{T-1} k(\frac{j}{M}) \hat{\gamma}(j), \ \hat{\gamma}(j) = \begin{cases} \frac{1}{T} \sum_{t=1}^{T-j} \hat{u}_{t+j} \hat{u}_t & \text{for } j \geq 0 \\ \frac{1}{T} \sum_{t=-j+1}^{T} \hat{u}_{t+j} \hat{u}_t & \text{for } j < 0 \end{cases} \tag{10}$$

involving the sample covariances $\hat{\gamma}(j)$. In (10), $k(\cdot)$ 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), Newey and West (1987, 1994)).

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

$$t_{\hat{\omega}(M)} = T^{1/2}(\hat{\beta} - \beta_0)/\hat{\omega}(M), \tag{11}$$

5

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

$$\hat{\omega}_b^2 = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{bT}\right)\hat{\gamma}(j), \tag{12}$$

and the associated t statistic is given by

$$t_b = T^{1/2}(\hat{\beta} - \beta_0)/\hat{\omega}_b. \tag{13}$$

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

$$\Xi_b = \int_0^1 \int_0^1 k_b(r - s)dV(r)dV(s), \tag{14}$$

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

$$t_b \Rightarrow W(1)\Xi_b^{-1/2}, \tag{15}$$

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

$$t_b \Rightarrow (\delta + W(1))\Xi_b^{-1/2}, \tag{16}$$

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. It is important to point out that in the nonstandard limit distribution $W(1)$ is independent Ξ_b. This is because $W(1)$ is uncorrelated with $V(r)$ for any $r \in [0,1]$ and both $W(1)$ and $V(r)$ are Gaussian.

In related work, the present authors (PSJ (2006, 2007)) propose using an estimator of ω^2 of the form

$$\hat{\omega}_\rho^2 = \sum_{j=-T+1}^{T-1} \left[k\left(\frac{j}{T}\right)\right]^\rho \hat{\gamma}(j), \tag{17}$$

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 (2006a, 2006b)).

3 Expansion of the Nonstandard Limit Theory

This section develops asymptotic expansions of the limit distributions given in (15) and (16) 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. The expansions 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.

The 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 $\int_0^\infty k(x)x dx < \infty$.*
(ii) The Parzen characteristic exponent defined by

$$q = \max\{q_0 : q_0 \in \mathbb{Z}^+, \; g_{q_0} = \lim_{x \to 0} \frac{1 - k(x)}{|x|^{q_0}} < \infty\} \tag{18}$$

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 $\int_0^\infty k(x)x dx < \infty$ ensures the integrals that appear frequently in later developments are finite and validates use of the Riemann-Lebesgue lemma in 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.

Let $F_\delta(z) := P\left\{\left|(\delta + W(1))\Xi_b^{-1/2}\right| \leq z\right\}$ be the nonstandard limiting distribution and $G_\lambda = G(\cdot; \lambda^2)$ be the cdf of a non-central $\chi_1^2(\lambda^2)$ variate with noncentrality parameter λ^2. The following theorem establishes the asymptotic expansion of $F_\delta(z)$ around $G_\delta(z^2)$.

Theorem 1 *Let Assumption 2 hold. Then*

$$F_\delta(z) = G_\delta(z^2) + p_\delta(z^2)b + q_\delta(z)b^2 + o\left(b^2\right), \tag{19}$$

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

$$p_\delta(z^2) = c_2 G_\delta''(z^2)z^4 - c_1 G_\delta'(z^2)z^2,$$
$$q_\delta(z^2) = -G_\delta'(z^2)z^2 c_3 + \frac{1}{2}G_\delta''(z^2)z^4(2c_4 - c_1^2) - G_\delta'''(z^2)z^6 c_1 c_2, \tag{20}$$

and

$$c_1 = \int_{-\infty}^{\infty} k(x)dx, \qquad c_2 = \int_{-\infty}^{\infty} k^2(x)dx,$$
$$c_3 = -\int_{-\infty}^{\infty} k(x)\,|x|\,dx, \quad c_4 = -\int_{-\infty}^{\infty} k^2(x)\,|x|\,dx. \tag{21}$$

As is apparent from the proof given in SPJ, 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 $\delta = 0$,

$$F_0(z) = D(z^2) + \left[c_2 D''(z^2)z^4 - c_1 D'(z^2)z^2 \right] b$$
$$+ \left[-D'(z^2)z^2 c_3 + \frac{1}{2}D''(z^2)z^4(2c_4 - c_1^2) - D'''(z^2)z^6 c_1 c_2 \right] b^2 + o\left(b^2\right), \tag{22}$$

where $D(\cdot) = G_0(\cdot)$ is the cdf of χ_1^2 distribution. For any $\alpha \in (0, 1)$, let $z_\alpha^2 \in \mathbb{R}^+$, $z_{\alpha,b}^2 \in \mathbb{R}^+$ 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 to $O(b)$ the second-order corrected critical value, and the critical value correct to $O(b^2)$ the third-order corrected critical value. We use this convention throughout the rest of the paper. The following quantities appear in the corollary:

$$k_1 = \left(c_1 + \frac{1}{2}c_2 \right) z_\alpha^2 + \frac{1}{2}c_2 z_\alpha^4, \quad k_3 = \frac{1}{2}\left(c_1 + \frac{1}{2}c_2 \right) z_\alpha + \frac{1}{4}c_2 z_\alpha^3,$$

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

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

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

$$z_{\alpha,b}^2 = z_\alpha^2 + k_1 b + o(b), \quad z_{\alpha,b} = z_\alpha + k_3 b + o(b), \tag{23}$$

(ii) the third-order corrected critical values are

$$z_{\alpha,b}^2 = z_\alpha^2 + k_1 b + k_2 b^2 + o(b^2), \quad z_{\alpha,b} = z_\alpha + k_3 b + k_4 b^2 + o(b^2), \tag{24}$$

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

8

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.

For the Bartlett, Parzen and QS kernels, we can compute c_1, c_2, c_3 and c_4 either analytically or numerically. Table I gives the high-order corrected critical values for these three kernels. These higher order corrected critical values provide excellent approximations to the exact ones when b is smaller than 0.5 and reasonably good approximations for other values of b.

Table I. 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	16.1447	2.5616	2.4446	6.0489	9.3014	1.8386	1.7997
Parzen	7.8964	10.5040	2.0144	1.6444	4.7337	6.0229	1.4388	1.2014
QS	14.1017	34.9061	3.5974	5.6033	8.3968	19.9344	2.5522	4.0792

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*

$$P\left\{\left|(\delta + W(1)) \Xi_b^{-1/2}\right| > z_{\alpha,b}\right\} = 1 - G_\delta(z_\alpha^2) - c_2 z_\alpha^4 K_\delta\left(z_\alpha^2\right) b + o(b), \qquad (25)$$

as $b \to 0$ where

$$K_\delta(z) = \sum_{j=0}^{\infty} \frac{\left(\delta^2/2\right)^j}{j!} e^{-\delta^2/2} \frac{z^{j-1/2} e^{-z/2}}{\Gamma(j+1/2) 2^{j+1/2}} \frac{j}{z} \qquad (26)$$

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. For a given critical value, we can show that $z_\alpha^4 K_\delta\left(z_\alpha^2\right)$ 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), Velasco and Robinson (2001)). The following facilitates the development of the higher order expansion.

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

We develop an asymptotic expansion of $P\{|\sqrt{T}(\hat{\beta} - \beta_0)/\hat{\omega}_b| \leq z\}$ 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 $\tilde{\beta} = (l_T'\Omega_T^{-1}l_T)^{-1} l_T'\Omega_T^{-1}y$ and

$$\hat{\beta} - \beta = \tilde{\beta} - \beta + (l_T'l_T)^{-1} l_T'\tilde{u}, \tag{27}$$

where $\tilde{u} = (I - l_T (l_T'\Omega_T^{-1}l_T)^{-1} l_T'\Omega_T^{-1})u$, 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

$$\omega_T^2 := var\left(\sqrt{T}(\hat{\beta} - \beta)\right) = T^{-1}l_T'\Omega_Tl_T = \omega^2 + O\left(T^{-1}\right), \tag{28}$$

and it follows from Grenander and Rosenblatt (1957) that

$$\tilde{\omega}_T^2 := var\left(\sqrt{T}(\tilde{\beta} - \beta)\right) = T\left(l_T'\Omega_T^{-1}l_T\right)^{-1} = \omega^2 + O\left(T^{-1}\right). \tag{29}$$

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

$$P\left\{\sqrt{T}\left(\hat{\beta} - \beta_0\right)/\hat{\omega}_b \leq z\right\}$$
$$= P\left\{\sqrt{T}\left(\tilde{\beta} - \beta\right)/\tilde{\omega}_T + c/\tilde{\omega}_T \leq z\hat{\omega}_b/\tilde{\omega}_T - T^{-1/2}l_T'\tilde{u}/\tilde{\omega}_T\right\}$$
$$= E\Phi\left(z\hat{\omega}_b/\tilde{\omega}_T - c/\tilde{\omega}_T - T^{-1/2}l_T'\tilde{u}/\tilde{\omega}_T\right)$$
$$= E\Phi\left(z\hat{\omega}_b/\tilde{\omega}_T - c/\tilde{\omega}_T\right) - T^{-1/2}E\left[\varphi\left(z\hat{\omega}_b/\tilde{\omega}_T - c/\tilde{\omega}_T\right)l_T'\tilde{u}/\tilde{\omega}_T\right] + O\left(T^{-1}\right)$$
$$= P\left\{\sqrt{T}\left(\tilde{\beta} - \beta\right)/\tilde{\omega}_T + c/\tilde{\omega}_T \leq z\hat{\omega}_b/\tilde{\omega}_T\right\} + O\left(T^{-1}\right), \tag{30}$$

10

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 $E\left[\varphi\left(z\hat{\omega}_b/\tilde{\omega}_T - c/\tilde{\omega}_T\right)l_T'\tilde{u}\right] = 0$. 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

$$\begin{aligned}
F_{T,\delta}(z) &:= P\left\{\left|\sqrt{T}\left(\hat{\beta} - \beta_0\right)/\hat{\omega}_b\right| \leq z\right\} \\
&= E\left\{G_\delta(z^2\hat{\omega}_b^2/\tilde{\omega}_T^2)\right\} = E\left\{G_\delta(z^2\varsigma_{bT})\right\} + O\left(T^{-1}\right),
\end{aligned} \tag{31}$$

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

Setting $\mu_{bT} = E(\varsigma_{bT})$, we have

$$\begin{aligned}
F_{T,\delta}(z) &= G_\delta(\mu_{bT}z^2) + \frac{1}{2}G_\delta''(\mu_{bT}z^2)E(\varsigma_{bT} - \mu_{bT})^2 z^4 \\
&\quad + \frac{1}{6}G_\delta'''(\mu_{bT}z^2)E(\varsigma_{bT} - \mu_{bT})^3 z^6 + O\left(E(\varsigma_{bT} - \mu_{bT})^4\right) + O\left(T^{-1}\right)
\end{aligned} \tag{32}$$

where the $O(\cdot)$ term holds uniformly over $z \in \mathbb{R}^+$. By developing asymptotic expansions of μ_{bT} and $E(\varsigma_{bT} - \mu_{bT})^m$ 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*

$$\begin{aligned}
F_{T,\delta}(z) &= G_\delta(z^2) + \left[c_2 G_\delta''(\mu_b z^2)z^4 - c_1 G_\delta'(z^2)z^2\right]b \\
&\quad - g_q d_{qT} G_\delta'(z^2)z^2(bT)^{-q} + o\left\{b + (bT)^{-q}\right\} + O\left(T^{-1}\right)
\end{aligned} \tag{33}$$

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

$$\begin{aligned}
F_{T,0}(z) &= D(z^2) + \left[c_2 D''(z^2)z^4 - c_1 D'(z^2)z^2\right]b \\
&\quad - g_q d_{qT} D'(z^2)z^2(bT)^{-q} + o\left\{b + (bT)^{-q}\right\} + O\left(T^{-1}\right).
\end{aligned} \tag{34}$$

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 (22) above. Thus, use of the second-order corrected critical values given in (23), which take account of terms up to order $O(b)$, should lead to size improvements when $T^q b^{q+1} \to \infty$.

The third term in the expansion (34) 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 (34), or equivalently those based on the nonstandard

limit theory, at least to order $O(b)$. Thus, the third term of (34) approximately measures how satisfactory the second-order corrected critical values given in (23) are for any given values of b and T.

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 $o\left((bT)^{-q} + b\right) + O\left(T^{-1}\right)$. Note that the above approximation depends on δ only through δ^2, so it is valid under both $H_1 : \beta - \beta_0 = c/\sqrt{T}$ and $H_1 : \beta - \beta_0 = -c/\sqrt{T}$.

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

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 two-sided t-test based on the second-order corrected critical values is

$$(1 - F_{T,0}(z_{\alpha,b})) - \alpha = g_q d_{qT} D'(z_\alpha^2)z_\alpha^2(bT)^{-q} + o\left((bT)^{-q} + b\right) + O\left(T^{-1}\right); \qquad (35)$$

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

$$\begin{aligned} 1 - F_{T,\delta}(z_{\alpha,b}) = 1 &- G_\delta(z_\alpha^2) - c_2 z_\alpha^4 K_\delta\left(z_\alpha^2\right)b \\ &+ g_q d_{qT} G_\delta'(z_\alpha^2)z_\alpha^2(bT)^{-q} + o\left((bT)^{-q} + b\right) + O\left(T^{-1}\right). \end{aligned} \qquad (36)$$

5 Accuracy of the Nonstandard Approximation

The previous section showed 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 (31), the error of the nonstandard approximation is given by

$$\begin{aligned} F_{T,0}(z) - F_0(z) &:= P\left\{\left|\sqrt{T}\left(\hat{\beta} - \beta\right)/\hat{\omega}_b\right| \le z\right\} - P\left\{\left|W(1)\Xi_b^{-1/2}\right| \le z\right\} \\ &= ED(z^2\varsigma_{bT}) - ED(z^2\Xi_b) + O\left(T^{-1}\right). \end{aligned} \qquad (37)$$

By comparing the cumulants of $\varsigma_{bT} - E\varsigma_{bT}$ with that of $\Xi_b - E\Xi_b$, we can show that $ED(z^2\varsigma_{bT}) - ED(z^2\Xi_b) = O\left(T^{-1}\right)$, which, combined with the above equation, gives Theorem (6) below. The requirement $b < 1/(16\int_{-\infty}^\infty |k(x)|dx)$ on b that appears in the theorem 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*

$$F_{T,0}(z) = F_0(z) + O\left(T^{-1}\right), \tag{38}$$

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 $bT \to \infty$,

$$F_{T,0}(z) = D(z^2) + O\left(T^{-q/(q+1)}\right) \tag{39}$$

for typical economic time series. Comparing (38) with (39), 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 (38) holds for fixed b while the $O(\cdot)$ term in (39) holds for diminishing b. Since the $O(\cdot)$ term in (38) 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

$$F_{T,0}(z) - D(z^2) = F_{T,0}(z) - F_0(z) + F_0(z) - D(z^2) = O(1). \tag{40}$$

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 (40) with (38), 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

$$\begin{aligned}
F_{T,0}(z) - F_0(z) &= \left[F_{T,0}(z) - D(z^2)\right] - \left[F_0(z) - D(z^2)\right] \\
&= -g_q d_{qT} D'(z^2)z^2(bT)^{-q} + o\left\{b + (bT)^{-q}\right\} + O\left(T^{-1}\right), \tag{41}
\end{aligned}$$

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 (41).

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 (35), we know that the type I error for a two-sided test with nominal size α can be expressed as

$$1 - F_{T,0}(z_{\alpha,b}) = \alpha + g_q d_{qT} D'(z_\alpha^2) z_\alpha^2 (bT)^{-q} + o\left((bT)^{-q} + b\right) + O\left(T^{-1}\right), \qquad (42)$$

Similarly, from (36), the type II error has the form

$$G_\delta(z_\alpha^2) + c_2 z_\alpha^4 K_\delta(z_\alpha^2) b - g_q d_{qT} G'_\delta(z_\alpha^2) z_\alpha^2 (bT)^{-q} + o\left((bT)^{-q} + b\right) + O\left(T^{-1}\right). \qquad (43)$$

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 (42); (ii) The magnitude of the type II error, as measured by the $O(b)$ and $O((bT)^{-q})$ terms in (43); 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 (43) 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. 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 (43) 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.

14

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, (42) shows that the type I error increases with b while the type II error may increase or decrease with b depending on which of the two terms in (43) dominates.

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 (42) and (43). Setting

$$e_T^I = \alpha + g_q d_{qT} D'(z^2) z^2 (bT)^{-q},$$
$$e_T^{II} = G_\delta(z_\alpha^2) + c_2 z_\alpha^4 K_\delta(z_\alpha^2) b - g_q d_{qT} G'_\delta(z_\alpha^2) z_\alpha^2 (bT)^{-q}, \tag{44}$$

we define the loss function to be

$$L\left(b; \delta, T, z_\alpha\right) = \frac{w_T(\delta)}{1 + w_T(\delta)} e_T^I + \frac{1}{1 + w_T(\delta)} e_T^{II}, \tag{45}$$

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 $L\left(b; \delta, T, z_\alpha\right)$ 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:

$$L\left(b; \delta, T, z_\alpha\right) = \left[g_q d_{qT} \left\{w_T D'(z_\alpha^2) - G'_\delta(z_\alpha^2)\right\} z_\alpha^2 (bT)^{-q} + c_2 z_\alpha^4 K_\delta(z_\alpha^2) b\right] \frac{1}{1 + w_T} + C_T, \tag{46}$$

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 $L\left(b; \delta, T, z_\alpha\right)$ is minimized for the following choice of b:

$$b_{\text{opt}} = \left\{\frac{q g_q d_{qT} \left[w_T D'(z_\alpha^2) - G'_\delta(z_\alpha^2)\right]}{c_2 z_\alpha^2 K_\delta(z_\alpha^2)}\right\}^{1/(q+1)} T^{-q/(q+1)}. \tag{47}$$

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

$$b_{\text{opt}} = \left\{\frac{q g_q d_{qT} D'(z_\alpha^2)}{z_\alpha^2 c_2 K_\delta(z_\alpha^2)}\right\}^{1/(q+1)} (w_T/T^q)^{1/(q+1)}. \tag{48}$$

Fixed b rules may then be interpreted as assigning relative weight $w_T = O\left(T^q\right)$ 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 (47), 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 $O\left(w_T T\right)^{-q/(q+1)}$ rather than $O\left(T^{-1}\right)$, 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 (46). 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 $O\left(T^{-2q/(2q+1)}\right)$.

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 $O\left(T^{-2/3}\right)$ 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 $L\left(b; \delta, T, z_\alpha\right)$ approaches C_T from above. Therefore, the loss $L\left(b; \delta, T, z_\alpha\right)$ 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 $d_{qT} < 0$, $L\left(b; \delta, T, z_\alpha\right)$ 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 practice, we may 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 $b = O\left(T^{-q/(q+1)}\right)$. This rate may be compared with the optimal rate of $b = O\left(T^{-2q/(2q+1)}\right)$ 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. This conclusion highlights the importance of

bias reduction in statistical testing.

In Monte Carlo experiments not reported here, we have compared the finite sample performance of the new plug-in procedure with that of the conventional plug-in procedure given in Andrews (1991). The findings from these simulations indicate that the new plug-in procedure works well in terms of incurring a smaller loss than the conventional plug-in procedure. Detailed results of this experiment are reported in Sun, Phillips and Jin (2006).

7 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 results in this paper suggest some important areas of future research. The present work has focused on two-sided tests for the Gaussian location model. But the ideas and methods explored here can be used as a foundation for tackling bandwidth choice problems in nonparametrically-studentized tests and confidence interval construction in general regression settings with linear instrumental variable and GMM estimation. Results for these general settings will be reported in later work. The additional complexity of the formulae in the general case complicates the search for an optimal truncation parameter, but the main conclusion of the present findings are the same, viz. that when the goal is testing or confidence interval construction it is advantageous to reduce bias in HAC estimation by undersmoothing.

References

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

[2] Gonçalves, S. and Vogelsang T.J. (2005): "Block Bootstrap Puzzles in HAC Robust Testing: The Sophistication of the Naive Bootstrap," Working paper, University of Montreal and Cornell University.

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

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

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

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

[7] ——— (2002b): "Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation," *Econometrica,* 70, 2093–2095.

[8] ——— (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory* 21, 1130-1164.

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

[10] ——— (1994): "Automatic Lag Selection in Covariance Estimation," *Review of Economic Studies*, 61, 631–654.

[11] 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.

[12] Phillips, P. C. B., Y. Sun and S. Jin (2005a): "Improved HAR Inference Using Power Kernels without Truncation," Yale University, mimeographed.

[13] ——— (2005b): "Balancing Size and Power in Non Parametric Studentized Testing with Quadratic Power Kernels without Truncation," Department of Economics, UCSD.

[14] ——— (2006): "Spectral Density Estimation and Robust Hypothesis Testing using Steep Origin Kernels without Truncation," *International Economic Review,* 21, 837-894.

[15] ——— (2007): "Long-run Variance Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation," *Journal of Statistical Planning and Inference,* 137(3), 985-1023.

[16] Sun, Y., Phillips, P. C. B., and S. Jin (2006): "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," Department of Economics, UCSD.

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