

# A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests

Nicholas M. Kiefer and Timothy J. Vogelsang\*

Departments of Economics and Statistical Science, Cornell University

April 2002, Revised March 23, 2005

## Abstract

A new first-order asymptotic theory for heteroskedasticity-autocorrelation (HAC) robust tests based on nonparametric covariance matrix estimators is developed. The bandwidth of the covariance matrix estimator is modeled as a fixed proportion of the sample size. This leads to a distribution theory for HAC robust tests that explicitly captures the choice of bandwidth and kernel. This contrasts with the traditional asymptotics (where the bandwidth increases slower than the sample size) where the asymptotic distributions of HAC robust tests do not depend on the bandwidth or kernel. Finite sample simulations show that the new approach is more accurate than the traditional asymptotics. The impact of bandwidth and kernel choice on size and power of  $t$ -tests is analyzed. Smaller bandwidths lead to tests with higher power but greater size distortions and large bandwidths lead to tests with lower power but less size distortions. Size distortions across bandwidths increase as the serial correlation in the data becomes stronger. Overall, the results clearly indicate that for bandwidth and kernel choice there is a trade-off between size distortions and power. Finite sample performance using the new asymptotics is comparable to the bootstrap which suggests the asymptotic theory in this paper could be useful in understanding the theoretical properties of the bootstrap when applied to HAC robust tests.

Keywords: Covariance matrix estimator, inference, autocorrelation, truncation lag, prewhitening, generalized method of moments, functional central limit theorem, bootstrap.

---

\*We thank an editor and a referee for constructive comments on a previous version of the paper. Helpful comments provided by Cliff Hurvich, Andy Levin, Jeff Simonoff and seminar participants at NYU (Statistics), U. Texas Austin, Yale, U. Montreal, UCSD, UC Riverside, UC Berkeley, U. of Pittsburgh, SUNY Albany, U. Aarhus, Brown U., NBER/NSF Time Series Conference and 2003 Winter Meetings of the Econometrics Society are gratefully acknowledged. We gratefully acknowledge financial support from the National Science Foundation through grant SES-0095211. We thank the Center for Analytic Economics at Cornell University. Corresponding Author: Tim Vogelsang, Department of Economics, Uris Hall, Cornell University, Ithaca, NY 14853-7601, Phone: 607-255-5108, Fax: 607-255-2818, email: tjv2@cornell.edu

## 1 Introduction

We provide a new and improved approach to the asymptotics of hypothesis testing in time series models with “arbitrary,” i.e. unspecified, serial correlation and heteroskedasticity. Our results are general enough to apply to stationary models estimated by generalized method of moments (GMM). Heteroskedasticity and autocorrelation consistent (HAC) estimation and testing in these models involves calculating an estimate of the spectral density at zero frequency of the estimating equations or moment conditions defining the estimator. We focus on the class of nonparametric spectral density estimators<sup>1</sup> which were originally proposed and analyzed in the time series statistics literature. See Priestley (1981) for the standard textbook treatment. Important contributions to the development of these estimators for covariance matrix estimation in econometrics include White (1984), Newey and West (1987), Gallant (1987), Gallant and White (1988), Andrews (1991), Andrews and Monahan (1992), Hansen (1992), Robinson (1998), de Jong and Davidson (2000) and Jansson (2002).

We stress at the outset that we are not proposing new estimators or statistics; rather we focus on improving the asymptotic distribution theory for existing techniques. Our results provide a framework that can be used to assess the impact of the choice of HAC estimator on the resulting test statistic.

Conventional asymptotic theory for HAC estimators is well established and has proved useful in providing practical formulas for estimating asymptotic variances. The ingenious “trick” is the assumption that the variance estimator depends on a fraction of sample autocovariances, with the number of sample autocovariances going to infinity, but the fraction going to zero as the sample size grows. Under this condition it has been shown that well-known HAC estimators of the asymptotic variance are consistent. Then, the asymptotic distribution of estimated coefficients can essentially be derived assuming the variance is known. That is, sampling bias and variance of the variance estimator does not appear in the first-order asymptotic distribution theory of test statistics regarding parameters of interest. While this is an extremely productive simplifying assumption that leads to standard asymptotic distribution theory for tests, the accuracy of the resulting asymptotic theory is often less than satisfactory. In particular there is a tendency for HAC robust tests to over reject (sometimes substantially) under the null hypothesis in finite samples; see Andrews (1991), Andrews and Monahan (1992), and the July 1996 special issue of *Journal of Business and Economic Statistics* for evidence.

There are two main sources of finite sample distortions. The first source is inaccuracy via the central limit theorem approximation to the sampling distribution of parameters of interest. This becomes a serious problem for data that has strong or persistent serial correlation. The second source is the bias and sampling variability of the HAC estimate of the asymptotic variance. This

second source is the focus of this paper. Simply appealing to a consistency result for the asymptotic variance estimator as is done under the standard approach does not capture these important small sample properties.

The assumption that the fraction of the sample autocovariances used in calculating the HAC variance estimator goes to zero as the sample size goes to infinity is a clever technical assumption that substantially simplifies asymptotic calculations. However, in practice there is a given sample size and some fraction of sample autocovariances is used to estimate the asymptotic variance. Even if a practitioner chooses the fraction based on a rule such that the fraction goes to zero as the sample size grows, it does not change the fact that a positive fraction is being used for a particular data set. The implications of this simple observation have been eloquently summarized by Neave (1970, p.70) in the context of spectral density estimation:

“When proving results on the asymptotic behavior of estimates of the spectrum of a stationary time series, it is invariably assumed that as the sample size  $T$  tends to infinity, so does the truncation point  $M$ , but at a slower rate, so that  $M/T$  tends to zero. This is a convenient assumption mathematically in that, in particular, it ensures consistency of the estimates, but it is unrealistic when such results are used as approximations to the finite case where the value of  $M/T$  cannot be zero”.

Based on this observation, Neave (1970) derived an asymptotic approximation for the sampling variance of spectral density estimates under the assumption that  $M/T$  is a constant and showed that his approximation was more accurate than the standard approximation.

In this paper, we effectively generalize the approach of Neave (1970) for zero frequency nonparametric spectral density estimators (HAC estimators). We derive the entire asymptotic distribution (rather than just the variance) of these estimators under the assumption that  $M = bT$  where  $b \in (0, 1]$  is a constant. We label asymptotics obtained under this nesting of the bandwidth, “fixed- $b$  asymptotics”. In contrast, under the standard asymptotics  $b$  goes to zero as  $T$  increases. Therefore, we refer to the standard asymptotics as “small- $b$  asymptotics”. We show that under fixed- $b$  asymptotics, the HAC robust variance estimators converge to a limiting random matrix that is proportional to the unknown asymptotic variance and has a limiting distribution that depends on the kernel and  $b$ . Under the fixed- $b$  asymptotics, HAC robust test statistics computed in the usual way are shown to have limiting distributions that are pivotal but depend on the kernel and  $b$ . This contrasts with small- $b$  asymptotics where the effects of the kernel and bandwidth are not captured.

Fixed- $b$  asymptotics is a natural generalization of ideas explored by Kiefer, Vogelsang and Bunzel (2000), Bunzel, Kiefer and Vogelsang (2001), Kiefer and Vogelsang (2002a), Kiefer and Vogelsang (2002b) and Vogelsang (2003). Those papers analyzed HAC robust tests for the case

where the HAC bandwidth is set equal to the sample size, i.e.  $b = 1$ . By considering values of  $b < 1$  in this paper, we follow the traditional approach where the bandwidth controls the downweighting on higher order sample autocovariances. An alternative has been proposed in recent work by Phillips, Sun and Jin (2003) and Phillips, Sun and Jin (2005) where downweighting is achieved by exponentiating kernels that use bandwidth equal to the sample size. They show that if the exponent increases with the sample size at a suitable rate, then consistent and asymptotically normal HAC variance estimators can be obtained. In an analysis analogous to fixed- $b$  asymptotics, they also develop an asymptotic theory under the assumption that the exponent is fixed as the sample size increases. Fixed-exponent asymptotics captures the choice of kernel and exponent in much the same way that fixed- $b$  asymptotics captures the choice of kernel and bandwidth.

While the fixed- $b$  assumption is a better reflection of practice in reality, that alone does not justify the new asymptotic theory. In fact, our asymptotic theory leads to two important innovations for HAC robust testing.

First, because the fixed- $b$  asymptotics explicitly captures the choice of bandwidth and kernel, a more accurate first-order asymptotic approximation is obtained for HAC robust tests. Finite sample simulations reported here and in the working paper, Kiefer and Vogelsang (2002c), show that in many situations the fixed- $b$  asymptotics provide a more accurate approximation than the standard small- $b$  asymptotics. There is also theoretical evidence by Jansson (2004) showing that fixed- $b$  asymptotics is more accurate than small- $b$  asymptotics in Gaussian location models for the special case of the Bartlett kernel with  $b = 1$ . Jansson (2004) proves that fixed- $b$  asymptotics delivers an error in rejection probability that is  $O(T^{-1} \log(T))$ . This contrasts with small- $b$  asymptotics where the error in rejection probability is no smaller than  $O(T^{-1/2})$  (see Velasco and Robinson (2001)). Again focusing on Gaussian location models, recent work by Phillips, Sun and Jin (2004) has shown that the error in rejection probability is  $O(T^{-1})$  for exponentiated kernels under fixed-exponent asymptotics. Phillips et al. (2004) conjecture that this result likely extends to traditional kernels under fixed- $b$  asymptotics. If this conjecture is true, then fixed- $b$  asymptotics has an error in rejection probability of smaller order than the standard normal approximation (small- $b$ ).

Second, fixed- $b$  asymptotic theory permits a local asymptotic power analysis for HAC robust tests that depends on the kernel and bandwidth. We can theoretically analyze how the choices of kernel and bandwidth affect the power of HAC robust tests. Such an analysis is not possible under the standard small- $b$  asymptotics because local asymptotic power does not depend on the choice of kernel or bandwidth. Because of this fact, the existing HAC robust testing literature has focused instead on minimizing the asymptotic truncated MSE of the asymptotic variance estimators when choosing the kernel and bandwidth. For the analysis of HAC robust tests, this is not a completely satisfying situation as noted by Andrews (1991, p.828)<sup>2</sup>.

An obvious alternative to asymptotic approximations is the bootstrap. Recently it has been

shown by Hall and Horowitz (1996), Götze and Künsch (1996), Inoue and Shintani (2004) and others that higher order refinements to the small- $b$  asymptotics are feasible when bootstrapping the distribution of HAC robust tests using blocking. Using finite sample simulations we compare fixed- $b$  asymptotics with the block bootstrap. Our results suggest some interesting properties of the bootstrap and indicate that fixed- $b$  asymptotics may be a useful analytical tool for understanding variation in bootstrap performance across bandwidths. One result is that the bootstrap without blocking performs almost identically to the fixed- $b$  asymptotics even when the data are dependent (in contrast the small- $b$  asymptotics performs relatively poorly). When blocking is used, the bootstrap can perform slightly better or slightly worse than fixed- $b$  asymptotics depending on the choice of block length. It may be the case that the block bootstrap delivers an asymptotic refinement over the fixed- $b$  first-order asymptotics if the block length is chosen in a suitable way. A higher order fixed- $b$  asymptotic theory is required before such a result can be formally established.

The remainder of the paper is organized as follows. Section 2 lays out the GMM framework and reviews standard results. Section 3 reports some small sample simulation results that illustrate the inaccuracies that can occur when using the small- $b$  asymptotic approximation. Section 4 introduces the new fixed- $b$  asymptotic theory. Section 5 analyzes the performance of the new asymptotic theory in terms of size distortions and local asymptotic power. The impact of the choice of bandwidth and kernel is analyzed and comparisons are made with the traditional small- $b$  asymptotics and the block bootstrap. Section 6 gives concluding comments. Proofs are provided in an appendix.

The following notation is used throughout the paper. The symbol  $\Rightarrow$  denotes weak convergence,  $B_j(r)$  denotes a  $j$  vector of standard Brownian motions (Wiener processes) defined on  $r \in [0, 1]$ ,  $\tilde{B}_j(r) = B_j(r) - rB_j(1)$  denotes a  $j$  vector of standard Brownian bridges, and  $[rT]$  denotes the integer part of  $rT$  for  $r \in [0, 1]$ .

## 2 Inference in GMM Models: The Standard Approach

We present our results in the GMM framework noting that this covers estimating equation methods (Heyde, 1997). Since the influential work of Hansen (1982), GMM is widely used in virtually every field of economics. Heteroskedasticity or autocorrelation of unknown form is often an important specification issue especially in macroeconomics and financial applications. Typically the form of the correlation structure is not of direct interest (if it is, it should be modeled directly). What is desired is an inference procedure that is robust to the form of the heteroskedasticity and serial correlation. HAC covariance matrix estimators were developed for exactly this setting.

Consider the  $p \times 1$  vector of parameters,  $\theta \in \Theta \subset R^p$ . Let  $\theta_0$  denote the true value of  $\theta$ , and assume  $\theta_0$  is an interior point of  $\Theta$ . Let  $v_t$  denote a vector of observed data and assume that  $q$

moment conditions hold that can be written as

$$E[f(v_t, \theta_0)] = \mathbf{0}, \quad t = 1, 2, \dots, T, \quad (1)$$

where  $f(\cdot)$  is a  $q \times 1$  vector of functions with  $q \geq p$  and  $\text{rank}(E[\partial f / \partial \theta']) = p$ . The expectation is taken over the endogenous variables in  $v_t$ , and may be conditional on exogenous elements of  $v_t$ . There is no need in what follows to make this conditioning explicit in the notation. Define

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

where  $g_T(\theta) = T^{-1} \sum_{t=1}^T f(v_t, \theta)$  is the sample analog to (1). The GMM estimator is defined as

$$\hat{\theta}_T = \arg \min_{\theta \in \Theta} g_T(\theta)' W_T g_T(\theta) \quad (2)$$

where  $W_T$  is a  $q \times q$  positive definite weighting matrix. Alternatively,  $\hat{\theta}_T$  can also be defined as an estimating equations estimator, the solution to the  $p$  first-order conditions associated with (2)

$$G_T(\hat{\theta}_T)' W_T g_T(\hat{\theta}_T) = \mathbf{0}, \quad (3)$$

where  $G_t(\theta) = T^{-1} \sum_{j=1}^t \partial f(v_j, \theta) / \partial \theta'$ . Of course, when the model is exactly identified and  $q = p$ , an exact solution to  $g_T(\hat{\theta}_T) = \mathbf{0}$  is attainable and the weighting matrix  $W_T$  is irrelevant. Application of the mean-value theorem implies that

$$g_t(\hat{\theta}_T) = g_t(\theta_0) + G_t(\hat{\theta}_T, \theta_0, \lambda_T)(\hat{\theta}_T - \theta_0) \quad (4)$$

where  $G_t(\hat{\theta}_T, \theta_0, \lambda_T)$  denotes the  $q \times p$  matrix whose  $i^{\text{th}}$  row is the corresponding row of  $G_t(\bar{\theta}_T^{(i)})$  where  $\bar{\theta}_T^{(i)} = \lambda_{i,T} \theta_0 + (1 - \lambda_{i,T}) \hat{\theta}_T$  for some  $0 \leq \lambda_{i,T} \leq 1$  and  $\lambda_T$  is the  $q \times 1$  vector with  $i^{\text{th}}$  element  $\lambda_{i,T}$ .

In order to focus on the new asymptotic theory for tests, we avoid listing primitive assumptions and make rather high-level assumptions on the GMM estimator  $\hat{\theta}_T$ . Lists of sufficient conditions for these to hold can be found in Hansen (1982) and Newey and McFadden (1994). Our assumptions are:

**Assumption 1**  $p \lim \hat{\theta}_T = \theta_0$ .

**Assumption 2**  $T^{-1/2} \sum_{t=1}^{\lceil rT \rceil} f(v_t, \theta_0) = T^{1/2} g_{\lceil rT \rceil}(\theta_0) \Rightarrow \Lambda B_q(r)$  where  $\Lambda \Lambda' = \Omega = \sum_{j=-\infty}^{\infty} \Gamma_j$ ,  $\Gamma_j = E[f(v_t, \theta_0) f(v_{t-j}, \theta_0)']$ .

**Assumption 3**  $p \lim G_{\lceil rT \rceil}(\hat{\theta}_T) = rG_0$  and  $p \lim G_{\lceil rT \rceil}(\hat{\theta}_T, \theta_0, \lambda_T) = rG_0$  uniformly in  $r \in [0, 1]$  where  $G_0 = E[\partial f(v_t, \theta_0) / \partial \theta']$ .

**Assumption 4**  $W_T$  is positive semi-definite and  $\text{plim } W_T = W_\infty$  where  $W_\infty$  is a matrix of constants and  $G_0'W_\infty G_0$  is positive definite.

These assumptions hold for many models in economics, and with the exception of Assumption 2 they are fairly standard. Assumption 2 requires that a functional central limit theorem hold for  $T^{1/2}g_t(\theta_0)$ . This is stronger than the central limit theorem for  $T^{1/2}g_T(\theta_0)$  that is required for asymptotic normality of  $\hat{\theta}_T$ . However, consistent estimation of the asymptotic variance of  $\hat{\theta}_T$  requires an estimate of  $\Omega$ . Conditions for consistent estimation of  $\Omega$  are typically stronger than Assumption 2 and often imply Assumption 2. For example, Andrews (1991) requires that  $f(v_t, \theta_0)$  is a mean zero fourth order stationary process that is  $\alpha$ -mixing. Phillips and Durlauf (1986) show that Assumption 2 holds under the weaker assumption that  $f(v_t, \theta_0)$  is a mean zero,  $2 + \delta$  order stationary process (for some  $\delta > 0$ ) that is  $\alpha$ -mixing. Thus our assumptions are slightly weaker than those usually given for asymptotic testing in HAC-estimated GMM models.

Under our assumptions  $\hat{\theta}_T$  is asymptotically normally distributed, as recorded in the following lemma which is proved in the appendix.

**Lemma 1** Under Assumptions 1 - 4, as  $T \rightarrow \infty$ ,

$$T^{1/2}(\hat{\theta}_T - \theta_0) \Rightarrow -(G_0'W_\infty G_0)^{-1}\Lambda^*B_p(1) \sim N(0, V),$$

where  $\Lambda^*\Lambda^{*'} = G_0'W_\infty\Lambda\Lambda'W_\infty G_0$  and  $V = (G_0'W_\infty G_0)^{-1}\Lambda^*\Lambda^{*'}(G_0'W_\infty G_0)^{-1}$ .

Under the standard approach, a consistent estimator of  $V$  is required for inference. Let  $\hat{\Omega}$  denote an estimator of  $\Omega$ . Then  $V$  can be estimated by

$$\hat{V} = [G_T(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T)]^{-1}G_T(\hat{\theta}_T)'W_T\hat{\Omega}W_T G_T(\hat{\theta}_T)[G_T(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T)]^{-1}. \quad (5)$$

The HAC literature builds on the spectral density estimation literature to suggest feasible estimators of  $\Omega$  and to find conditions under which such estimators are consistent. The widely used class of nonparametric estimators of  $\Omega$  take the form

$$\hat{\Omega} = \sum_{j=-(T-1)}^{T-1} k(j/M)\hat{\Gamma}_j \quad (6)$$

with

$$\hat{\Gamma}_j = T^{-1} \sum_{t=j+1}^T f(v_t, \hat{\theta}_T)f(v_{t-j}, \hat{\theta}_T)' \text{ for } j \geq 0, \quad \hat{\Gamma}_j = \hat{\Gamma}'_{-j} \text{ for } j < 0,$$

where  $k(x)$  is a kernel function  $k : R \rightarrow R$  satisfying  $k(x) = k(-x)$ ,  $k(0) = 1$ ,  $|k(x)| \leq 1$ ,  $k(x)$  continuous at  $x = 0$  and  $\int_{-\infty}^{\infty} k^2(x)dx < \infty$ . Often  $k(x) = 0$  for  $|x| > 1$  so  $M$  “trims” the sample autocovariances and acts as a truncation lag. Some popular kernel functions do not truncate, and

$M$  is often called a bandwidth parameter in those cases. For kernels that truncate, the cutoff at  $|x| = 1$  is arbitrary and is essentially a normalization. For kernels that do not truncate, a normalization must be made since the weights generated by the kernel  $k(x)$  and bandwidth,  $M$  are the same as those generated by kernel  $k(ax)$  with bandwidth  $aM$ . Therefore, there is an interaction between bandwidth and kernel choice. We focus on kernels that yield positive definite  $\widehat{\Omega}$  for the obvious practical reasons although many of our theoretical results hold without this restriction.

Standard asymptotic analysis proceeds under the assumption that  $M \rightarrow \infty$  and  $M/T \rightarrow 0$  as  $T \rightarrow \infty$  in which case  $\widehat{\Omega}$  is a consistent estimator. Thus,  $b \rightarrow 0$  as  $T \rightarrow \infty$  and hence the label, small- $b$  asymptotics. Because in practical settings  $b$  is strictly positive, the assumption that  $b$  shrinks to zero has little to do with econometric practice; rather it is an ingenious technical assumption allowing an estimable asymptotic approximation to the asymptotic distribution of  $\widehat{\theta}_T$  to be calculated. The difficulty in practice is that any choice of  $M$  for a given sample size,  $T$ , can be made consistent with the above rate requirement. Although the rate requirement can be refined if one is interested in minimizing the MSE of  $\widehat{\Omega}$  (e.g.  $M$  must increase at rate  $T^{1/3}$  for the Bartlett kernel), these refinements do not deliver specific choices for  $M$ . This fact has long been recognized in the spectral density and HAC literatures and data dependent methods for choosing  $M$  have been proposed. See Andrews (1991) and Newey and West (1994). These papers suggest choosing  $M$  to minimize the truncated MSE of  $\widehat{\Omega}$ . However, because the MSE of  $\widehat{\Omega}$  depends on the serial correlation structure of  $f(v_t, \theta_0)$ , the practitioner must estimate the serial correlation structure of  $f(v_t, \theta_0)$  either nonparametrically or with an “approximate” parametric model. While data dependent methods are a significant improvement over the basic case for empirical implementation, the practitioner is still faced with either a choice of approximating parametric model or the choice of bandwidth in a preliminary nonparametric estimation problem. Even these bandwidth rules do not yield unique bandwidth choices in practice because, for example, if a bandwidth,  $M_A$  satisfies the optimality criterion of Andrews (1991), then the bandwidth,  $M_A + d$  where  $d$  is a finite constant is also optimal. See den Haan and Levin (1997) for details and additional practical challenges.

As the previous discussion makes clear, the standard approach addresses the choice of bandwidth by determining the rate by which  $M$  must grow to deliver a consistent variance estimator so that the usual standard normal approximation can be justified. In contrast, for a given data set and sample size, we take the choice of  $M$  as given and provide an asymptotic theory that reflects to some extent how that choice of  $M$  affects the sampling distribution of the HAC robust test.

### 3 Motivation: Finite Sample Performance of Standard Asymptotics

While the poor finite sample performance of HAC robust test in the presence of strong serial correlation is well documented, it will be useful for later comparisons to illustrate some of these problems briefly in a simple environment. Consider the most basic univariate time series model

with ARMA(1,1) errors,

$$\begin{aligned}
 y_t &= \mu + u_t, \\
 u_t &= \rho u_{t-1} + \varepsilon_t + \varphi \varepsilon_{t-1}, \\
 \varepsilon_t &\sim iidN(0, 1), \\
 u_0 &= \varepsilon_0 = 0.
 \end{aligned} \tag{7}$$

Define the HAC robust  $t$ -statistic for  $\mu$  as

$$t = \frac{\sqrt{T}(\hat{\mu} - \mu)}{\sqrt{\hat{\Omega}}},$$

where  $\hat{\mu} = \bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$  and  $\hat{\Omega}$  is computed using  $f(y_t, \hat{\mu}) = y_t - \hat{\mu}$ . We set  $\mu = 0$  and generated data according to (7) for a wide variety of parameter values and computed empirical rejection probabilities of the  $t$ -statistic for testing the null hypothesis that  $\mu \leq 0$  against the alternative that  $\mu > 0$ . Note that because the test statistic has a symmetric distribution, the performance of a two sided version of the test is qualitatively similar. We report results for the sample size  $T = 50$  and 5,000 replications were used in all cases. To illustrate how well the standard normal approximation works as the bandwidth varies in this finite sample, we computed rejection probabilities for the  $t$ -statistic implemented using  $M = 1, 2, 3, \dots, 49, 50$ . We set the asymptotic significance level to 0.05 and used the usual standard normal critical value of 1.645 for all values of  $M$ . To conserve on space, we report results for six error configurations: *iid* errors ( $\rho = \varphi = 0$ ), AR(1) errors with  $\rho = 0.7, 0.9$ , MA(1) errors with  $\varphi = -0.5, 0.5$  and ARMA(1,1) errors with  $\rho = 0.8, \varphi = -0.4$ . We also give results where  $\hat{\Omega}$  is implemented with AR(1) prewhitening. We report results for the popular Bartlett and quadratic spectral (QS) kernels. Results for other kernels are similar.

The results are depicted in Figures 1 and 2. In each figure, the line with the label, N(0,1), plots rejection probabilities when the critical value 1.645 is used and there is no prewhitening. In the case of prewhitening, the label is N(0,1), PW. The figures also depict plots of rejection probabilities using the fixed- $b$  asymptotic critical values but for now we only focus on the results using the standard small- $b$  asymptotics.

Consider first the case of *iid* errors. When  $M$  is small rejection probabilities are only slightly above 0.05 as one might expect for *iid* data. However, as  $M$  increases, rejection probabilities steadily rise. When  $M = T$ , rejection probabilities are nearly 0.2 for the Bartlett kernel and exceed 0.25 for the QS kernel. Somewhat different patterns occur for AR(1) errors. When  $M$  is small, there are nontrivial over-rejection problems. When prewhitening is not used, the rejections fall as  $M$  increases but then rise again as  $M$  increases further. When prewhitening is used, rejection probabilities essentially increase as  $M$  increases. Prewhitening reduces the extent of over-rejection but does not remove it. The patterns for MA(1) errors are similar to the *iid* case and the patterns for ARMA(1,1) errors are combinations of the AR(1) and MA(1) cases.

These plots clearly illustrate a fundamental finite sample property of HAC robust tests: the choice of  $M$  in a given sample matters and can greatly affect the extent of size distortion depending on the serial correlation structure of the data. And, when  $M$  is not very small, the choice of kernel also matters for the over-rejection problem. Given that the sampling distribution of the test depends on the bandwidth and kernel, it would seem to be a minimal requirement that the asymptotic approximation reflect this dependence. Whereas the standard small- $b$  asymptotics is too crude for this purpose, our results in the next section show that the fixed- $b$  asymptotics naturally captures the dependence on the bandwidth and kernel and captures the dependence in a simple and elegant manner<sup>3</sup>.

Before moving on, it is useful to discuss in some detail the reason that rejection probabilities have the non-monotonic pattern for AR(1) errors. When  $M$  is small,  $\widehat{\Omega}$  is biased but has relatively small variance. Because the bias is downward, that leads to over-rejection. As  $M$  initially increases, bias falls and variance rises. The bias effect is more important for over-rejection (see Simonoff (1993)) and the extent of over-rejection decreases. According to the conventional, but wrong, wisdom, the story would be that as  $M$  increases further bias continues to fall but variance increases so much that over-rejections become worse<sup>4</sup>. This is not what is happening. In fact, as  $M$  increases further, bias eventually increases and variance falls. It is the increase in bias that leads to the large over-rejections when  $M$  is large. The reason that bias increases and variance shrinks as  $M$  gets large is easy to explain. When  $M$  is large, high weights are placed on high order sample autocovariances. In the extreme case where full weight is placed on all of the sample autocovariances, it is well known that  $\widehat{\Omega} = 0$ , and this occurs because  $\widehat{\Omega}$  is computed using residuals that have a sample average of zero. Obviously,  $\widehat{\Omega} = 0$  it is an estimator with large bias and zero variance. Thus, as  $M$  increases,  $\widehat{\Omega}$  is pushed closer to the full weight case.

## 4 A New Asymptotic Theory

### 4.1 Distribution of $\widehat{\Omega}$ in the Fixed- $b$ Asymptotic Framework

We now develop a distribution theory for  $\widehat{\Omega}$  in the fixed- $b$  asymptotic framework<sup>5</sup>. We proceed under the asymptotic nesting that  $M = bT$  where  $b \in (0, 1]$  is fixed. The limiting distribution of  $\widehat{\Omega}$  in the fixed- $b$  asymptotic framework can be written in terms of  $Q_i(b)$ , an  $i \times i$  random matrix that takes on one of three forms depending on the second derivative of the kernel. The following definition gives the forms of  $Q_i(b)$ .

**Definition 1** *Let the  $i \times i$  random matrix,  $Q_i(b)$  be defined as follows. Case (i): if  $k(x)$  is twice continuously differentiable everywhere,*

$$Q_i(b) = - \int_0^1 \int_0^1 \frac{1}{b^2} k''\left(\frac{r-s}{b}\right) \widetilde{B}_i(r) \widetilde{B}_i(s)' dr ds.$$

Case (ii): if  $k(x)$  is continuous,  $k(x) = 0$  for  $|x| \geq 1$ , and  $k(x)$  is twice continuously differentiable everywhere except for  $|x| = 1$ ,

$$Q_i(b) = - \iint_{|r-s|<b} \frac{1}{b^2} k''\left(\frac{r-s}{b}\right) \tilde{B}_i(r) \tilde{B}_i(s)' dr ds \\ + \frac{k'_-(1)}{b} \int_0^{1-b} \left( \tilde{B}_i(r+b) \tilde{B}_i(r)' + \tilde{B}_i(r) \tilde{B}_i(r+b)' \right) dr.$$

where  $k'_-(1) = \lim_{h \rightarrow 0} [(k(1) - k(1-h))/h]$ , i.e.  $k'_-(1)$  is the derivative of  $k(x)$  from the left at  $x = 1$ . Case (iii): if  $k(x)$  is the Bartlett kernel

$$Q_i(b) = \frac{2}{b} \int_0^1 \tilde{B}_i(r) \tilde{B}_i(r)' dr - \frac{1}{b} \int_0^{1-b} \left( \tilde{B}_i(r+b) \tilde{B}_i(r)' + \tilde{B}_i(r) \tilde{B}_i(r+b)' \right) dr.$$

The moments of  $Q_i(1)$  have been derived by Phillips et al. (2003) for the case of positive definite kernels. Hashimzade, Kiefer and Vogelsang (2003) have generalized those results for  $Q_i(b)$  while also relaxing the positive definite requirement. The moments are

$$E(Q_i(b)) = I_i \left( 1 - \iint k\left(\frac{r-s}{b}\right) dr ds \right),$$

$$\text{var}(\text{vec}(Q_i(b))) = \nu(b) (I_{i^2} + \kappa_{ii}),$$

$$\nu(b) = \left[ \iint k\left(\frac{r-s}{b}\right) dr ds \right]^2 - 2 \iiint k\left(\frac{r-s}{b}\right) k\left(\frac{r-q}{b}\right) dr ds dq + \iint k\left(\frac{r-s}{b}\right)^2 dr ds,$$

where  $I_i$  is the  $i \times i$  identity matrix,  $\kappa_{ii}$  is the standard commutation matrix<sup>6</sup> and the range of all integrals is 0 to 1. Using these moment results, Hashimzade et al. (2003) prove that  $\lim_{b \rightarrow 0} E(Q_i(b)) = I_i$  and  $\lim_{b \rightarrow 0} \text{var}(\text{vec}(Q_i(b))) = \mathbf{0}$ . As an illustrative example, consider the Bartlett kernel where

$$E(Q_i(b)) = I_i \left( 1 - b + \frac{1}{3} b^2 \right),$$

$$\nu(b) = \frac{4}{3} b - \frac{7}{3} b^2 + \frac{14}{15} b^3 + \frac{2}{9} b^4 \text{ for } b \leq \frac{1}{2}.$$

We first consider the asymptotic distribution of  $\hat{\Omega}$  for the case of exactly identified models.

**Theorem 1** (*Exactly Identified Models*) Suppose that  $q = p$ . Let  $M = bT$  where  $b \in (0, 1]$  is fixed. Let  $Q_p(b)$  be given by Definition 1 for  $i = p$ . Then, under Assumptions 1-4, as  $T \rightarrow \infty$ ,

$$\hat{\Omega} \Rightarrow \Lambda Q_p(b) \Lambda'.$$

Several useful observations can be made regarding this theorem. Under fixed- $b$  asymptotics,  $\hat{\Omega}$  converges to a matrix of random variables (rather than constants) that is proportional to  $\Omega$  through

$\Lambda$  and  $\Lambda'$ . This contrasts with the small- $b$  asymptotic approximation where  $\widehat{\Omega}$  is approximated by the constant  $\Omega$ . As  $b \rightarrow 0$ , it follows from Lemma 1 that  $p \lim_{b \rightarrow 0} \Lambda Q_p(b) \Lambda' = \Omega$ . Thus, the fixed- $b$  asymptotics coincides with the standard small- $b$  asymptotics as  $b$  goes to zero. The advantage of the fixed- $b$  asymptotic result is that limit of  $\widehat{\Omega}$  depends on the kernel through  $k''(x)$  and  $k'_-(1)$  and on the bandwidth through  $b$  but are otherwise nuisance parameter free. Therefore, it is possible to obtain a first-order asymptotic distribution theory that explicitly captures the choice of kernel and bandwidth. Under fixed- $b$  asymptotics, any choice of bandwidth leads to asymptotically pivotal tests of hypotheses regarding  $\theta_0$  when using  $\widehat{\Omega}$  to construct standard errors (details are given below). Note that Theorem 1 generalizes results obtained by Vogelsang (2003) where the focus was  $b = 1$ .

When  $q > p$  and the model is overidentified, the limiting expressions for  $\widehat{\Omega}$  are more complicated and asymptotic proportionality to  $\Omega$  no longer holds. This was established for the special case of  $b = 1$  by Vogelsang (2003). This does not imply, however, that valid testing is not possible when using  $\widehat{\Omega}$  in overidentified models because the required asymptotic proportionality does hold for  $G_T(\widehat{\theta}_T)' W_T \widehat{\Omega} W_T G_T(\widehat{\theta}_T)$ , the middle term in  $\widehat{V}$ . The following theorem provides the relevant result.

**Theorem 2** (*Over-identified Models*) *Suppose that  $q > p$ . Let  $M = bT$  where  $b \in (0, 1]$  is fixed. Let  $Q_p(b)$  be given by Definition 1 for  $i = p$ . Define  $\Lambda^* = G'_0 W_\infty \Lambda$ . Under Assumptions 1-4, as  $T \rightarrow \infty$ ,*

$$G_T(\widehat{\theta}_T)' W_T \widehat{\Omega} W_T G_T(\widehat{\theta}_T) \Rightarrow \Lambda^* Q_p(b) \Lambda^{*'}.$$

This theorem shows that  $G_T(\widehat{\theta}_T)' W_T \widehat{\Omega} W_T G_T(\widehat{\theta}_T)$  is asymptotically proportional to  $\Lambda^* \Lambda^{*'}$  and otherwise only depends on the random matrix  $Q_p(b)$ . It directly follows that  $\widehat{V}$  is asymptotically proportional to  $V$ , and asymptotically pivotal tests can be obtained.

## 4.2 Inference

We now examine the limiting null distributions of tests regarding  $\theta_0$  under fixed- $b$  asymptotics. Consider the hypotheses

$$\begin{aligned} H_0 &: r(\theta_0) = 0 \\ H_1 &: r(\theta_0) \neq 0 \end{aligned}$$

where  $r(\theta)$  is an  $m \times 1$  vector ( $m \leq p$ ) of continuously differentiable functions with first derivative matrix,  $R(\theta) = \partial r(\theta) / \partial \theta'$ . Applying the delta method to Lemma 1 we obtain

$$T^{1/2} r(\widehat{\theta}_T) \Rightarrow -R(\theta_0) V^{1/2} B_p(1) \equiv N(\mathbf{0}, V_R), \quad (8)$$

where  $V_R = R(\theta_0)VR(\theta_0)'$ . Using (8) one can construct the standard HAC robust Wald test of the null hypothesis or a  $t$ -test in the case of  $m = 1$ . To remain consistent with earlier work, we consider the F-test version of the Wald statistic defined as

$$F = Tr(\widehat{\theta}_T)' \left( R(\widehat{\theta}_T)\widehat{V}R(\widehat{\theta}_T)' \right)^{-1} r(\widehat{\theta}_T)/m,$$

When  $m = 1$  the usual  $t$ -statistic can be computed as

$$t = \frac{T^{1/2}r(\widehat{\theta}_T)}{\sqrt{R(\widehat{\theta}_T)\widehat{V}_{M=bT}R(\widehat{\theta}_T)'}}$$

Often, the significance of individual statistics are of interest which leads to  $t$ -statistics of the form

$$t = \frac{\widehat{\theta}_{iT}}{se(\widehat{\theta}_{iT})},$$

where  $se(\widehat{\theta}_i) = \sqrt{T^{-1}\widehat{V}^{ii}}$  and  $\widehat{V}^{ii}$  is the  $i^{th}$  diagonal element of the  $\widehat{V}$  matrix. To avoid any confusion, note that these statistics are being computed in exactly the same way as under the standard approach. Only the asymptotic approximation to the sampling distribution is different.

Note that some kernels, including the Tukey-Hanning, allow negative variance estimates. In this case some convention must be adopted in calculating the denominator of the test statistics. Equally arbitrary conventions include reflection of negative values through the origin or setting negatives to a small positive value. Although our results apply to kernels that are not positive definite, we see no merit in using a kernel allowing negative estimated variances absent a compelling argument in a specific case. Nevertheless, we have experimented with the Tukey-Hanning and trapezoid kernels and results not reported here do not support their consideration over a kernel guaranteeing positive variance estimates.

The following theorem provides the asymptotic null distributions of  $F$  and  $t$ .

**Theorem 3** *Let  $b \in (0, 1]$  be a constant and suppose  $M = bT$ . Let  $Q_i(b)$  be given by Definition 1 for  $i = m$ . Then, under Assumptions 1-4 and  $H_0$ , as  $T \rightarrow \infty$ ,*

$$F \Rightarrow B_m(1)'Q_m(b)^{-1}B_m(1)/m,$$

$$t \Rightarrow \frac{B_1(1)}{\sqrt{Q_1(b)}}.$$

Theorem 3 shows that under fixed- $b$  asymptotics, asymptotically pivotal tests are obtained and the asymptotic distributions reflect the choices of kernel and bandwidth. This contrasts asymptotic results under the standard approach where  $F$  would have a limiting  $\chi_m^2/m$  distribution and  $t$  a limiting  $N(0, 1)$  distribution regardless of the choice of  $M$  and  $k(x)$ . As shown by the corollary to

Theorem 2, as  $b \rightarrow 0$ ,  $p \lim Q_m(b) = I_m$  and the fixed- $b$  asymptotics reduces to the standard small- $b$  asymptotics. Therefore, if a traditional bandwidth rule is used in conjunction with the fixed- $b$  asymptotics, in large samples the two asymptotic theories will coincide. However, because the value of  $b$  is strictly greater than zero in practice, it is natural to expect fixed- $b$  asymptotics to deliver a more accurate approximation. The simulations results reported in Section 5 and in the working paper, Kiefer and Vogelsang (2002c), indicate that this is true in some Gaussian models. Finite sample results reported by Ravikumar, Ray and Savin (2004) indicate that the fixed- $b$  asymptotic approximation can substantially reduce size distortions in tests of joint hypotheses especially when the number of hypotheses being tested is large.

A theoretical comparison of the accuracy of the fixed- $b$  asymptotics with the small- $b$  asymptotics is not currently available because existing methods in higher order asymptotic expansions, such as Edgeworth expansions, do not directly apply to the fixed- $b$  asymptotic nesting given the nonstandard nature of the distribution theory. Obtaining such theoretical results appears difficult although for the Gaussian local model and the special case of the Bartlett kernel with  $b = 1$ , Jansson (2004) has shown that the error in rejection probability of  $F$  is  $O(T^{-1} \log(T))$ . Phillips et al. (2004) conjecture that this result can be strengthened to  $O(T^{-1})$  and can be generalized to include  $b < 1$  and other kernels besides the Bartlett kernel.

### 4.3 Asymptotic Critical Values

The limiting distributions given by Theorem 3 are nonstandard. Analytical forms of the densities are not available with the exception of  $t$  for the case of the Bartlett kernel with  $b = 1$  (see Abadir and Paruolo, 2002 and Kiefer and Vogelsang, 2002b). However, because the limiting distributions are simple functions of standard Brownian motions, critical values are easily obtained using simulations. In the working paper we provide critical values for the  $t$  statistic for a selection of popular kernels (see the formula appendix for formulas for the kernels). Additional critical values for the  $F$  test will be made available in a follow-up paper.

To make the use of the fixed- $b$  asymptotics easy for practitioners we provide critical value functions for the  $t$  statistic using the cubic equation

$$cv(b) = a_0 + a_1b + a_2b^2 + a_3b^3.$$

For a selection of well-known kernels, we computed the  $cv(b)$  function for the percentage points 90%, 95%, 97.5% and 99%. Critical values for the left tail follow by symmetry around zero. The  $a_i$  coefficients are given in Table 1. They were obtained as follows. For each kernel and the grid  $b = 0.02, 0.04, \dots, 0.98, 1.0$  critical values were calculated via simulation methods using 50,000 replications. Normalized partial sums of 1,000 *i.i.d.*  $N(0, 1)$  random deviates were used to approximate the standard Brownian motions in the respective distributions given by Theorem 3.

For each percentage point, the simulated critical values were used to fit the  $cv(b)$  function by OLS. The intercept was constrained to yield the standard normal critical value so that  $cv(0)$  coincides with the standard asymptotics. Table 1 also reports the  $R^2$  from the regressions and the fits are excellent in all cases ( $R^2$  ranging from 0.9825 to 0.9996).

## 5 Choice of Kernel and Bandwidth: Performance

In this section we analyze the choice of kernel and bandwidth on the performance of HAC robust tests. We focus on accuracy of the asymptotic approximation under the null and on local asymptotic power. As far as we know, our analysis is the first to explore theoretically the effects of kernel and bandwidth choice on power of HAC robust tests.

### 5.1 Accuracy of the Asymptotic Approximation under the Null and Comparison with the Bootstrap

The way to evaluate the accuracy of an asymptotic approximation to a null distribution, or indeed any approximation, is to compare the approximate distribution to the exact distribution. Sometimes this can be done analytically; more commonly the comparison can be made by simulation. We argued above that our approximation to the distribution of HAC robust tests was likely to be better than the usual approximation, since ours takes into account the randomness in the estimated variance. However, as noted, that argument is unconvincing in the absence of evidence on the approximation's performance. We provide results for two popular positive definite kernels: Bartlett and QS. Results for the Parzen, Bohman and Daniell kernel are similar and are not reported here. The working paper, Kiefer and Vogelsang (2002c), contains additional finite sample simulation results that are similar to what is reported here.

The simulations were based on the simple location model (7) with the same design as described previously. Figures 1 and 2 provide plots of empirical rejection probabilities using the fixed- $b$  asymptotics. Compared to the standard small- $b$  asymptotics, the results are striking. In nearly all cases, the tendency to over-reject is substantially reduced. The exceptions are for small values of  $M$  where the fixed- $b$  asymptotics show only small improvements. But, this is to be expected given that the two asymptotic theories coincide as  $b$  goes to 0. In the case of *iid* errors, the fixed- $b$  asymptotics is remarkably accurate. In the case of MA errors, fixed- $b$  asymptotics gives tests with rejection probabilities close to 0.05. In the case of AR errors, over-rejections decrease as  $M$  increases. Intuitively, the fixed- $b$  asymptotics is capturing the variance of  $\widehat{\Omega}$  and the downward bias in  $\widehat{\Omega}$  that can be substantial when  $M$  is large. As  $M$  increases, the tendency to over-reject falls. The QS kernel delivers tests with less size distortions than the Bartlett kernel especially when  $M$  is large<sup>7</sup>. The results in Figures 1 and 2 strongly suggest that the use of fixed- $b$  asymptotic critical values greatly improves the performance of HAC robust t-tests.

Note however, that as serial correlation in the errors becomes stronger, the tendency to over-reject increases even when fixed- $b$  critical values are used. The reason is that the functional central limit theorem approximation begins to erode. Therefore, there may be potential for the bootstrap to provide further refinements in accuracy. The fixed- $b$  asymptotics suggests that the bootstrap could work because the HAC robust tests are asymptotically pivotal for any kernel and bandwidth. Ideally, one would like to establish theoretically whether or not the bootstrap provides an asymptotic refinement. Such a theoretical exercise is well beyond the scope of this paper because higher order expansions for fixed- $b$  asymptotics have not been developed. Obtaining such expansions is an important topic that deserves attention.

To show the potential of the bootstrap for HAC robust tests, we expanded our simulation study to include the block bootstrap. We implemented the block bootstrap resampling scheme following Götze and Künsch (1996) as follows. The original data,  $y_1, y_2, \dots, y_T$  is divided into  $T - l + 1$  overlapping blocks of length,  $l$ . For simplicity, we report results for  $l = 1, 4$  both of which factor into  $T = 50$ .  $T/l$  blocks are drawn randomly with replacement to form a bootstrap sample labeled  $y_1^*, y_2^*, \dots, y_T^*$ . The bootstrap  $t$  statistic is computed as

$$t_{Naive}^* = \frac{\sqrt{T}(\hat{\mu}^* - \bar{y})}{\sqrt{\hat{\Omega}^*}},$$

where  $\hat{\mu}^* = \bar{y}^* = T^{-1} \sum_{t=1}^T y_t^*$  and  $\hat{\Omega}^*$  is computed using formula (6) with  $f(y_t^*, \hat{\mu}^*) = \hat{u}_t^* = y_t^* - \hat{\mu}^*$ . We use the label "naive" because the bootstrap  $t$ -statistic is computed using the same formula for  $\hat{\Omega}$  as was used for the original data<sup>8</sup>.

Empirical rejection probabilities using the naive bootstrap critical values are reported in Figures 1 and 2 and are labelled NBoot. There are some compelling patterns in the figures. Contrary to theoretical results by Davison and Hall (1993) and Götze and Künsch (1996) suggesting otherwise, the naive bootstrap performs quite well; it dominates the standard normal approximation and has rejection probabilities that are very similar to the fixed- $b$  asymptotics. This is true even for the *iid* ( $l = 1$ ) bootstrap when the data is not *iid*. In addition, the fact that the naive bootstrap works well when using the Bartlett kernel contrasts with conventional wisdom that claims bootstrapping Bartlett kernel tests is not more accurate than the standard normal approximation.

It is quite striking how closely the naive bootstrap with  $l = 1$  (the *iid* bootstrap) follows the fixed- $b$  approximation regardless of the error structure. Recent work by Goncalves and Vogelsang (2004) has shown that this pattern is systematic. Goncalves and Vogelsang (2004) prove that the naive bootstrap has the same first-order asymptotic distribution as the  $t$ -statistic under fixed- $b$  asymptotics including the case when  $l = 1$ . The simulations also suggest that the naive bootstrap could be more accurate than fixed- $b$  asymptotics with suitable choice of the block length. Focusing on the cases of AR(1) errors with  $l = 4$ , notice that the naive bootstrap often has rejection probabilities closer to 0.05 than fixed- $b$ . While the naive bootstrap may provide a refinement

over the fixed- $b$  asymptotics, such a refinement has not been formally established. Higher order properties of the naive bootstrap under fixed- $b$  asymptotics is a very interesting, but potentially difficult, area where more work is needed.

## 5.2 Local Asymptotic Power

Whereas the existing HAC literature has almost exclusively focused on the MSE of the HAC estimator to guide the choice of kernel and bandwidth, a more relevant metric is to focus on the power of the resulting tests. In this section we compare power of HAC robust  $t$ -tests using a local asymptotic power analysis within the fixed- $b$  asymptotic framework. Our analysis permits comparison of power across bandwidths and across kernels. Such a comparison is not possible using the traditional first-order small- $b$  asymptotics because local asymptotic power is the same for all bandwidths and kernels.

For clarity, we restrict attention to linear regression models. Given the results in Theorem 1, the derivations in this section are very simple extensions of results given by Kiefer and Vogelsang (2002b). Therefore, details are kept to a minimum. Consider the regression model

$$y_t = x_t' \theta_0 + u_t \tag{9}$$

with  $\theta_0$  and  $x_t$   $p \times 1$  vectors. In terms of the general model we have  $f(v_t, \theta_0) = x_t'(y_t - x_t' \theta_0)$ . Without loss of generality, we focus on  $\theta_{i0}$ , one element of  $\theta$ , and consider null and alternative hypotheses

$$\begin{aligned} H_0 &: \theta_{i0} \leq 0 \\ H_1 &: \theta_{i0} = cT^{-1/2} \end{aligned}$$

where  $c > 0$  is a constant. If the regression model satisfies Assumptions 1 through 4, then we can use the results of Theorem 1 and results from Kiefer and Vogelsang (2002b) to easily establish that under the local alternative,  $H_1$ , as  $T \rightarrow \infty$ ,

$$t \Rightarrow \frac{\delta + B_1(1)}{\sqrt{Q_1(b)}}, \tag{10}$$

where  $\delta = c/\sqrt{V^{ii}}$ .

Asymptotic power curves can be computed for given bandwidths and kernels by simulating the asymptotic distribution of  $t$  based on (10) for a range of values for  $\delta$  and computing rejection probabilities with respect to the relevant null critical value. Using the same simulation methods as for the asymptotic critical values, local asymptotic power was computed for  $\delta = 0, 0.2, 0.4, \dots, 4.8, 5.0$  using 5% asymptotic null critical values.

The power results are reported in two ways. Figures 3-10 plot power across the kernels for a given value of  $b$ . Figures 11-15 plot power across values of  $b$  for a given kernel. Figures 3-10 show that for small bandwidths, power is essentially the same across kernels. As  $b$  increases, it becomes clear that the Bartlett kernel has the highest power while the QS and Daniell kernels have the lowest power. If power is the criterion used to choose a test, then the Bartlett kernel is the best choice within this set of five kernels. If we compare the Bartlett and QS kernels, we see that the power ranking of these kernels is the reverse of their ranking based on accuracy of the asymptotic approximation under the null.

Figures 11-15 show how the choice of bandwidth affects power. Regardless of the kernel, power is highest for small bandwidths and lowest for large bandwidths and power is decreasing in  $b$ . These figures also show that power of the Bartlett kernel is least sensitive to  $b$  whereas power of the QS and Daniell kernels is the most sensitive to  $b$ . Again, power rankings of  $b$  are the opposite of rankings of  $b$  based on accuracy of the asymptotic approximation under the null.

### 5.3 Size-Power Trade-off

The finite sample simulations and local asymptotic power calculations clearly indicate a trade-off between size distortions and power with regard to choice of kernel and bandwidth when using fixed- $b$  critical values. Smaller bandwidths lead to tests with higher power but at the cost of greater size distortions whereas larger bandwidths lead to tests with less size distortions but lower power. Similar trade-offs occur across kernels<sup>9</sup>. Balancing the size-power trade-off in some systematic way could lead to the development of useful bandwidth rules that deliver tests with desirable properties<sup>10</sup>. Indeed, for the exponentiated Bartlett kernel, Phillips et al. (2004) propose a data dependent rule for the choice of exponent that minimizes a loss function that is a weighted sum of type I and type II errors. They develop higher-order expansions within the fixed-exponent asymptotic framework and use these expansions to analytically quantify their loss function. The authors conjecture that their results can be extended to traditional kernels via fixed- $b$  asymptotics. Thus it seems promising that fixed- $b$  asymptotics can be used to develop data dependent bandwidth rules that balance size distortion and power.

## 6 Conclusions

We have provided a new approach to the asymptotic theory of HAC robust testing. We consider tests based on the popular nonparametric kernel estimates of the standard errors. We are not proposing new tests but rather we propose a new asymptotic theory for these well-known tests. Our results are general enough to apply to stationary models estimated by GMM. In our approach,  $b$ , the ratio of bandwidth to sample size is held constant when deriving the asymptotic behavior of the relevant covariance matrix estimator (i.e. zero frequency spectral density estimator). Thus

we label our asymptotic framework “fixed- $b$ ” asymptotics. In standard asymptotics,  $b$  is sent to zero and can be viewed as a “small- $b$ ” asymptotic framework. Fixed- $b$  asymptotics improves upon two well-known problems with the standard approach. First, as has been well documented in the literature, the standard asymptotic approximation of the sampling behavior of tests is often poor. Second, the kernel and bandwidth choice do not appear in the approximate distribution, leaving the standard theory silent on the choice of kernel and bandwidth with respect to properties of the tests. Our theory leads to approximate distributions that explicitly depend on the kernel and bandwidth. The new approximation performs much better and gives insight into the choice of kernel and bandwidth with respect to test behavior. Once the higher-order results of Phillips et al. (2004) have been extended to the fixed- $b$  framework, data dependent bandwidth rules that balance size distortion and power can be developed. In addition, fixed- $b$  asymptotics should be useful in explaining the performance of the naive bootstrap when applied to HAC robust tests.

The new approximations should be used for HAC robust test statistics for any choice of kernel and bandwidth. The fixed- $b$  approximation is an unambiguous improvement over the standard normal approximation in most cases considered. We show that size distortions are reduced when large bandwidths are used, but so is asymptotic power. Generally there is a trade-off in bandwidth and kernel choice between size (the accuracy of the approximation) and power. Among a group of popular kernels, the QS kernel leads to the least size distortion, while the Bartlett kernel leads to tests with highest power (and generally acceptable size distortion when large bandwidths are used).

## 7 Appendix: Proofs

We first define some relevant functions and derive preliminary results before proving the lemma and theorems. Define the functions

$$\begin{aligned} k^*(x) &= k\left(\frac{x}{b}\right), \\ K_{ij} &= k\left(\frac{i-j}{bT}\right) = k^*\left(\frac{i-j}{T}\right), \\ \Delta^2 K_{ij} &= (K_{ij} - K_{i,j+1}) - (K_{i+1,j} - K_{i+1,j+1}), \\ D_T^*(r) &= T^2 \left[ \left( k^*\left(\frac{[rT]+1}{T}\right) - k^*\left(\frac{[rT]}{T}\right) \right) - \left( k^*\left(\frac{[rT]}{T}\right) - k^*\left(\frac{[rT]-1}{T}\right) \right) \right]. \end{aligned}$$

Notice that

$$T^2 \Delta^2 K_{ij} = -D_T^*\left(\frac{i-j}{T}\right).$$

Because  $k(r)$  is an even function around  $r = 0$ ,  $D_T^*(-r) = D_T^*(r)$ . If  $k^{*''}(r)$  exists then  $\lim_{T \rightarrow \infty} D_T^*(r) = k^{*''}(r)$  by the definition of the second derivative. If  $k^{*''}(r)$  is continuous, then  $D_T^*(r)$  converges to  $k^{*''}(r)$  uniformly in  $r$ . Define the stochastic process

$$X_T(r) = G_T(\hat{\theta}_T)' W_T T^{1/2} g_{[rT]}(\theta_0)$$

for  $\frac{1}{T} \leq r \leq 1$  and  $X_T(r) = 0$  for  $0 \leq r < \frac{1}{T}$  (i.e. set  $g_0(\theta_0) = 0$ ). It directly follows from Assumptions 2, 3 and 4 that

$$X_T(r) \Rightarrow G'_0 W_\infty \Lambda B_q(r) \equiv \Lambda^* B_p(r). \quad (11)$$

**Proof of Lemma 1:** Setting  $t = T$ , multiplying both sides of (4) by  $G_T(\hat{\theta}_T)'W_T$ , and using the first-order condition  $G_T(\hat{\theta}_T)'W_T g_T(\hat{\theta}_T) = \mathbf{0}$  gives

$$\mathbf{0} = G_T(\hat{\theta}_T)'W_T g_T(\theta_0) + G_T(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T, \theta_0, \lambda_T)(\hat{\theta}_T - \theta_0). \quad (12)$$

Solving (12) for  $(\hat{\theta}_T - \theta_0)$  and scaling by  $T^{1/2}$  gives

$$\begin{aligned} T^{1/2}(\hat{\theta}_T - \theta_0) &= -[G'_T(\hat{\theta}_T)W_T G_T(\hat{\theta}_T, \theta_0, \lambda_T)]^{-1} G'_T(\hat{\theta}_T)W_T T^{1/2} g_T(\theta_0) \\ &= -[G_T(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T, \theta_0, \lambda_T)]^{-1} X_T(1). \end{aligned} \quad (13)$$

Because  $p \lim G_T(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T, \theta_0, \lambda_T) = G'_0 W_\infty G_0$  by Assumptions 3 and 4, it follows from (11) that

$$T^{1/2}(\hat{\theta}_T - \theta_0) \Rightarrow - (G'_0 W_\infty G_0)^{-1} \Lambda^* B_p(1).$$

We now prove Theorem 2. The proof of Theorem 1 follows using similar arguments and is omitted.

**Proof of Theorem 2:** Define the random process  $\tilde{X}_T(r) = G_T(\hat{\theta}_T)'W_T T^{1/2} g_{[rT]}(\hat{\theta}_T)$  for  $\frac{1}{T} \leq r \leq 1$  and  $\tilde{X}_T(r) = 0$  for  $0 \leq r < \frac{1}{T}$ . Plugging in for  $g_{[rT]}(\hat{\theta}_T)$  using (4) gives

$$\begin{aligned} \tilde{X}_T(r) &= X_T(r) + G_T(\hat{\theta}_T)'W_T G_{[rT]}(\hat{\theta}_T, \theta_0, \lambda_T) T^{1/2}(\hat{\theta}_T - \theta_0) \\ &= X_T(r) - G_T(\hat{\theta}_T)'W_T G_{[rT]}(\hat{\theta}_T, \theta_0, \lambda_T) \left[ G_T(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T, \theta_0, \lambda_T) \right]^{-1} X_T(1), \end{aligned}$$

using (13). It directly follows from Assumptions 3 and 4 and (11) that

$$\begin{aligned} \tilde{X}_T(r) &\Rightarrow \Lambda^* B_p(r) - r G'_0 W_\infty G_0 (G'_0 W_\infty G_0)^{-1} \Lambda^* B_p(1) \\ &= \Lambda^* (B_p(r) - r B_p(1)) \equiv \Lambda^* \tilde{B}_p(r). \end{aligned} \quad (14)$$

Straightforward algebra gives

$$\hat{\Omega} = \sum_{j=-(T-1)}^{T-1} k \left( \frac{j}{bT} \right) \hat{\Gamma}_j = T^{-1} \sum_{i=1}^T \sum_{j=1}^T f(v_i, \hat{\theta}_T) K_{ij} f(v_j, \hat{\theta}_T)'$$

Using algebraic arguments similar to those used by Kiefer and Vogelsang (2002b), it is straightforward to show that

$$\begin{aligned}\widehat{\Omega} &= T^{-1} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} \Delta^2 K_{ij} T g_i(\widehat{\theta}_T) T g_j(\widehat{\theta}_T)' \\ &\quad + g_T(\widehat{\theta}_T) \sum_{i=1}^{T-1} (K_{Ti} - K_{T,i+1}) T g_i(\widehat{\theta}_T)' + \left( \sum_{j=1}^T f(v_j, \widehat{\theta}_T) K_{jT} \right) g_T(\widehat{\theta}_T)'.\end{aligned}\quad (15)$$

To simplify notation, define  $C_{ij} = G_T(\widehat{\theta}_T)' W_T T^{1/2} g_i(\widehat{\theta}_T) T^{1/2} g_j(\widehat{\theta}_T)' W_T G_T(\widehat{\theta}_T)$ . Using (15) and the facts that

$$\begin{aligned}G_T(\widehat{\theta}_T)' W_T T g_T(\widehat{\theta}_T) &= 0, \\ T g_T(\widehat{\theta}_T)' W_T G_T(\widehat{\theta}_T) &= 0,\end{aligned}$$

which follow from (3), it directly follows that

$$\begin{aligned}G_T(\widehat{\theta}_T)' W_T \widehat{\Omega} W_T G_T(\widehat{\theta}_T) &= T^{-1} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} \Delta^2 K_{ij} G_T(\widehat{\theta}_T)' W_T T g_i(\widehat{\theta}_T) T g_j(\widehat{\theta}_T)' W_T G_T(\widehat{\theta}_T) \\ &= T^{-2} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} T^2 \Delta^2 K_{ij} C_{ij} = T^{-2} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} -D_T^* \left( \frac{i-j}{T} \right) C_{ij}.\end{aligned}\quad (16)$$

The rest of the proof is divided into three cases.

Case 1:  $k(x)$  is twice continuously differentiable. Using (16) it follows that

$$\begin{aligned}&G_T(\widehat{\theta}_T)' W_T \widehat{\Omega} W_T G_T(\widehat{\theta}_T) \\ &= - \int_0^1 \int_0^1 D_T^*(r-s) G_T(\widehat{\theta}_T)' W_T T^{1/2} g_{[rT]}(\widehat{\theta}_T) T^{1/2} g_{[sT]}(\widehat{\theta}_T)' W_T G_T(\widehat{\theta}_T) dr ds \\ &= - \int_0^1 \int_0^1 D_T^*(r-s) \widetilde{X}_T(r) \widetilde{X}_T(s)' dr ds \\ &\Rightarrow -\Lambda^* \int_0^1 \int_0^1 k^{*''}(r-s) \widetilde{B}_p(r) \widetilde{B}_p(s)' dr ds \Lambda^{*'},\end{aligned}$$

using the continuous mapping theorem. The final expression is obtained using  $k^{*''}(x) = \frac{1}{b^2} k''(\frac{x}{b})$ .

Case 2:  $k(x)$  is continuous,  $k(x) = 0$  for  $|x| \geq 1$ , and  $k(x)$  is twice continuously differentiable everywhere except for  $|x| = 1$ . Let  $1(\bullet)$  denote the indicator function. Noting that  $\Delta^2 K_{ij} = 0$  for  $|i-j| > [bT] + 1$ ,  $\Delta^2 K_{ij} = -k\left(\frac{[bT]}{bT}\right) = -k^*\left(\frac{[bT]}{T}\right)$  for  $|i-j| = [bT] + 1$ , and  $\Delta^2 K_{ij} =$

$\left[ k \left( \frac{[bT]}{bT} \right) - k \left( \frac{[bT]}{bT} - \frac{1}{bT} \right) \right] + k \left( \frac{[bT]}{bT} \right) = \left[ k^* \left( \frac{[bT]}{T} \right) - k^* \left( \frac{[bT]}{T} - \frac{1}{T} \right) \right] + k^* \left( \frac{[bT]}{T} \right)$  for  $|i - j| = [bT]$ , break up the double sum in the second line of (16) to obtain

$$\begin{aligned}
& G_T(\widehat{\theta}_T)' W_T \widehat{\Omega} W_T G_T(\widehat{\theta}_T) \\
&= T^{-2} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} 1(|i - j| < [bT]) T^2 \Delta^2 K_{ij} C_{ij} \\
&+ T \left[ k^* \left( \frac{[bT]}{T} \right) - k^* \left( \frac{[bT]}{T} - \frac{1}{T} \right) \right] T^{-1} \sum_{i=1}^{T-[bT]-1} (C_{i+[bT],i} + C_{i,i+[bT]}) \\
&+ k^* \left( \frac{[bT]}{T} \right) \sum_{i=1}^{T-[bT]-1} (C_{i+[bT],i} + C_{i,i+[bT]}) - k^* \left( \frac{[bT]}{T} \right) \sum_{i=1}^{T-[bT]-2} (C_{i+[bT]+1,i} + C_{i,i+[bT]+1}). \quad (17)
\end{aligned}$$

Using the fact that  $C_{T,i} = C_{i,T} = 0$ , it follows that

$$\sum_{i=1}^{T-[bT]-2} (C_{i+[bT]+1,i} + C_{i,i+[bT]+1}) = \sum_{i=1}^{T-[bT]-1} (C_{i+[bT]+1,i} + C_{i,i+[bT]+1}).$$

By definition

$$\begin{aligned}
g_{i+[bT]+1}(\widehat{\theta}_T) &= g_{i+[bT]}(\widehat{\theta}_T) + T^{-1} f(v_{i+[bT]+1}, \widehat{\theta}_T), \\
C_{i+[bT]+1,i} &= G_T(\widehat{\theta}_T)' W_T T^{1/2} g_{i+[bT]+1}(\widehat{\theta}_T) T^{1/2} g_i(\widehat{\theta}_T)' W_T G_T(\widehat{\theta}_T)
\end{aligned}$$

in which case we can write

$$C_{i+[bT]+1,i} = C_{i+[bT],i} + G_T(\widehat{\theta}_T)' W_T T^{1/2} f(v_{i+[bT]+1}, \widehat{\theta}_T) T^{1/2} g_i(\widehat{\theta}_T)' W_T G_T(\widehat{\theta}_T).$$

Therefore it follows that

$$\begin{aligned}
& -k^* \left( \frac{[bT]}{T} \right) \sum_{i=1}^{T-[bT]-2} (C_{i+[bT]+1,i} + C_{i,i+[bT]+1}) = -k^* \left( \frac{[bT]}{T} \right) \sum_{i=1}^{T-[bT]-1} (C_{i+[bT]+1,i} + C_{i,i+[bT]+1}) \\
&= -k^* \left( \frac{[bT]}{T} \right) \sum_{i=1}^{T-[bT]-1} (C_{i+[bT],i} + C_{i,i+[bT]}) \\
&-k^* \left( \frac{[bT]}{T} \right) T^{-1} \sum_{i=1}^{T-[bT]-1} G_T(\widehat{\theta}_T)' W_T f(v_{i+[bT]+1}, \widehat{\theta}_T) T g_i(\widehat{\theta}_T)' W_T G_T(\widehat{\theta}_T) \\
&-k^* \left( \frac{[bT]}{T} \right) T^{-1} \sum_{i=1}^{T-[bT]-1} G_T(\widehat{\theta}_T)' W_T T g_i(\widehat{\theta}_T) f(v_{i+[bT]+1}, \widehat{\theta}_T)' W_T G_T(\widehat{\theta}_T) \quad (18)
\end{aligned}$$

$$= -k^* \left( \frac{[bT]}{T} \right) \sum_{i=1}^{T-[bT]-1} (C_{i+[bT],i} + C_{i,i+[bT]}) + o_p(1) \quad (19)$$

where the second two terms of (18) are  $o_p(1)$  because

$$T^{-1} \sum_{i=1}^{T-[bT]-1} G_T(\hat{\theta}_T)' W_T f(v_{i+[bT]+1}, \hat{\theta}_T) T g_i(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T) = O_p(1),$$

$$\lim_{T \rightarrow \infty} k^* \left( \frac{[bT]}{T} \right) = k^*(b) = k(1) = 0.$$

Note that the  $o_p(1)$  terms become identically zero when  $bT$  is an integer. Using (19) the right-hand side of (17) simplifies to

$$\begin{aligned} & T^{-2} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} 1(|i-j| < [bT]) T^2 \Delta^2 K_{ij} C_{ij} \\ & + T \left[ k^* \left( \frac{[bT]}{T} \right) - k^* \left( \frac{[bT]}{T} - \frac{1}{T} \right) \right] T^{-1} \sum_{i=1}^{T-[bT]-1} (C_{i+[bT],i} + C_{i,i+[bT]}) + o_p(1). \end{aligned} \quad (20)$$

Using this expression we can write

$$\begin{aligned} & G_T(\hat{\theta}_T)' W_T \hat{\Omega} W_T G_T(\hat{\theta}_T) = \\ & - \iint_{|r-s|<b} D_T^*(r-s) \tilde{X}_T(r) \tilde{X}_T(s)' dr ds \\ & + T \left[ k^* \left( \frac{[bT]}{T} \right) - k^* \left( \frac{[bT]}{T} - \frac{1}{T} \right) \right] \int_0^{1-b} \left( \tilde{X}_T(r+b) \tilde{X}_T(r)' + \tilde{X}_T(r) \tilde{X}_T(r+b)' \right) dr + o_p(1). \end{aligned}$$

Let  $k'_-(b)$  denote the first derivative of  $k^*(x)$  from the left at  $x = b$ . By definition

$$\frac{k'_-(1)}{b} = k'_-(b) = \lim_{T \rightarrow \infty} T \left[ k^*(b) - k^* \left( b - \frac{1}{T} \right) \right] = \lim_{T \rightarrow \infty} T \left[ k^* \left( \frac{[bT]}{T} \right) - k^* \left( \frac{[bT]}{T} - \frac{1}{T} \right) \right].$$

Therefore, by the continuous mapping theorem

$$\begin{aligned} & G_T(\hat{\theta}_T)' W_T \hat{\Omega} W_T G_T(\hat{\theta}_T) dr \Rightarrow \\ & \Lambda^* \left[ - \iint_{|r-s|<b} k^{*''}(r-s) \tilde{B}_p(r) \tilde{B}_p(s)' dr ds + k'_-(b) \int_0^{1-b} \left( \tilde{B}_p(r+b) \tilde{B}_p(r)' + \tilde{B}_p(r) \tilde{B}_p(r+b)' \right) dr \right] \Lambda^{*'} \end{aligned}$$

Case 3:  $k(x)$  is the Bartlett kernel. It is easy to calculate that  $\Delta^2 K_{ij} = \frac{2}{bT}$  for  $|i-j| = 0$ ,  $\Delta^2 K_{ij} = -\frac{1}{bT} + 1 - \frac{[bT]}{bT}$  for  $|i-j| = [bT]$ ,  $\Delta^2 K_{ij} = -\left(1 - \frac{[bT]}{bT}\right)$  for  $|i-j| = [bT] + 1$  and  $\Delta^2 K_{ij} = 0$  otherwise. Using similar algebraic calculations as was done in Case 2, it is straightforward to show that

$$G_T(\hat{\theta}_T)' W_T \hat{\Omega} W_T G_T(\hat{\theta}_T) = \frac{2}{bT} \sum_{i=1}^{T-1} C_{ii} - \frac{1}{bT} \sum_{i=1}^{T-[bT]-1} (C_{i+[bT],i} + C_{i,i+[bT]}) + o_p(1)$$

$$\begin{aligned}
&= \frac{2}{b} \int_0^1 \tilde{X}_T(r) \tilde{X}_T(r)' dr - \frac{1}{b} \int_0^{1-b} \left( \tilde{X}_T(r+b) \tilde{X}_T(r)' + \tilde{X}_T(r) \tilde{X}_T(r+b)' \right) dr + o_p(1) \\
&\Rightarrow \Lambda^* \left[ \frac{2}{b} \int_0^1 \tilde{B}_p(r) \tilde{B}_p(r)' dr - \frac{1}{b} \int_0^{1-b} \left( \tilde{B}_p(r+b) \tilde{B}_p(r)' + \tilde{B}_p(r) \tilde{B}_p(r+b)' \right) dr \right] \Lambda^{*'}
\end{aligned}$$

**Proof of Theorem 3:** We only give the proof for  $F$  as the proof for  $t$  follows using similar arguments. Applying the delta method to the result in Lemma 1 and using the fact that  $B_q(1)$  is a vector of independent standard normal random variables gives

$$\begin{aligned}
T^{1/2}r(\hat{\theta}_T) &\Rightarrow -R(\theta_0) (G_0'W_\infty G_0)^{-1} G_0'W_\infty \Lambda B_q(1) \\
&\equiv -R(\theta_0) (G_0'W_\infty G_0)^{-1} \Lambda^* B_p(1) \\
&\equiv \Lambda^{**} B_m(1),
\end{aligned} \tag{21}$$

where  $\Lambda^{**}$  is the matrix square root of  $R(\theta_0) (G_0'W_\infty G_0)^{-1} \Lambda^* \Lambda^{*'} (G_0'W_\infty G_0)^{-1} R(\theta_0)'$ . Using the results in Theorem 2, it directly follows that

$$\begin{aligned}
R(\hat{\theta}_T) \hat{V}_{M=bT} R(\hat{\theta}_T)' &= R(\hat{\theta}_T) \left[ G_T(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T, \theta_0, \lambda_T) \right]^{-1} G_T(\hat{\theta}_T)' W_T \hat{\Omega}_{M=T} W_T G_T(\hat{\theta}_T) \\
&\quad \times \left[ G_T(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T, \theta_0, \lambda_T) \right]^{-1} R(\hat{\theta}_T)' \\
&\Rightarrow R(\theta_0) (G_0'W_\infty G_0)^{-1} \Lambda^* Q_p(b) \Lambda^{*'} (G_0'W_\infty G_0)^{-1} R(\theta_0)' \\
&\equiv \Lambda^{**} Q_m(b) \Lambda^{*'},
\end{aligned} \tag{22}$$

where we use the fact that

$$\begin{aligned}
R(\theta_0) (G_0'W_\infty G_0)^{-1} \Lambda^* \tilde{B}_p(r) &= R(\theta_0) (G_0'W_\infty G_0)^{-1} \Lambda^* (B_p(r) - rB_p(1)) \\
&\equiv \Lambda^{**} (B_m(r) - rB_m(1)) \\
&= \Lambda^{**} \tilde{B}_m(r).
\end{aligned}$$

Using (21) and (22) it directly follows that

$$\begin{aligned}
F &= Tr(\hat{\theta}_T)' \left( R(\hat{\theta}_T) \hat{V}_{M=bT} R(\hat{\theta}_T)' \right)^{-1} r(\hat{\theta}_T)/m \\
&= T^{1/2}r(\hat{\theta}_T)' \left( R(\hat{\theta}_T) \hat{V}_{M=T} R(\hat{\theta}_T)' \right)^{-1} T^{1/2}r(\hat{\theta}_T)/m \\
&\Rightarrow (\Lambda^{**} B_m(1))' (\Lambda^{**} Q_m(b) \Lambda^{*'})^{-1} (\Lambda^{**} B_m(1))/m \\
&\equiv B_m(1)' Q_m(b)^{-1} B_m(1)/m,
\end{aligned}$$

which completes the proof.

## Notes

<sup>1</sup>An alternative to the nonparametric approach has been advocated by den Haan and Levin (1997,1998). Following Berk (1974) and others in the time-series statistics literature, they propose estimating the zero frequency spectral density parametrically using vector autoregression (VAR) models. They show that this parametric approach can achieve essentially the same generality as the nonparametric approach if the VAR lag length increases with the sample size at a suitable rate.

<sup>2</sup>Additional discussion of this point is given by Cushing and McGarvey (1999, p. 80).

<sup>3</sup>An alternative approach to the fixed- $b$  asymptotics is to consider higher order asymptotic approximations in the small- $b$  framework. For example, one could take the Edgeworth expansions from the recent work by Velasco and Robinson (2001) and obtain a second order asymptotic approximation for HAC robust tests that depends on the bandwidth and kernels. While this is a potentially fruitful approach, it is complicated by the need to estimate the second derivative of the spectral density.

<sup>4</sup>This common misperception is an unfortunate result of a folk-lore in the econometrics literature that states that “as  $M$  increases, bias in  $\widehat{\Omega}$  decreases but variance increases.” This folk-lore is quite misleading although the source is easy to pinpoint. A careful reader of Priestley (1981) will repeatedly see the phrase “as  $M$  increases, bias in  $\widehat{\Omega}$  decreases but variance increases”. This phrase is completely correct if, as in Priestley (1981), one is discussing the properties of spectral density estimators at non-zero frequencies or, in the case of known mean zero data, at the zero frequency. However, this phrase does not apply to zero frequency estimators computed using demeaned data as is the case in GMM models. This fact is well known and has been nicely illustrated by Figure 2 in Ng and Perron (1996) where plots of the exact bias and variance of  $\widehat{\Omega}$  are given for AR(1) processes.

<sup>5</sup>The results in this section are a fixed- $b$  asymptotic analysis of nonparametric zero-frequency spectral density estimators. Fixed- $b$  asymptotic results for spectral density estimators at non-zero frequencies have been obtained by Hashimzade and Vogelsang (2003).

<sup>6</sup>The standard notation for the commutation matrix is usually  $K_{ii}$  (see Magnus and Neudecker (1999, p.46)). We are using alternative notation because  $K_{ij}$  is used for a different matrix in our proofs.

<sup>7</sup>A recent paper by Müller (2004) provides a novel way of theoretically quantifying the size robustness properties of HAC robust tests as a function of the kernel and  $b$ . The robustness measure proposed by Müller (2004) closely matches the patterns seen in the finite sample simulations.

<sup>8</sup>We considered implementing the recentered and rescaled version of the bootstrap as proposed by Götze and Künsch (1996). Because Götze and Künsch (1996) recommend using the block length,  $l$ , equal to the HAC bandwidth,  $M$ , it is difficult to implement this bootstrap for the full range of  $M$  values given in Figures 1 and 2. The problem is that for large values of  $M$ ,  $l$  is so large that the resampling is based on a very small number of blocks and it breaks down. We leave a comparison of the naive bootstrap and the Götze and Künsch (1996) bootstrap to future research.

<sup>9</sup>Simulations reported by Phillips et al. (2003) and Phillips et al. (2005) found a similar trade-off between size distortion and power with respect to the exponent of exponentiated kernels.

<sup>10</sup>In the context of trend function inference analyzed using fixed- $b$  asymptotics, Bunzel and Vogelsang (2005) propose a data dependent bandwidth rule that maximizes integrated power. They focus only on power because the class of tests they consider are very robust in terms of size distortions. Their approach does not naturally apply to the models in this paper because their size correction technique does not easily apply to models with stochastic regressors.

## References

- Abadir, K. M. and Paruolo, P. (2002), Simple Robust Testing of Regression Hypotheses: a Comment, *Econometrica* **70**, 2097–2099.
- Andrews, D. W. K. (1991), Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica* **59**, 817–854.
- Andrews, D. W. K. and Monahan, J. C. (1992), An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator, *Econometrica* **60**, 953–966.
- Berk, K. N. (1974), Consistent Autoregressive Spectral Estimates, *The Annals of Statistics* **2**, 489–502.
- Bunzel, H. and Vogelsang, T. J. (2005), Powerful Trend Function Tests That are Robust to Strong Serial Correlation with an Application to the Prebisch-Singer Hypothesis, *Journal of Business and Economic Statistics*, *forthcoming*.
- Bunzel, H., Kiefer, N. M. and Vogelsang, T. J. (2001), Simple Robust Testing of Hypotheses in Non-linear Models, *Journal of American Statistical Association* **96**, 1088–1098.
- Cushing, M. J. and McGarvey, M. G. (1999), Covariance Matrix Estimation, in L. Matyas (ed.), *Generalized Method of Moments Estimation*, Cambridge University Press, New York.
- Davison, A. C. and Hall, P. (1993), On Studentizing and Blocking Methods for Implementing the Bootstrap with Dependent Data, *Australian Journal of Statistics* **35**, 215–224.
- de Jong, R. M. and Davidson, J. (2000), Consistency of Kernel Estimators of Heteroskedastic and Autocorrelated Covariance Matrices, *Econometrica* **68**, 407–424.
- den Haan, W. J. and Levin, A. (1997), A Practitioner’s Guide to Robust Covariance Matrix Estimation, in G. Maddala and C. Rao (eds), *Handbook of Statistics: Robust Inference*, Volume 15, Elsevier, New York, pp. 291–341.
- den Haan, W. J. and Levin, A. (1998), Vector Autoregressive Covariance Matrix Estimation, working paper, International Finance Division, FED Board of Governors.
- Gallant, A. (1987), *Nonlinear Statistical Models*, Wiley, New York.
- Gallant, A. and White, H. (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Basil Blackwell, New York.

- Goncalves, S. and Vogelsang, T. J. (2004), Block Bootstrap Puzzles in HAC Robust Testing: The Sophistication of the Naive Bootstrap, Working Paper, Department of Economics, Cornell University.
- Götze, F. and Künsch, H. R. (1996), Second-Order Correctness of the Blockwise Bootstrap for Stationary Observations, *Annals of Statistics* **24**, 1914–1933.
- Hall, P. and Horowitz, J. L. (1996), Bootstrap Critical Values for Tests Based on Generalized Method of Moments Estimators, *Econometrica* **64**, 891–916.
- Hansen, B. E. (1992), Consistent Covariance Matrix Estimation for Dependent Heterogenous Processes, *Econometrica* **60**, 967–972.
- Hansen, L. P. (1982), Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* **50**, 1029–1054.
- Hashimzade, N. and Vogelsang, T. J. (2003), A New Asymptotic Approximation for the Sampling Behavior of Spectral Density Estimators, Working Paper, Department of Economics, Cornell University.
- Hashimzade, N., Kiefer, N. M. and Vogelsang, T. J. (2003), Moments of HAC Robust Covariance Matrix Estimators Under Fixed- $b$  Asymptotics, Working Paper, Department of Economics, Cornell University.
- Heyde, C. (1997), *Quasi-Likelihood and its Application. A General Approach to Optimal Parameter Estimation*, Springer, New York.
- Inoue, A. and Shintani, M. (2004), Bootstrapping GMM Estimators for Time Series, *Journal of Econometrics*. forthcoming.
- Jansson, M. (2002), Consistent Covariance Estimation for Linear Processes, *Econometric Theory* **18**, 1449–1459.
- Jansson, M. (2004), The Error Rejection Probability of Simple Autocorrelation Robust Tests, *Econometrica* **72**, 937–946.
- Kiefer, N. M. and Vogelsang, T. J. (2002)a, Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel Without Truncation, *Econometrica* **70**, 2093–2095.
- Kiefer, N. M. and Vogelsang, T. J. (2002)b, Heteroskedasticity-Autocorrelation Robust Testing Using Bandwidth Equal to Sample Size, *Econometric Theory* **18**, 1350–1366.

- Kiefer, N. M. and Vogelsang, T. J. (2005), A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests, *Econometric Theory*, *forthcoming*.
- Kiefer, N. M., Vogelsang, T. J. and Bunzel, H. (2000), Simple Robust Testing of Regression Hypotheses, *Econometrica* **68**, 695–714.
- Magnus, J. R. and Neudecker, H. (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, New York.
- Müller, U. K. (2004), A Theory of Robust Long-Run Variance Estimation, mimeo, Department of Economics, Princeton University.
- Neave, H. R. (1970), An Improved Formula for the Asymptotic Variance of Spectrum Estimates, *Annals of Mathematical Statistics* **41**, 70–77.
- Newey, W. K. and McFadden, D. L. (1994), Large Sample Estimation and Hypothesis Testing, in R. Engle and D. L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, Elsevier Science Publishers, Amsterdam, The Netherlands, pp. 2113–2247.
- Newey, W. K. and West, K. D. (1987), A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* **55**, 703–708.
- Newey, W. K. and West, K. D. (1994), Automatic Lag Selection in Covariance Estimation, *Review of Economic Studies* **61**, 631–654.
- Ng, S. and Perron, P. (1996), The Exact Error in Estimating the Spectral Density at the Origin, *Journal of Time Series Analysis* **17**, 379–408.
- Phillips, P. C. B. and Durlauf, S. N. (1986), Multiple Regression with Integrated Processes, *Review of Economic Studies* **53**, 473–496.
- Phillips, P. C. B., Sun, Y. and Jin, S. (2003), Consistent HAC Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation, Working Paper, Department of Economics, Yale University.
- Phillips, P. C. B., Sun, Y. and Jin, S. (2004), Improved HAR Inference Using Power Kernels without Truncation, Working Paper, Department of Economics, Yale University.
- Phillips, P. C. B., Sun, Y. and Jin, S. (2005), Spectral Density Estimation and Robust Hypothesis Testing using Steep Origin Kernels without Truncation, *International Economic Review*, *forthcoming*.
- Priestley, M. B. (1981), *Spectral Analysis and Time Series*, Vol. 1, Academic Press, New York.

- Ravikumar, B., Ray, S. and Savin, N. E. (2004), Robust Wald Tests and the Curse of Dimensionality, Working paper, Department of Economics, University of Iowa.
- Robinson, P. (1998), Inference-Without Smoothing in the Presence of Nonparametric Autocorrelation, *Econometrica* **66**, 1163–1182.
- Simonoff, J. (1993), The Relative Importance of Bias and Variability in the Estimation of the Variance of a Statistic, *The Statistician* **42**, 3–7.
- Velasco, C. and Robinson, P. M. (2001), Edgeworth Expansions for Spectral Density Estimates and Studentized Sample Mean, *Econometric Theory* **17**, 497–539.
- Vogelsang, T. J. (2003), Testing in GMM Models Without Truncation, *Advances in Econometrics, Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later*, Vol. 17, pp. 199–233.
- White, H. (1984), *Asymptotic Theory for Econometricians*, Academic Press, New York.

**Table I: Fixed- $b$  Asymptotic Critical Value Function Coefficients for  $t$ .**

$$cv(b) = a_0 + a_1b + a_2b^2 + a_3b^3$$

Bartlett	$a_0$	$a_1$	$a_2$	$a_3$	$R^2$
90%	1.2816	1.3040	0.5135	-0.3386	0.9995
95%	1.6449	2.1859	0.3142	-0.3427	0.9991
97.5%	1.9600	2.9694	0.4160	-0.5324	0.9980
99%	2.3263	4.1618	0.5368	-0.9060	0.9957
Parzen	$a_0$	$a_1$	$a_2$	$a_3$	$R^2$
90%	1.2816	0.9729	0.5514	0.0011	0.9993
95%	1.6449	1.5184	1.0821	-0.0660	0.9993
97.5%	1.9600	2.0470	1.7498	-0.1076	0.9993
99%	2.3263	2.5794	3.9580	-0.7012	0.9984
QS	$a_0$	$a_1$	$a_2$	$a_3$	$R^2$
90%	1.2816	1.6269	2.6366	-0.4329	0.9996
95%	1.6449	2.7098	4.5885	-0.6984	0.9992
97.5%	1.9600	3.0002	10.5805	-3.3454	0.9984
99%	2.3263	5.4054	14.1281	-2.3440	0.9969
Daniell	$a_0$	$a_1$	$a_2$	$a_3$	$R^2$
90%	1.2816	1.4719	2.1942	-0.1981	0.9991
95%	1.6449	2.4986	3.9948	-0.4587	0.9990
97.5%	1.9600	2.8531	9.6484	-3.0756	0.9984
99%	2.3263	5.0506	14.1258	-3.2775	0.9952
Tuk-Han	$a_0$	$a_1$	$a_2$	$a_3$	$R^2$
90%	1.2816	1.1147	1.9782	-0.5142	0.9940
95%	1.6449	1.5479	4.4153	-1.4993	0.9957
97.5%	1.9600	1.6568	8.2454	-2.6136	0.9892
99%	2.3263	1.1261	18.3270	-7.1177	0.9825
Bohman	$a_0$	$a_1$	$a_2$	$a_3$	$R^2$
90%	1.2816	1.0216	0.7906	-0.1121	0.9993
95%	1.6449	1.5927	1.5151	-0.2925	0.9994
97.5%	1.9600	2.2432	2.0441	-0.1358	0.9989
99%	2.3263	2.6213	5.4876	-1.6575	0.9982

Notes: Given the kernel and a percentage point, for a given value of  $b = \frac{M}{T}$  the critical value for  $t$  is given by the polynomial  $cv(b) = a_0 + a_1b + a_2b^2 + a_3b^3$ . The  $R^2$  indicates the fit of the polynomial through the simulated asymptotic critical values.

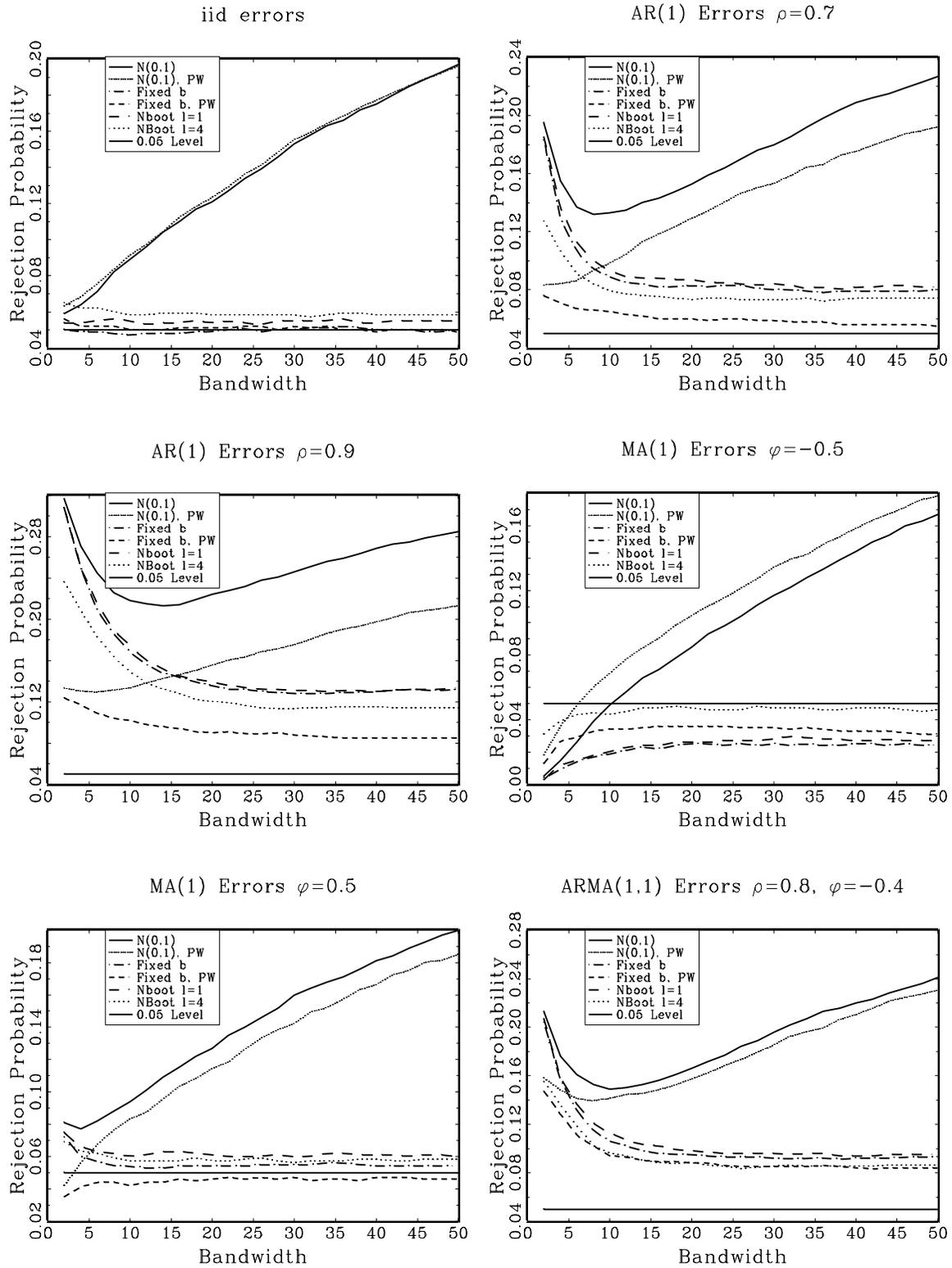


Figure 1: Empirical Null Rejection Probabilities  
Bartlett Kernel, T=50, 5% Nominal Level

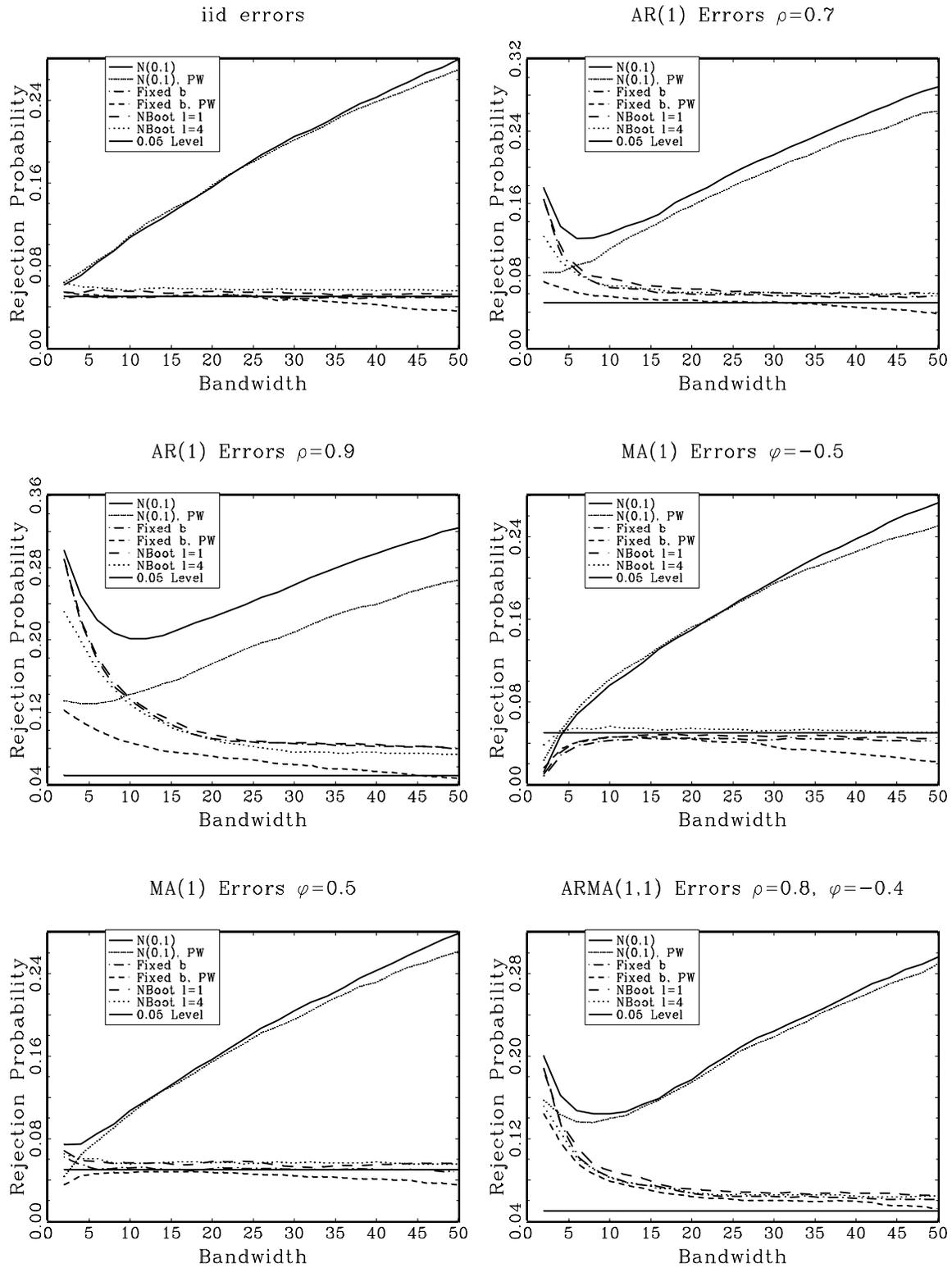


Figure 2: Empirical Null Rejection Probabilities  
 QS Kernel, T=50, 5% Nominal Level

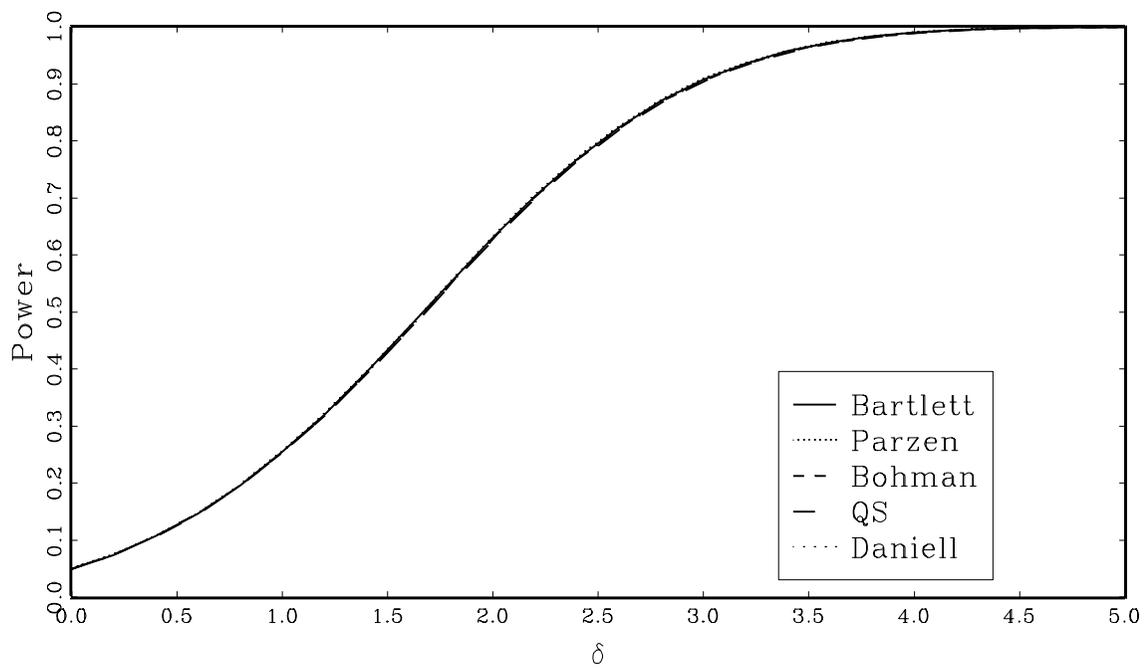


Figure 3: Local Asymptotic Power of  $t$ ,  $b = 0.02$

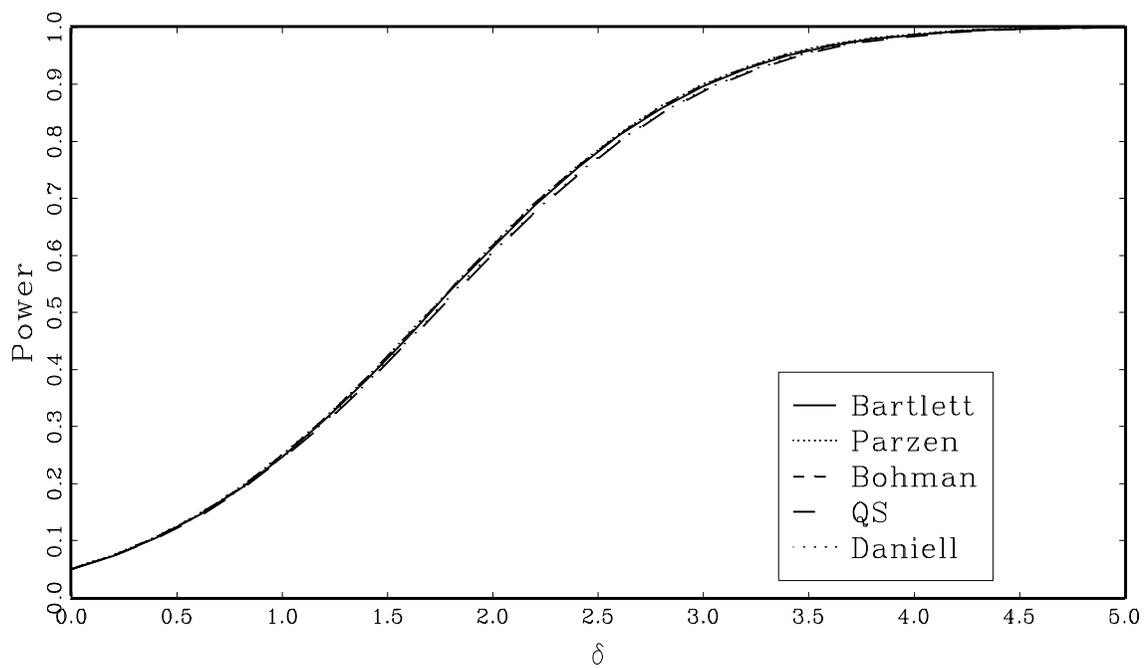


Figure 4: Local Asymptotic Power of  $t$ ,  $b = 0.06$

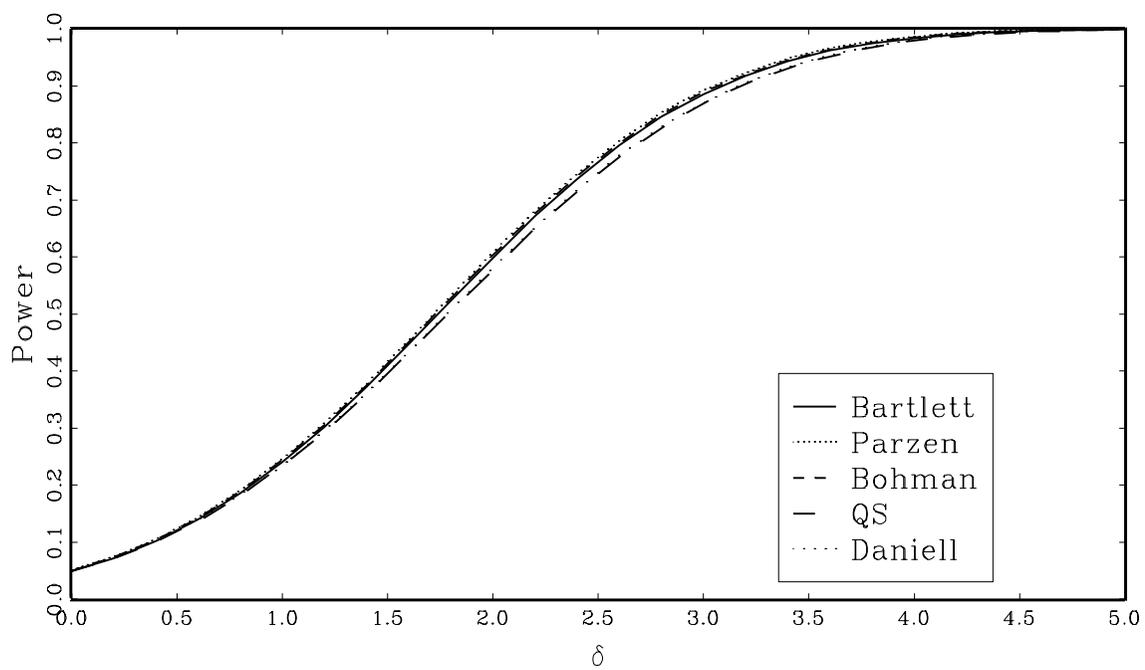


Figure 5: Local Asymptotic Power of t,  $b = 0.1$

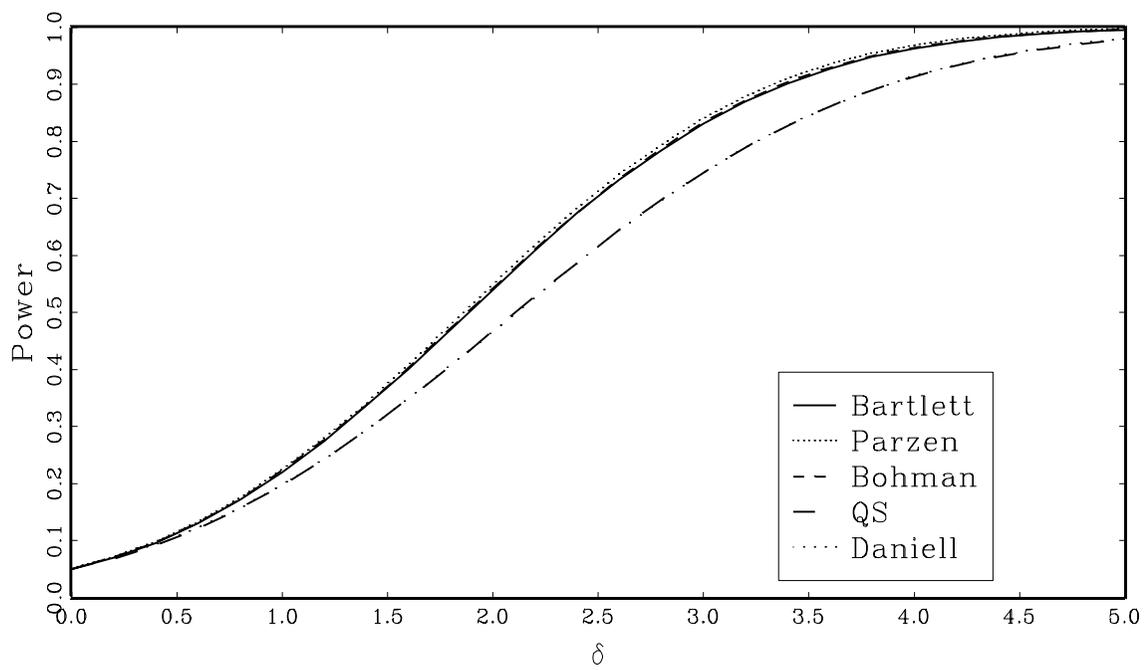


Figure 6: Local Asymptotic Power of t,  $b = 0.3$

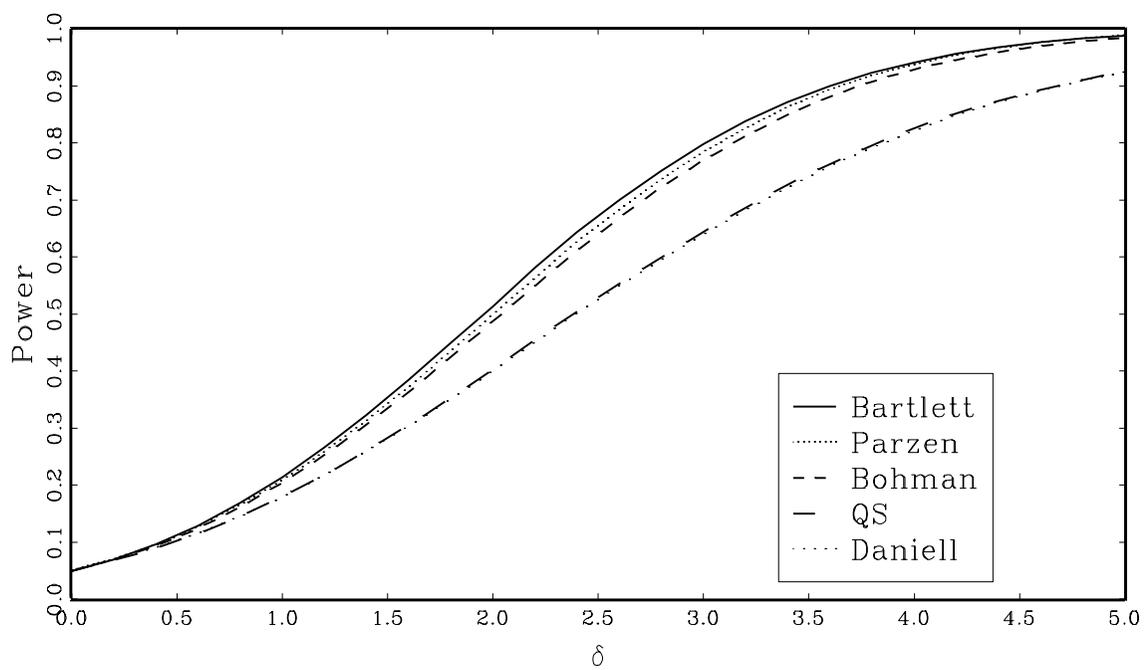


Figure 7: Local Asymptotic Power of t,  $b = 0.5$

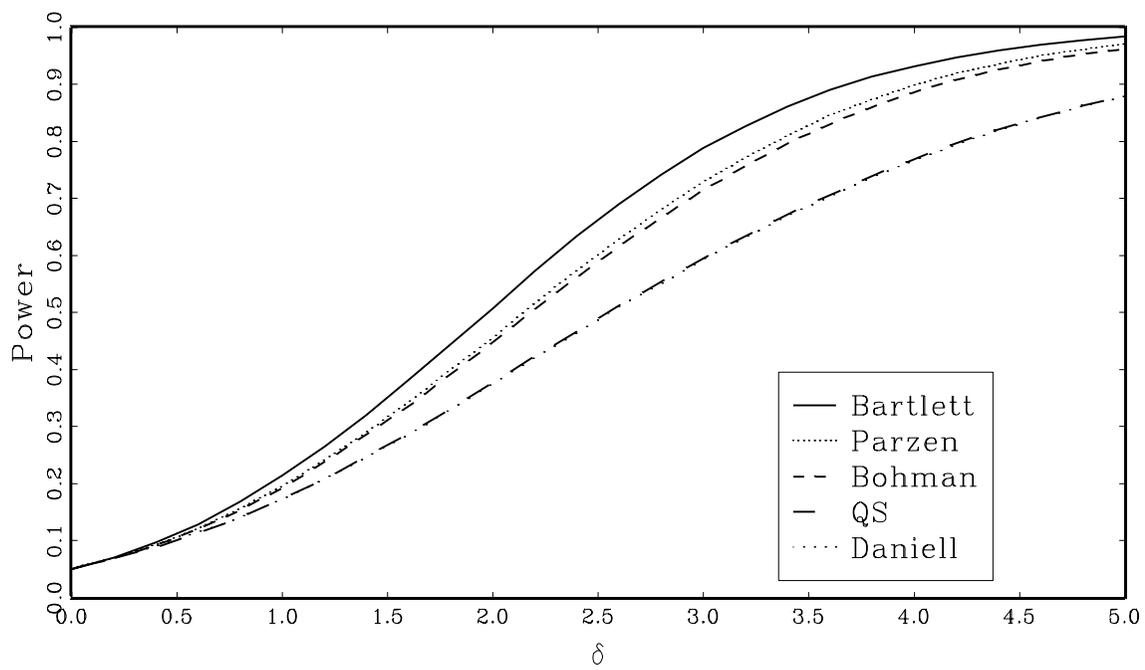


Figure 8: Local Asymptotic Power of t,  $b = 0.7$

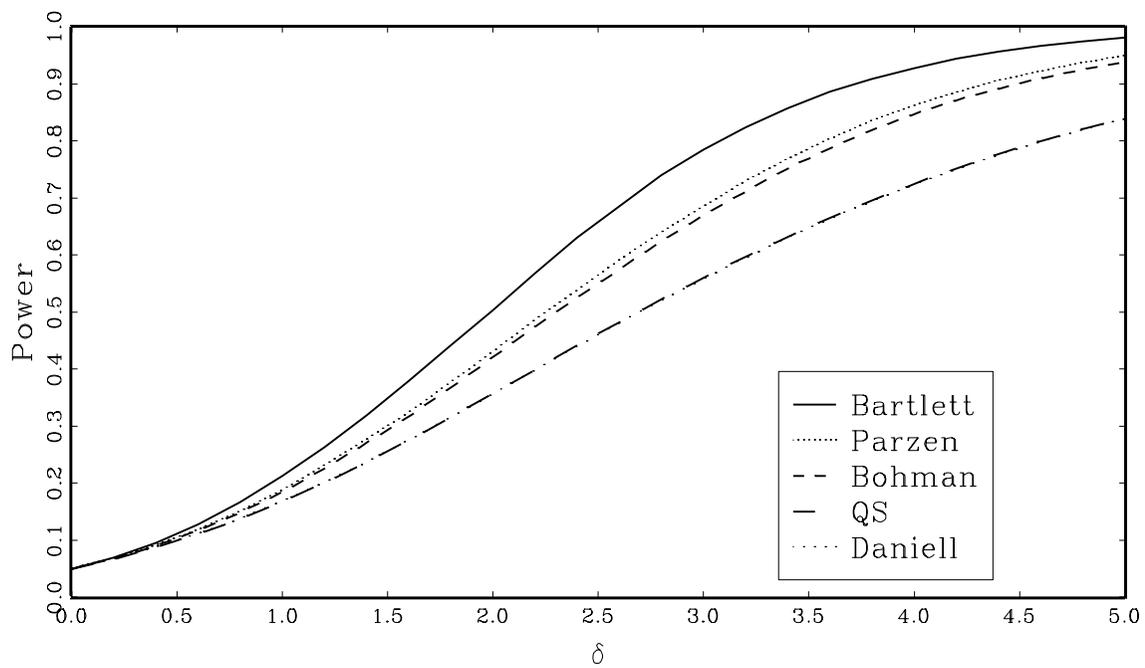


Figure 9: Local Asymptotic Power of  $t$ ,  $b = 0.9$

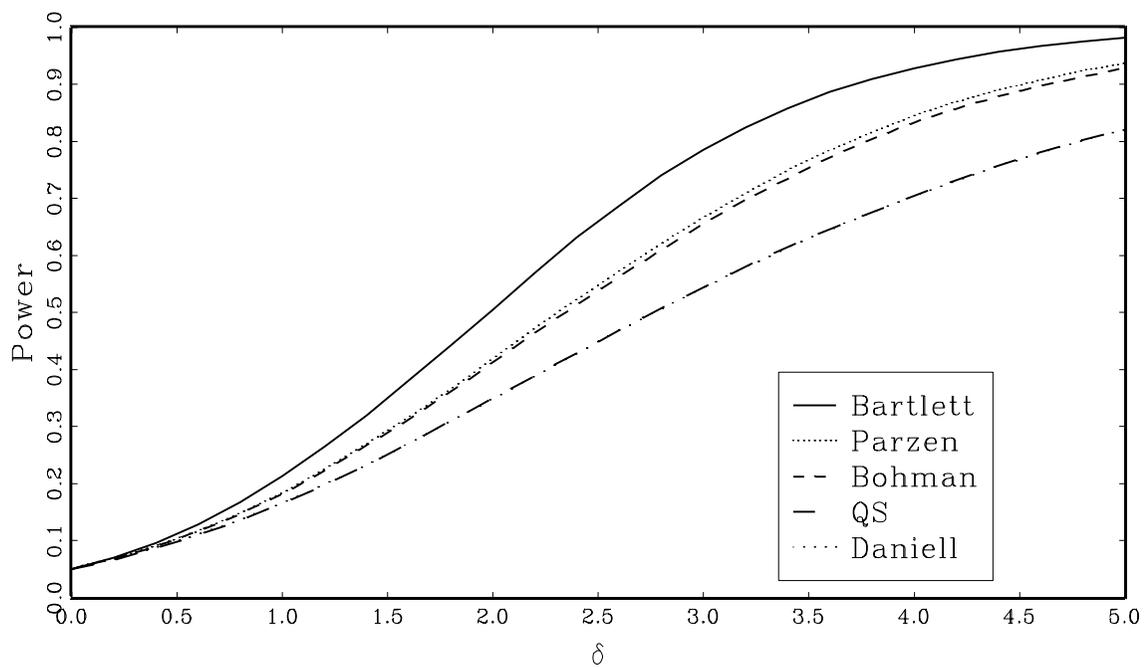


Figure 10: Local Asymptotic Power of  $t$ ,  $b = 1.0$

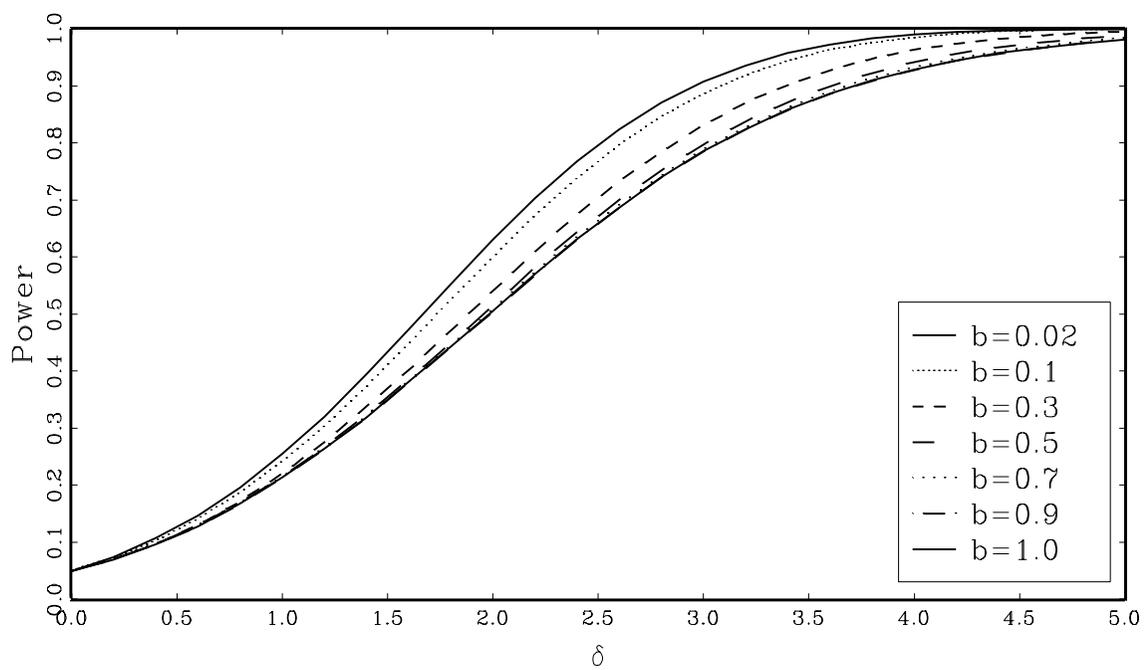


Figure 11: Local Asymptotic Power of t, Bartlett Kernel

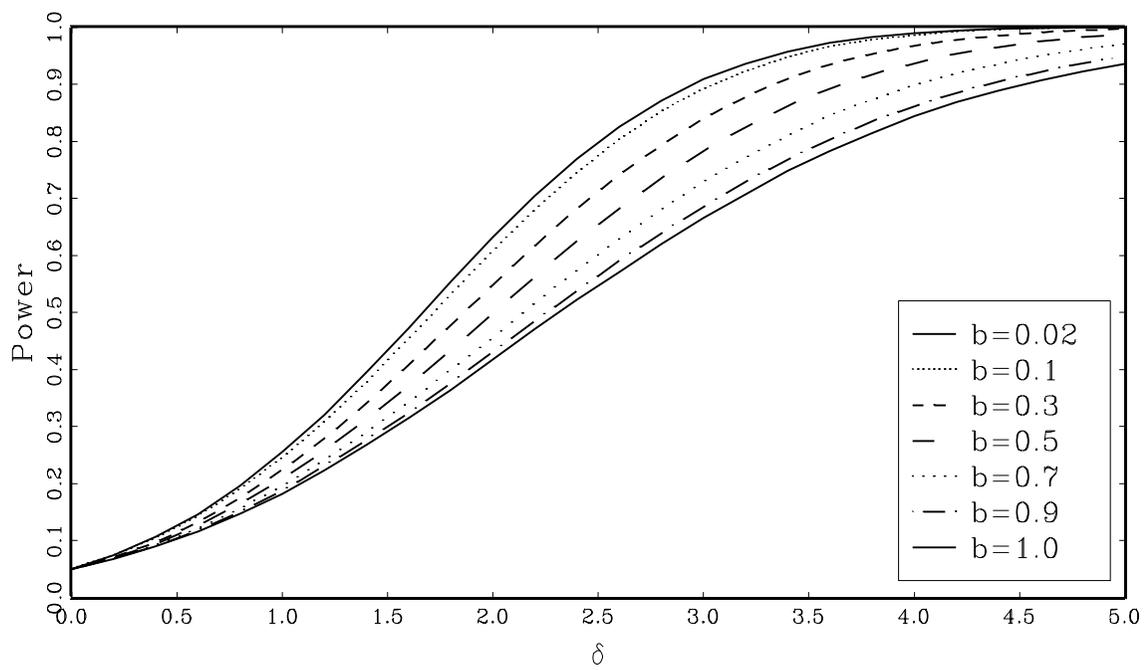


Figure 12: Local Asymptotic Power of t, Parzen Kernel

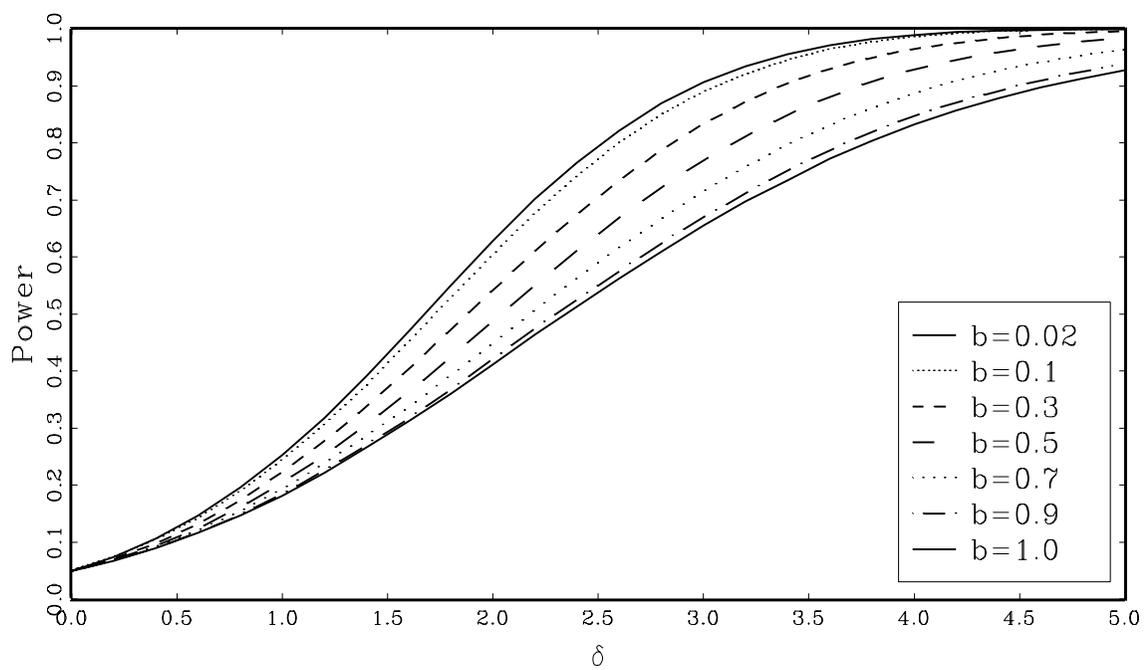


Figure 13: Local Asymptotic Power of t, Bohman Kernel

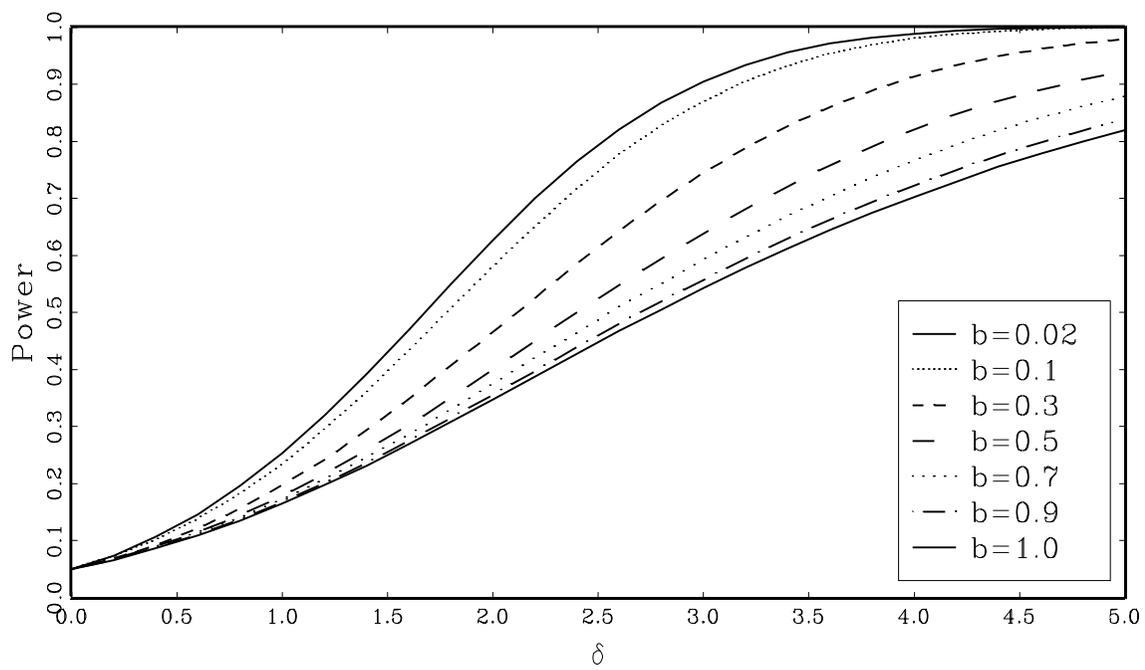


Figure 14: Local Asymptotic Power of t, Daniell Kernel

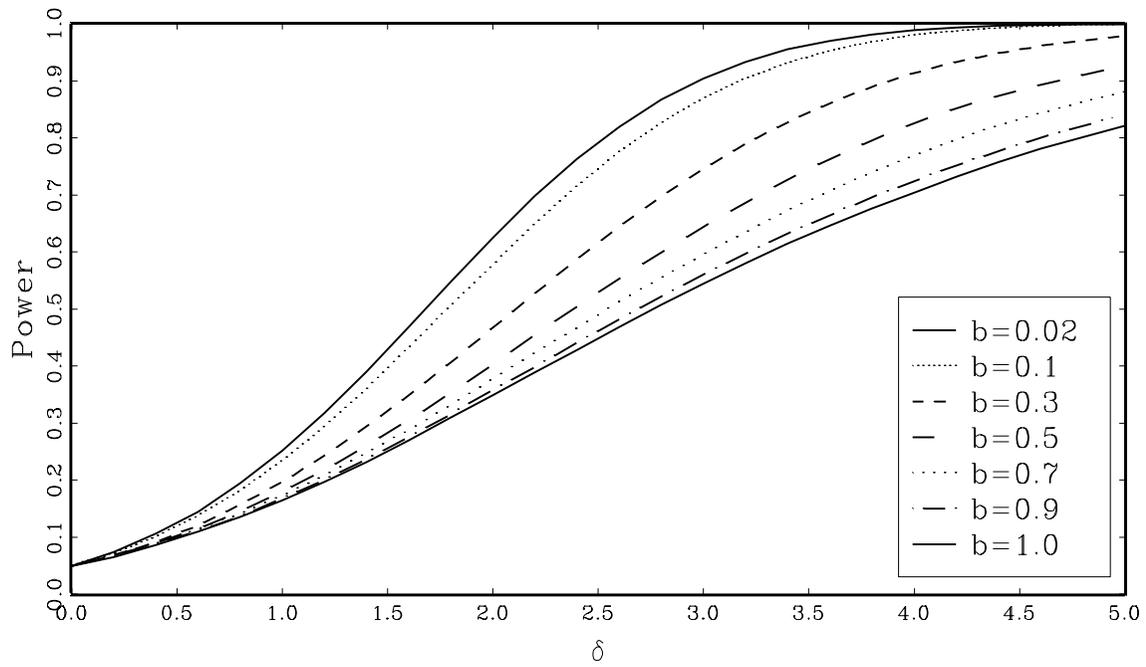


Figure 15: Local Asymptotic Power of  $t$ , QS Kernel