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 August 7, 2003

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.

*Helpful comments provided by Cliff Hurvich, Andy Levin, Jeff Simonoff and seminar participants at NYU (Statistics), U. Texas Austin, NBER/NSF Time Series Conference and 2003 Winter Meetings of the Econometrics Society are gratefully acknowledged. We thank Joel Horowitz and three referees for thoughtful comments that helped us improve the paper. 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 generalized of method of moments (GMM) models. 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\(^1\) 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) and de Jong and Davidson (2000).

We stress at the outset that we are not proposing new estimators or test statistics; rather we focus on improving the asymptotic distribution theory for existing techniques. Our results, however, do provide some guidance on the choice of HAC estimator.

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

---

\(^{1}\) 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.
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 asymptotic variance 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\(^2\). 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.


2
While the assumption that the proportion of sample autocovariances remains fixed as the sample size grows 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 (2002)c, 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 (2002) showing that fixed-b asymptotics is more accurate than small-b asymptotics in Gaussian location models in the special case of the Bartlett kernel with \( b = 1 \). Jansson (2002) proves that fixed-b asymptotics delivers an error in rejection probability that is \( O(T^{-1}) \). 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)).

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)\(^3\).

An obvious alternative to asymptotic approximations is the bootstrap. Recently it has been shown by Hall and Horowitz (1996), Gotze and Kunsch (1996) and Inoue and Shintani (2001) 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 that the bootstrap without blocking performs almost as well as 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 needs to be developed before such a result can be obtained.

Because we are approaching the distribution theory of HAC robust testing using a perspective that differs from the conventional approach, we think it is important to stress here that an

\(^3\)Additional discussion of this point is given by Cushing and McGarvey (1999, p. 80).
important purpose of asymptotic theory is to provide approximations to sampling distributions. While sampling distributions can be obtained exactly under precise distributional assumptions by a change of variables, this approach often requires difficult if not impossible calculations that may be required on a case by case basis. A unifying approach, giving results applicable in a wide variety of settings, must involve approximations. The usual approach is to consider asymptotic approximations. Whether these are useful in any particular setting is a matter of how well the approximation mimics the exact sampling distribution. Although exact comparisons can be made in simple cases, typically, this approximation is most efficiently assessed by finite sample Monte Carlo experiments, and that is our approach here. This point is emphasized by Barndorff-Nielsen and Cox (1989, p. ix)

“The approximate arguments are developed by supposing that some defining quantity, often a sample size but more generally an amount of information, becomes large: it must be stressed that this is a technical device for generating approximations whose adequacy always needs assessing, rather than a 'physical' limiting notion.”

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. 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 discusses how the kernel and bandwidth affect local asymptotic power. Section 7 gives concluding comments. Proofs and some formulas are provided in two appendices.

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 \( \Theta \). Let \( v_t \) denote a vector of observed data and assume that \( q \) moment conditions hold that can be written as

\[
E[f(v_t, \theta_0)] = 0, \quad t = 1, 2, ..., T, \tag{1}
\]

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

\[
\hat{\theta}_T = \arg\min_{\theta \in \Theta} g_T(\theta)'W_T g_T(\theta) \tag{2}
\]

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)

\[
G_T(\hat{\theta}_T)'W_T g_T(\hat{\theta}_T) = 0, \tag{3}
\]

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 \( g_T(\hat{\theta}_T) = 0 \) is attainable and the weighting matrix \( W_T \) is irrelevant. Application of the mean value theorem implies that

\[
g_t(\hat{\theta}_T) = g_t(\theta_0) + G_t(\hat{\theta}_T, \theta_0, \lambda_T)(\hat{\theta}_T - \theta_0) \tag{4}
\]

where \( G_t(\theta_T, \theta_0, \lambda_T) \) denotes the \( q \times p \) matrix whose \( i^{th} \) row is the corresponding row of \( G_t(\theta_T) \) where \( \theta_T = \lambda_{i,T}\theta_0 + (1 - \lambda_{i,T})\theta_T \) for some \( 0 \leq \lambda_{i,T} \leq 1 \) and \( \lambda_T \) is the \( q \times 1 \) vector with \( i^{th} \) 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** \( T^{-1/2} \sum_{t=1}^{[T]} f(v_t, \theta_0) = T^{1/2}g_{[T]}(\theta_0) \Rightarrow \Lambda B_q(r) \) 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 \( \text{p}\lim G_{[:T]}(\hat{\theta}_T) = rG_0 \) and \( \text{p}\lim G_{[:T]}(\hat{\theta}_T, \theta_0, \lambda_T) = rG_0 \) 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 \( \text{p}\lim W_T = W_\infty \) where \( W_\infty \) is a matrix of constants.

These assumptions hold in wide generality for the models seen 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 \( b_T \). 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 \( T \to \infty \),

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

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

\[
\hat{V} = [G_T(\hat{\theta}_T)'W_TG_T(\hat{\theta}_T)]^{-1}G_T(\hat{\theta}_T)'W_T\hat{\Omega}W_TG_T(\hat{\theta}_T)[G_T(\hat{\theta}_T)'W_TG_T(\hat{\theta}_T)]^{-1} \sim N(0, V). \tag{5}
\]

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

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

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 $\hat{\Omega}$ 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 $\hat{\theta}_T$ 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 been 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,

\[ y_t = \mu + u_t, \]
\[ u_t = \rho u_{t-1} + \varepsilon_t + \varphi \varepsilon_{t-1}, \]
\[ \varepsilon_t \sim iid N(0,1), \]
\[ u_0 = \varepsilon_0 = 0. \]

Define the HAC robust t-statistic for \( \mu \) as

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

where \( \hat{\mu} = \overline{y} = \frac{1}{T} \sum_{t=1}^{T} y_t \) 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 \( \mu > 0 \). 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 \( M = 1, 2, 3, \ldots, 49, 50 \). We set the asymptotic significance level to 0.05 and used the usual standard normal critical value of 1.96 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 1.96 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. Similar, although 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 essential increase as $M$ increases. Prewhitening reduces the extent of over-rejection but not remove it. The patterns for MA(1) errors are similar to the iid case and the patterns for ARMA(1,1) errors are combination 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$^4$.

Before moving on, it is useful to discuss in some detail the reason that rejection probabilities have the nonmonotonic pattern for AR(1) errors. When $M$ is small, $\hat{\Omega}$ 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$^5$. This is not what is happening. In fact, as $M$ increases further, bias eventually starts to increase and variance begins to fall. It is the increase in bias the 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 being 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 $\hat{\Omega}$ is identically equal to zero, and this occurs because $\hat{\Omega}$ is

$^4$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. This estimate would introduce additional sampling variability that could offset gains in accuracy from using a higher order approximation. In contrast, the fixed-$b$ asymptotic framework requires no additional estimates.

$^5$This common misperception is an unfortunate result of a folk-lore in the econometrics literature that states that “as $M$ increases, bias in $\hat{\Omega}$ 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 $\hat{\Omega}$ 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 $\hat{\Omega}$ are given for AR(1) processes.
computed using residuals that have a sample average of zero. Obviously, \( \hat{\Omega} = 0 \) it is an estimator with large bias and zero variance. Thus, as \( M \) increases, \( \hat{\Omega} \) is pushed closer to the full weight case.

### 4 A New Asymptotic Theory

#### 4.1 Distribution of \( \hat{\Omega} \) in the Fixed-\( b \) Asymptotic Framework

We now develop a distribution theory for \( \hat{\Omega} \) in the fixed-\( b \) asymptotic framework. We proceed under the asymptotic nesting that \( M = bT \) where \( b \in (0, 1] \) is fixed. The limiting distribution of \( \hat{\Omega} \) 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,

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

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

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

where \( k'_- (1) = \lim_{h \to 0} [(k(1) - k(1-h))/h] \), 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 (see the formula appendix)

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

The moments of \( Q_i(1) \) have been derived by Phillips, Sun and Jin (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

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

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

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

10
where \( \kappa_{ii} \) is the standard commutation matrix\(^6\) and the range of all integrals is 0 to 1. Using these moment results, Hashimzade et al. (2003) prove that

\[
\lim_{b \to 0} E(Q_i(b)) = I_i
\]

and

\[
\lim_{b \to 0} \text{var}(\vec{Q}_i(b)) = 0.
\]

As an illustrative example, consider the Bartlett kernel where

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

\[
\nu(b) = \frac{2}{3}b - \frac{7}{6}b^2 + \frac{7}{15}b^3 + \frac{1}{9}b^4 \quad \text{for} \quad b \leq \frac{1}{2}.
\]

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 \( T \to \infty \),

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

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 \( \hat{\Omega} \) is approximated by the constant \( \Omega \). As \( b \to 0 \), it follows from Lemma 1 that

\[
p \lim_{b \to 0} \Lambda Q_p(b)\Lambda' = \Omega.
\]

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 \( \hat{\Omega} \) 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 \( \hat{\Omega} \) to construct standard errors (details are given below).

Note that Theorem 1 generalizes results obtained by Kiefer and Vogelsang (2002)a and Kiefer and Vogelsang (2002)b where the focus was \( b = 1 \).

When \( q > p \) and the model is overidentified, the limiting expressions for \( \hat{\Omega} \) are more complicated and asymptotic proportionality to \( \Omega \) no longer holds. This was established for the special case of \( b = 1 \) by Vogelsang (2000). This does not mean, however, that valid testing is not possible when using \( \hat{\Omega} \) in overidentified models because the required asymptotic proportionality does hold for \( G_T(\hat{\theta}_T)'W_T\hat{\Omega}W_TG_T(\hat{\theta}_T) \), the middle term in \( \hat{V} \). 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

\(^6\)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.
\[ T \to \infty, \]
\[ G_T(\hat{\theta}_T)'W_TW_TG_T(\hat{\theta}_T) \Rightarrow \Lambda^*Q_p(b)\Lambda'' \]

This theorem shows that \( G_T(\hat{\theta}_T)'W_TW_TG_T(\hat{\theta}_T) \) is asymptotically proportional to \( \Lambda^*\Lambda'' \) and otherwise only depends on the random matrix \( Q_p(b) \). It directly follows that \( \hat{V} \) 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

\[ H_0 : r(\theta_0) = 0 \]
\[ H_1 : r(\theta_0) \neq 0 \]

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

\[ T^{1/2}r(\hat{\theta}_T) \Rightarrow -R(\theta_0)V^{1/2}B_p(1) \equiv N(0, V_R), \tag{8} \]

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

\[ F = T r(\hat{\theta}_T)' \left( R(\hat{\theta}_T)\hat{V}R(\hat{\theta}_T)' \right)^{-1} r(\hat{\theta}_T)/m. \]

When \( m = 1 \) the usual \( t \)-statistic can be computed as

\[ t = \frac{T^{1/2}r(\hat{\theta}_T)}{ \sqrt{R(\hat{\theta}_T)\hat{V}M=-bT R(\hat{\theta}_T)'}}. \]

Often, the significance of individual statistics are of interest which leads to \( t \)-statistics of the form

\[ t = \frac{\hat{\theta}_T}{se(\hat{\theta}_T)}, \]

where \( se(\hat{\theta}_i) = \sqrt{T^{-1}V_{ii}} \) and \( V_{ii} \) is the \( i^{th} \) diagonal element of the \( \hat{V} \) matrix. To avoid any confusion, please 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 $T \to \infty$, 

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

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

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 $\chi^2_m/m$ distribution and $t$ a limiting $N(0, 1)$ distribution regardless of the choice of $M$ and $k(x)$. It is natural to expect that the fixed-$b$ asymptotics provide a more accurate approximation in finite samples than the traditional asymptotics. As shown by the corollary to Theorem 2, as $b \to 0$, $p\lim Q_m(b) = I_m$ 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 the fixed-$b$ asymptotics to deliver a more accurate approximation. The simulations results reported in Section 5 and in the working paper, Kiefer and Vogelsang (2002)c, indicate that this is true in some simple Gaussian models.

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 because of the nonstandard nature of the distribution theory. Obtaining such theoretical results appears very difficult although it may be possible to obtain bounds on the rates at which the error in rejection probability shrinks using the lines of argument in Jansson (2002). Jansson (2002) showed that for a Gaussian location model the error in rejection probability of $F$ is $O(T^{-1})$ when $M = T$ ($b = 1$) and the Bartlett kernel is used. While it remains an open question whether this results holds more generally, our simulations are consistent with this possibility. Obviously, this is an area of current research.
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

$$cv(b) = a_0 + a_1 b + a_2 b^2 + a_3 b^3.$$ 

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 $b = 0.02, 0.04, ..., 0.98, 1.0$ 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 (2002)c, 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 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 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. This pattern is consistent with Hashimzade and Vogelsang (2003) who found in finite sample simulations that the fixed-\(b\) asymptotic approximation for \(\hat{\Omega}\) improves as \(M\) increases. Intuitively, the fixed-\(b\) asymptotics is capturing the downward bias in \(\hat{\Omega}\) 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.

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.

The improvement of the asymptotic approximation using fixed-\(b\) asymptotics occurs because fixed-\(b\) asymptotics captures much of the bias and sampling variability induced by the kernel and bandwidth. In the case of iid errors, the fixed-\(b\) asymptotics is remarkably accurate. As serial correlation in the errors becomes stronger, the tendency to over-reject appears. 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 following Gotze and Kunsch (1996) as follows. The original data, \(y_1, y_2, \ldots, y_T\) is divided into \(T - l + 1\) overlapping blocks of length, \(l\). For simplicity, we report results for \(l = 1, 4, 10\) all of which factor into \(T = 50\).
$T/l$ blocks are drawn randomly with replacement to form a bootstrap sample labeled $y_1^*, y_2^*, ..., y_T^*$. The bootstrap $t$ statistic is computed in two ways. The first statistic, labelled the naive bootstrap test, is defined as

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

where $\hat{\mu}^* = \bar{y}^* = T^{-1} \sum_{t=1}^{T} y_t^*$ and $\Omega^*$ is computed using formula (6) with $f(y_t^*, \hat{\mu}^*) = \bar{u}_t^* = y_t^* - \hat{\mu}^*$. The second statistic, labelled the Gotze-Kunsch (GK) bootstrap test, is defined as

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

where

$$\bar{y} = (T - l + 1)^{-1} \sum_{j=0}^{T-l} \left( l^{-1} \sum_{i=1}^{l} y_{j+i} \right), \quad \Omega^* = T^{-1} \sum_{j=1}^{T/l} \left( \sum_{i=1}^{l} \bar{u}_{l(j-1) + i} \right)^2.$$

The centering and standardization of the GK test is required for the bootstrap to attain second order accuracy under small-$b$ asymptotics.

Empirical rejection probabilities using the bootstrap critical values are reported in Figures 1, 2, 3 and 4. The label NBoot refers to the naive bootstrap and the label GKBoot refers to the GK bootstrap. In Figures 1 and 2 the block length is 4. In figures 3 and 4 results are reported for the naive bootstrap for block lengths 1, 4 and 10.

There are some very compelling patterns in the Figures. In figures 1 and 2 it is obvious that the GK bootstrap provides a more accurate approximation than the standard normal approximation although for AR(1) errors the prewhitened standard normal approximation outperforms the GK bootstrap. On the other hand, the GK bootstrap is inferior to the fixed-$b$ asymptotic approximation except when $b$ is very close to zero ($M = 1, 2$) in which case differences are small. These simulation results suggest that while the GK bootstrap provides the expected asymptotic refinement over the standard normal approximation, it does not approach the accuracy of the fixed-$b$ asymptotics when $M$ is not very small. This is not that surprising given that the theoretical results for the GK bootstrap were obtained under the small-$b$ asymptotic nesting. It is well known that the GK bootstrap does not attain second order correctness when the Bartlett kernel is used. Yet, the simulations suggest that GK bootstrap appears to provide refinements when the Bartlett kernel is used.

Contrary to theoretical results by Davison and Hall (1993) in the bootstrap literature, the naive bootstrap performs quite well and yields rejection probabilities that are very similar to the fixed-$b$ asymptotics. It appears that the naive bootstrap may provide a higher order refinement over the fixed-$b$ asymptotics although in the case of AR(1) errors, prewhitening and fixed-$b$ asymptotics is more accurate than the naive bootstrap.
Regardless of the data generating process or kernel, it is striking how the naive bootstrap closely follows the fixed-\(b\) asymptotics while the GK bootstrap closely follows the standard small-\(b\) asymptotics. Clearly, there are some very interesting properties of the bootstrap that warrant additional research. It appears that fixed-\(b\) asymptotics could play an important role in understanding the strong performance of the naive bootstrap in spite of the negative theoretical results obtained in the literature. Our conjecture is that while the GK bootstrap nails the second order term of the small-\(b\) Edgeworth expansion, the Edgeworth expansion itself does not provide a very accurate approximation when the bandwidth isn’t small. On the other hand, we conjecture that the naive bootstrap may capture a second order term in the fixed-\(b\) asymptotics. We hope these interesting puzzles will stimulate new theoretical work in the bootstrap literature.

Finally, we report some results in Figures 3 and 4 illustrating the sensitivity of the naive bootstrap to the choice of block length. For AR(1) processes, a block length of 10 does fairly well, almost as well as the prewhitened test with fixed-\(b\) critical values. In contrast for MA(1) errors, a large block length can cause the bootstrap to perform more poorly than the fixed-\(b\) asymptotics.

6 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 (2002)a. Therefore, details are kept to a minimum. Consider the regression model

\[ y_t = x_t' \theta_0 + u_t \]  

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

\[ H_0 : \theta_{i0} \leq 0 \]
\[ H_1 : \theta_{i0} = cT^{-1/2} \]

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 (2002)a to easily establish that
under the local alternative, $H_1$, as $T \to \infty$,
\[ t^*_b \Rightarrow \frac{\delta + B_1(1)}{\sqrt{Q_1(b)}}, \tag{10} \]
where $\delta = \frac{c}{\sqrt{V_{ii}}}$, $V_{ii}$ is the $i^{th}$ diagonal element of $V$, and $Q_1(b)$ is given by Definition 1 for $i = 1$.

Asymptotic power curves can be computed for given bandwidths and kernels by simulating the asymptotic distribution of $t^*_b$ 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 $\delta = 0, 0.2, 0.4, ..., 4.8, 5.0$ using 5% asymptotic null critical values.

The power results are reported in two ways. Figures 5-12 plot power across the kernels for a given value of $b$. Figures 13-17 plot power across values of $b$ for a given kernel. Figures 5-12 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 13-17 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.

7 Conclusions

We have provided a new approach to the asymptotic theory of HAC robust testing. We consider tests based on the popular kernel based nonparametric 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. 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. Our approximation is an unambiguous improvement over the standard 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).

8 Appendix: Proofs

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

\[ k^*(x) = k\left(\frac{x}{b}\right), \]

\[ K_{ij} = k\left(\frac{i-j}{bT}\right) = k^*\left(\frac{i-j}{T}\right), \]

\[ \Delta^2 K_{ij} = (K_{ij} - K_{i,j+1}) - (K_{i+1,j} - K_{i+1,j+1}), \]

\[ D_T^*(r) = T^2 \left[ \left( k^*\left(\frac{rT}{T} + 1\right) - k^*\left(\frac{rT}{T}\right) \right) - \left( k^*\left(\frac{rT}{T}\right) - k^*\left(\frac{rT}{T} - 1\right) \right) \right]. \]

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 \(k^{*\prime\prime}(r)\) exists then \(\lim_{T \to \infty} D_T^*(r) = k^{*\prime\prime}(r)\) by the definition of the second derivative. If \(k^{*\prime\prime}(r)\) is continuous, then \(D_T^*(r)\) converges to \(k^{*\prime\prime}(r)\) uniformly in \(r\). Define the stochastic process

\[ X_T(r) = G_T(\theta_T)' W_T T^{1/2} g_{[rT]}(\theta_0). \]

It directly follows from Assumptions 2, 3 and 4 that

\[ X_T(r) \Rightarrow G_0' W_\infty AB_q(r) \equiv \Lambda^* B_p(r). \]
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 \( G_T(\hat{\theta}_T)'W_T g_T(\hat{\theta}_T) = 0 \) gives

\[
0 = G_T(\hat{\theta}_T)'W_T g_T(\theta_0) + G_T(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T, \theta_0, \lambda_T)(\hat{\theta}_T - \theta_0). \tag{12}
\]

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

\[
T^{1/2}(\hat{\theta}_T - \theta_0) = -[G_T(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T, \theta_0, \lambda_T)]^{-1}G_T(\hat{\theta}_T)'W_T T^{1/2}g_T(\theta_0)
= -[G_T(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T, \theta_0, \lambda_T)]^{-1}X_T(1). \tag{13}
\]

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

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

The proof of Theorem 1 follows the same arguments as the proof for Theorem 2 and is omitted.

Proof of Theorem 2: Define the random process

\[
\tilde{X}_T(r) = G_T(\hat{\theta}_T)'W_T T^{1/2} g_{[rT]}(\hat{\theta}_T).
\]

Plugging in for \( g_{[rT]}(\hat{\theta}_T) \) using (4) gives

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

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

\[
\tilde{X}_T(r) \Rightarrow \Lambda^* B_p(r) - rG_0'W_\infty G_0 (G_0'W_\infty G_0)^{-1} \Lambda^* B_p(1)
= \Lambda^* (B_p(r) - rB_p(1)) \equiv \Lambda^* \tilde{B}_p(r). \tag{14}
\]

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 (2002)a, it is straightforward to show that

\[
\hat{\Omega} = T^{-1} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} \Delta^2 K_{ij} T g_i(\hat{\theta}_T) T g_j(\hat{\theta}_T)'
+ g_T(\hat{\theta}_T) \sum_{i=1}^{T-1} (K_{Ti} - K_{T,i+1}) T g_i(\hat{\theta}_T)' + \left( \sum_{j=1}^{T} f(v_j, \hat{\theta}_T) K_{JT} \right) g_T(\hat{\theta}_T)' \tag{15}
\]
Using (15) it directly follows that
\[
G_T(\hat{\theta}_T)'W_TW_T\hat{\Omega}W_TG_T(\hat{\theta}_T)
\]
\[
= T^{-1} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} \Delta^2 K_{ij} G_T(\hat{\theta}_T)'W_T T g_i(\hat{\theta}_T) T g_j(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T)
\]
\[
= T^{-2} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} T^2 \Delta^2 K_{ij} G_T(\hat{\theta}_T)'W_T T^{1/2} g_i(\hat{\theta}_T) T^{1/2} g_j(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T)
\]
\[
= T^{-2} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} -D_T^* \left( \frac{i-j}{T} \right) G_T(\hat{\theta}_T)'W_T T^{1/2} g_i(\hat{\theta}_T) T^{1/2} g_j(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T)
\]
(16)
where the second and third terms of (15) vanish because from (3) we have
\[
G_T(\hat{\theta}_T)'W_T T g_T(\hat{\theta}_T) = 0,
\]
\[
T g_T(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T) = 0.
\]
The rest of proof is divided into three cases.

Case 1: \(k(x)\) is twice continuously differentiable. Using (16) it follows that
\[
G_T(\hat{\theta}_T)'W_T \hat{\Omega}W_T G_T(\hat{\theta}_T)
\]
\[
= - \int_0^1 \int_0^1 D_T^*(r-s) G_T(\hat{\theta}_T)'W_T T^{1/2} g_{[rT]}(\hat{\theta}_T) T^{1/2} g_{[sT]}(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T) dr ds
\]
\[
= - \int_0^1 \int_0^1 D_T^*(r-s) \tilde{X}_T(r) \tilde{X}_T(s)' dr ds
\]
\[
\Rightarrow - \Lambda^* \int_0^1 \int_0^1 k''(r-s) \tilde{B}_p(r) \tilde{B}_p(s)' dr ds \Lambda^*,
\]
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 \(|i-j| > [bT]\) and \(\Delta^2 K_{ij} = -k^* (b - \frac{1}{2})\) for \(|i-j| = [bT]\), break up the double sum in the second line of the expression for \(G_T(\hat{\theta}_T)'W_T \hat{\Omega}W_T G_T(\hat{\theta}_T)\) into three pieces corresponding to \(|i-j| < [bT]\), \(|i-j| = [bT]\), and \(|i-j| > [bT]\) to obtain
\[
G_T(\hat{\theta}_T)'W_T \hat{\Omega}W_T G_T(\hat{\theta}_T) =
\]
\[
T^{-2} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} 1(|i-j| < |bT|) T^2 \Delta^2 K_{ij} G_T(\hat{\theta}_T)' W_T T_{ij}^{1/2} g_i(\hat{\theta}_T) T_{ij}^{1/2} g_j(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T) \\
- Tk^* \left( b - \frac{1}{T} \right) T^{-1} \sum_{i=1}^{T-|bT|-1} G_T(\hat{\theta}_T)' W_T T_{ij}^{1/2} g_i(\hat{\theta}_T) T_{ij}^{1/2} g_j(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T) \\
- Tk^* \left( b - \frac{1}{T} \right) T^{-1} \sum_{j=1}^{T-|bT|-1} G_T(\hat{\theta}_T)' W_T T_{ij}^{1/2} g_i(\hat{\theta}_T) T_{ij}^{1/2} g_j(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T). 
\]

It directly follows that

\[
G_T(\hat{\theta}_T)' W_T \hat{\Omega} W_T G_T(\hat{\theta}_T) = \\
- \int \int_{|r-s|<b} D_{r-s}^T r(s) \hat{X}_T(r) \hat{X}_T(s)' dr ds - Tk^* \left( b - \frac{1}{T} \right) \int_0^{1-b} \left( \hat{X}_T(r+b) \hat{X}_T(r)' + \hat{X}_T(r) \hat{X}_T(r+b)' \right) dr.
\]

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

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

Therefore, by the continuous mapping theorem

\[
G_T(\hat{\theta}_T)' W_T \hat{\Omega} W_T G_T(\hat{\theta}_T) = \\
\Lambda^* \left[ - \int \int_{|r-s|<b} k'''(r-s) \tilde{B}_p(r) \tilde{B}_p(s)' dr ds \right. \\
\left. + k'''(b) \int_0^{1-b} \left( \tilde{B}_p(r+b) \tilde{B}_p(r)' + \tilde{B}_p(r) \tilde{B}_p(r+b)' \right) dr \right] \Lambda''.
\]

Case 3: \( k(x) \) is the Bartlett kernel. It is easy to calculate that for the Bartlett kernel, \( \Delta^2 K_{ij} = \frac{2}{bT} \) for \( |i-j| = 0 \), \( \Delta^2 K_{ij} = -\frac{1}{T} \) for \( |i-j| = |bT| \) and \( \Delta^2 K_{ij} = 0 \) otherwise. Therefore we have

\[
G_T(\hat{\theta}_T)' W_T \hat{\Omega} W_T G_T(\hat{\theta}_T) = \\
\frac{2}{bT} \sum_{i=1}^{T-1} G_T(\hat{\theta}_T)' W_T T_{ij}^{1/2} g_i(\hat{\theta}_T) T_{ij}^{1/2} g_j(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T) \\
- \frac{1}{bT} \sum_{i=1}^{T-|bT|-1} G_T(\hat{\theta}_T)' W_T T_{ij}^{1/2} g_i(\hat{\theta}_T) T_{ij}^{1/2} g_j(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T) \\
- \frac{1}{bT} \sum_{j=1}^{T-|bT|-1} G_T(\hat{\theta}_T)' W_T T_{ij}^{1/2} g_i(\hat{\theta}_T) T_{ij}^{1/2} g_j(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T).
\]
\[
= \frac{2}{b} \int_{0}^{1} \bar{X}_T(r) \bar{X}_T(r)' dr - \frac{1}{b} \int_{0}^{1-b} \left( \bar{X}_T(r+b) \bar{X}_T(r) + \bar{X}_T(r) \bar{X}_T(r+b) \right) dr
\]

\[
\Rightarrow \Lambda^* \left[ \frac{2}{b} \int_{0}^{1} \tilde{B}_p(r) \tilde{B}_p(r)' dr - \frac{1}{b} \int_{0}^{1-b} \left( \tilde{B}_p(r+b) \tilde{B}_p(r) + \tilde{B}_p(r) \tilde{B}_p(r+b) \right) dr \right] \Lambda^{**}.
\]

**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

\[
T^{1/2_r}(\hat{\theta}_T) \Rightarrow -R(\theta_0) \left( G_0' W_\infty G_0 \right)^{-1} G_0' W_\infty \Lambda B_q(1)
\]

\[
\equiv -R(\theta_0) \left( G_0' W_\infty G_0 \right)^{-1} \Lambda^* B_p(1)
\]

\[
\equiv \Lambda^{**} B_m(1),
\]

(17)

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

\[
R(\hat{\theta}_T) \tilde{V}_{M=b_T} R(\hat{\theta}_T) = R(\hat{\theta}_T) \left[ G_T(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T, \theta_0, \lambda_T) \right]^{-1} G_T(\hat{\theta}_T)' W_T \hat{\Omega}_{M=T W_T G_T(\hat{\theta}_T)}
\]

\[
\times \left[ G_T(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T, \theta_0, \lambda_T) \right]^{-1} R(\hat{\theta}_T)'
\]

\[
\Rightarrow R(\theta_0) \left( G_0' W_\infty G_0 \right)^{-1} \Lambda^* Q_p(b) \Lambda^{**} \left( G_0' W_\infty G_0 \right)^{-1} R(\theta_0)'
\]

\[
\equiv \Lambda^{**} Q_m(b) \Lambda^{**},
\]

(18)

where we use the fact that

\[
R(\theta_0) \left( G_0' W_\infty G_0 \right)^{-1} \Lambda^* \tilde{B}_p(r) = R(\theta_0) \left( G_0' W_\infty G_0 \right)^{-1} \Lambda^* \left( B_p(r) - r B_p(1) \right)
\]

\[
\equiv \Lambda^{**} (B_m(r) - r B_m(1))
\]

\[
= \Lambda^{**} \tilde{B}_m(r).
\]

Using (17) and (18) it directly follows that

\[
F^* = Tr(\hat{\theta}_T) \left( R(\hat{\theta}_T) \tilde{V}_{M=b_T} R(\hat{\theta}_T) \right)^{-1} r(\hat{\theta}_T)/m
\]

\[
= T^{1/2_r}(\hat{\theta}_T) \left( R(\hat{\theta}_T) \tilde{V}_{M=b_T} R(\hat{\theta}_T) \right)^{-1} T^{1/2_r}(\hat{\theta}_T)/m
\]

\[
\Rightarrow (\Lambda^{**} B_m(1))' \left( \Lambda^{**} Q_m(b) \Lambda^{**} \right)^{-1} (\Lambda^{**} B_m(1))/m
\]

\[
\equiv B_m(1)' Q_m(b)^{-1} B_m(1)/m,
\]

which completes the proof.
9 Appendix: Kernel Formulas

The formulas for the kernels analyzed in this paper are

\[
\text{Bartlett } k(x) = \begin{cases} 
1 - |x| & \text{for } |x| \leq 1, \\
0 & \text{otherwise},
\end{cases}
\]

\[
\text{Parzen } k(x) = \begin{cases} 
1 - 6x^2 + 6|x|^3 & \text{for } |x| \leq \frac{1}{2}, \\
2(1 - |x|^3) & \text{for } \frac{1}{2} \leq |x| \leq 1 \\
0 & \text{otherwise},
\end{cases}
\]

\[
\text{Bohman } k(x) = \begin{cases} 
(1 - |x|)\cos(\pi x) + \sin(\pi |x|)/\pi & \text{for } |x| \leq 1, \\
0 & \text{otherwise},
\end{cases}
\]

\[
\text{Quadratic Spectral (QS) } k(x) = \frac{25}{12\pi^2 x^2} \left( \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right),
\]

\[
\text{Daniell } k(x) = \frac{\sin(\pi x)}{\pi x}.
\]
References


Table I: Asymptotic Critical Value Function Coefficients for $t$.

\[ cv(b) = a_0 + a_1 b + a_2 b^2 + a_3 b^3 \]

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

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 \( cv(b) = a_0 + a_1 b + a_2 b^2 + a_3 b^3 \). 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: Empirical Null Rejection Probabilities
Bartlett Kernel, T=50, 5% Nominal Level
Figure 4: Empirical Null Rejection Probabilities
QS Kernel, T=50, 5% Nominal Level
Figure 5: Local Asymptotic Power of $t$, $b = 0.02$

Figure 6: Local Asymptotic Power of $t$, $b = 0.06$
Figure 7: Local Asymptotic Power of $t$, $b = 0.1$

Figure 8: Local Asymptotic Power of $t$, $b = 0.3$
Figure 9: Local Asymptotic Power of $t$, $b = 0.5$

Figure 10: Local Asymptotic Power of $t$, $b = 0.7$
Figure 11: Local Asymptotic Power of $t$, $b = 0.9$

Figure 12: Local Asymptotic Power of $t$, $b = 1.0$
Figure 13: Local Asymptotic Power of t, Bartlett Kernel

Figure 14: Local Asymptotic Power of t, Parzen Kernel
Figure 15: Local Asymptotic Power of t, Bohman Kernel

Figure 16: Local Asymptotic Power of t, Daniell Kernel
Figure 17: Local Asymptotic Power of t, QS Kernel