Block Bootstrap HAC Robust Tests: The Sophistication of the Naive Bootstrap

Sílvia Gonçalves
Département de sciences économiques, CIREQ and CIRANO, Université de Montréal

and

Timothy J. Vogelsang*
Department of Economics, Michigan State University

May 14, 2007

Abstract

This paper studies the properties of naive block bootstrap tests that are scaled by zero frequency spectral density estimators (long run variance estimators). The naive bootstrap is a bootstrap where the formula used in the bootstrap world to compute standard errors is the same as the formula used on the original data. Simulation evidence shows that the naive bootstrap can be much more accurate than the standard normal approximation. The larger the HAC bandwidth, the greater the improvement. This improvement holds for a large number of popular kernels, including the Bartlett kernel, and it holds when the i.i.d. bootstrap is used and yet the data are serially correlated. Using recently developed fixed-b asymptotics for HAC robust tests, we provide theoretical results that can explain the finite sample patterns. We show that the block bootstrap, including the special case of the i.i.d. bootstrap, has the same limiting distribution as the fixed-b asymptotic distribution. For the special case of a location model with a Bartlett kernel HAC variance estimator, we provide theoretical results that suggest the naive bootstrap is more accurate than the standard normal approximation. Our theoretical results lay the foundation for a bootstrap asymptotic theory that is an alternative to the traditional approach based on Edgeworth expansions.

*For helpful comments and suggestions we thank Lutz Kilian, Guido Kuersteiner, Nour Meddahi, Ulrich Mueller, Pierre Perron, Yixiao Sun, Hal White, and seminar participants at Boston University, Queen's University, U. Toronto, Johns Hopkins Biostatistics, Chicago GSB, UCLA, UCSD, U. Michigan, U. Laval, U. Pittsburgh, U. Wisconsin, Cornell U. the 2005 Winter Meetings of the Econometrics Society in Philadelphia, the 2005 European Economic Association Meetings in Amsterdam and the 2004 Forecasting Conference at Duke University. Vogelsang thanks the Center for Analytic Economics at Cornell and gratefully acknowledges financial support from the NSF through grant SES-0525707.

1 Introduction

The bootstrap, originally proposed by **?**, has become a standard tool used in statistics and as a way to provide critical values for test statistics. The impact of the bootstrap on econometrics has been substantial. The bootstrap is a computationally intensive alternative to using critical values based on asymptotic approximations. Interest in the bootstrap has increased over time for two reasons. First, the computational power of computers has increased to the point where the bootstrap can often be quickly implemented in practice. Second, in simulations studies it is often found that the approximation delivered by the bootstrap is more accurate than the approximation given by first order asymptotics. Theoretical explanations for this improvement in accuracy are typically established using higher order asymptotics based on Edgeworth expansions.

With time series data, implementation of the bootstrap is more complicated than in the i.i.d. case because of the dependence structure in the data. Many variants of the bootstrap have been proposed for dependent data including the well known moving blocks bootstrap originally proposed by **?**. Theoretical conditions under which the block bootstrap can be expected to provide refinements have been established by **?**, **?**, **?** and others. Refinements of the block bootstrap in generalized method of moments (GMM) models have been shown by **?** and **?**. The theoretical results in these papers have been established using Edgeworth expansions with leading terms that are distributed standard normal.

When the moving blocks bootstrap (MBB) is applied to tests based on heteroskedasticity autocorrelation (HAC) robust variance estimators, a particular version of the MBB has been labeled "naive" by **?**. The naive bootstrap uses the same formula for the HAC estimator in the bootstrap world as is used on the original data. While this may seem to be a natural way to proceed with the MBB, **?** and **?** have shown that the naive MBB will not provide higher order accuracy as measured by edgeworth expansions. To obtain higher order edgeworth results they show that the HAC estimator in the bootstrap world needs to be computed using a formula that reflects the constraint on the correlation structure of the bootstrap data imposed by the moving blocks scheme. Recent work by **?** extends the Edgeworth analysis to certain testing problems in the GMM framework.

In a recent paper, **?** reported small sample simulation results for HAC robust t-statistics for testing hypotheses about the sample mean of a stationary univariate time series. They found that the naive bootstrap, including the i.i.d. bootstrap, can dramatically outperform the standard normal approximation, and this improvement over the standard normal approximation occurs for many kernels including the Bartlett kernel. The case of the Bartlett kernel is interesting because the edgeworth expansion results in the literature suggest that even the non-naive version of the MBB will not be more accurate than the standard normal approximation in the Bartlett kernel case. The simulations reported by **?** also exhibited an interesting and persistent pattern: the naive

MBB, especially the i.i.d. bootstrap, closely mimics rejections that are obtained when using the fixed-b asymptotic approximation proposed in their paper. Because these finite sample patterns are not predicted by the existing Edgeworth theory, an alternative theory is needed to understand the finite sample performance of the naive MBB.

In this paper we develop a theoretical framework that can be used to explain the finite sample patterns reported by **?**. We make two theoretical contributions. First, we provide sufficient conditions under which the naive MBB has the same first order fixed-b asymptotic distribution as the original statistic. This result holds for fixed block lengths (including the special case of the i.i.d. bootstrap) and for block lengths that increase with the sample size but at a slower rate. This result explains why rejections using the naive MBB closely follow the rejections using fixed-b asymptotic critical values. Second, in a simple location model we develop a higher order asymptotic theory to show that the i.i.d. bootstrap has an error in rejection probability (ERP) that converges to zero faster than the ERP of the standard normal approximation. *Ex ante*, it is not intuitively obvious that the i.i.d. bootstrap could be more accurate than the standard normal approximation (see **?**). *Ex post*, this property is no longer surprising when viewed from within the fixed-b asymptotic framework.

In establishing the higher order properties of the i.i.d. bootstrap, we show that the fixed-b approximation is more accurate than the standard normal approximation in the simple location model. This finding on its own is a contribution to the literature on the higher order asymptotic properties of the fixed-b and related asymptotic theories obtained by **?** and **?** for the simple location model. Because **?** and **?** obtain results under the assumption that the data is Gaussian, their results cannot be applied to the bootstrap because the bootstrap data cannot be Gaussian by construction. When dropping the Gaussian assumption, we find that one-tail tests remain more accurate than the standard normal approximation, but we find that the rate at which the ERP converges to zero is not as fast as the rates found by **?** and **?** for the Gaussian case. While it is possible that their results extend to the non-Gaussian case studied here, establishing such results appear very difficult.

The remainder of the paper is organized as follows. In the next section we describe the model and test statistics. We review the fixed-b asymptotic approximation. Section 3 reports simulation results for the simple location model and for a stationary regression model. The simulations illustrate the performance of the naive MBB relative to the standard normal and fixed-b approximations. In Sections 4 and 5 we provide theoretical explanations for several of the patterns that emerge from the simulations. Section 4 focuses on stationary regression models and establishes the first order asymptotic equivalence between the naive bootstrap and the fixed-b asymptotic approximation. These results could be generalized in straightforward ways to nonlinear models estimated by generalized method of moments. In Section 5 we narrow the focus to the simple location model and we provide higher order asymptotic results for Bartlett kernel based tests. These results establish that the fixed-b asymptotic approximation and the naive i.i.d. bootstrap have ERPs that converge to zero at rates faster than the standard normal approximation. In Section 6 we discuss heuristic comparisons between fixed-b asymptotic approximations and the Edgeworth approximations derived by **?** in an effort to shed some light on the relative performance of Edgeworth approximations and the naive bootstrap/fixed-b asymptotics in the simple location model. Proofs are given in two mathematical appendices.

2 Model and Test Statistics

Throughout the paper we focus on stationary regression models of the form

$$y_t = x_t'\beta + u_t, \quad t = 1, 2, \ldots, T, \tag{1}$$

where x_t and β are $s \times 1$ vectors. The stationary time series $\{x_t\}$ and $\{u_t\}$ are autocorrelated and possibly conditionally heteroskedastic. It is assumed that u_t is mean zero and is uncorrelated with x_t.

The parameter of interest is β and its estimator is $\hat{\beta} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t y_t$, the ordinary least squares (OLS) estimator. Let $Q = E(x_t x_t')$ and $\Omega = \lim_{T\to\infty} Var\left(T^{-1/2}\sum_{t=1}^{T} v_t\right)$, where $v_t = x_t u_t$. For HAC robust testing we require estimates of Q and Ω. The usual estimate of Q is $\hat{Q} = T^{-1}\sum_{t=1}^{T} x_t x_t'$. Estimation of Ω is often implemented with a kernel variance estimator such as

$$\hat{\Omega} = \sum_{j=-(T-1)}^{T-1} k\left(\frac{j}{M}\right)\hat{\Gamma}_j, \tag{2}$$

where $k(x)$ is a kernel function such that $k(x) = k(-x)$, $k(0) = 1$, $|k(x)| \le 1$, $k(x)$ is continuous at $x = 0$, and $\int_{-\infty}^{\infty} k^2(x) < \infty$. Here, for $j \ge 0$, $\hat{\Gamma}_j = T^{-1}\sum_{t=j+1}^{T} \hat{v}_t \hat{v}_{t-j}'$ are the sample autocovariances of the score vector $\hat{v}_t = x_t \hat{u}_t$, with $\hat{u}_t = y_t - x_t'\hat{\beta}$ the OLS residuals, and $\hat{\Gamma}_j = \hat{\Gamma}'_{-j}$ for $j < 0$. M is the bandwidth parameter, which can act as a truncation lag for kernels such that $k(x) = 0$ for $|x| > 1$.

Consider testing the null hypothesis $H_0 : R\beta = r$ against $H_1 : R\beta \neq r$, where R is a $q \times s$ matrix of rank q and r is a $q \times 1$ vector. We consider the following F-type statistic:

$$F_T = T\left(R\hat{\beta} - r\right)'\left[R\hat{Q}^{-1}\hat{\Omega}\hat{Q}^{-1}R'\right]^{-1}\left(R\hat{\beta} - r\right)/q.$$

In the case where $q = 1$ we can consider $t-$statistics of the form

$$t_T = \frac{\sqrt{T}\left(R\hat{\beta} - r\right)}{\sqrt{R\hat{Q}^{-1}\hat{\Omega}\hat{Q}^{-1}R'}}.$$

Under suitable regularity conditions (described subsequently), $\sqrt{T}\left(R\hat{\beta}-r\right)$ can be approximated by a vector of normal random variables with variance-covariance matrix $RQ^{-1}\Omega Q^{-1}R'$. Given that $p\lim \hat{Q} = Q$, the traditional asymptotic approach seeks to establish consistency of $\hat{\Omega}$ to justify approximating $\hat{\Omega}$ by Ω. Consistency of $\hat{\Omega}$ requires that $M \to \infty$ as $T \to \infty$, but $M/T \to 0$. Under the traditional approach, F_T has a limiting chi-square distribution and t_T has a limiting standard normal distribution.

An alternative approximation for $\hat{\Omega}$ has been proposed by **?**. Suppose the bandwidth is modelled as $M = bT$, with b a fixed constant in $(0, 1]$. Because b is held fixed in this asymptotic nesting of M, this approach has been labelled fixed-b asymptotics. Under fixed-b asymptotics, $\hat{\Omega}$ converges to a random variable (rather than a constant) that depends on the kernel and bandwidth. As a consequence, F_T and t_T have nonstandard distributions. These limiting distributions are useful for testing because they reflect the choice of bandwidth and kernel but are otherwise asymptotically pivotal (i.e. independent of nuisance parameters) and critical values can be tabulated. For example, under suitable regularity conditions (to be described subsequently), **?** showed that

$$\begin{aligned} F_T &\Rightarrow W_q(1)'Q_q(b)^{-1}B_q(1)/q, \\ t_T &\Rightarrow \frac{W_1(1)}{\sqrt{Q_1(b)}}, \end{aligned} \tag{3}$$

where $\Rightarrow$ denotes weak convergence, $W_i(r)$ is an $i \times 1$ vector of independent standard Wiener processes and $Q_i(b)$ is a random matrix that depends on the kernel. For example, in the case of the Bartlett kernel,

$$Q_i(b) = \frac{2}{b}\int_0^1 \widetilde{W}_i(r)\widetilde{W}_i(r)'dr - \frac{1}{b}\int_0^{1-b}\left(\widetilde{W}_i(r+b)\widetilde{W}_i(r)' + \widetilde{W}_i(r)\widetilde{W}_i(r+b)'\right)dr \tag{4}$$

where $\widetilde{W}_i(r) = W_i(r) - rW_i(1)$.

An alternative to asymptotic approximations is the bootstrap. In this paper we focus on the MBB of **?** and **?**. Define the vector $w_t = (y_t, x_t')'$ that collects the dependent and the explanatory variables for each observation. Let $\ell \in \mathbb{N}$ $(1 \leq \ell < T)$ be a block length, and let $B_{t,\ell} = \{w_t, w_{t+1}, \ldots, w_{t+\ell-1}\}$ be the block of ℓ consecutive observations starting at w_t. Note that $\ell = 1$ gives the standard i.i.d. bootstrap. For simplicity take $T = k_0\ell$. The MBB draws $k_0 = T/\ell$ blocks randomly with replacement from the set of overlapping blocks $\{B_{1,\ell}, \ldots, B_{T-\ell+1,\ell}\}$. Let F_T^* and t_T^* denote the naive bootstrap versions of F_T and t_T. F_T^* and t_T^* are computed as follows. Given a bootstrap resample $w_t^* = (y_t^*, x_t^{*\prime})'$, let $\widehat{\beta}^*$ denote the OLS estimate from the regression of y_t^* on x_t^* and let $\widehat{Q}^* = T^{-1}\sum_{t=1}^T x_t^* x_t^{*\prime}$. Let $\widehat{\Omega}^*$ denote the bootstrap version of $\widehat{\Omega}$ where $\hat{v}_t^* = x_t^*\hat{u}_t^* = x_t^*(y_t^* - x_t^{*\prime}\hat{\beta}^*)$ is used in place of $\widehat{v}_t$. The naive bootstrap statistics are defined as

$$F_T^* = T\left(R\hat{\beta}^* - r^*\right)'\left[R\hat{Q}^{*-1}\hat{\Omega}^*\hat{Q}^{*-1}R'\right]^{-1}\left(R\hat{\beta}^* - r^*\right)/q$$

where $r^* = R\widehat{\beta}$, and in the case of $q = 1$,

$$t_T^* = \frac{\sqrt{T}\left(R\hat{\beta}^* - r^*\right)}{\sqrt{R\hat{Q}^{*-1}\hat{\Omega}^*\hat{Q}^{*-1}R'}}.$$

The bootstrap statistics are naive in the sense that, except for the centering around r^* instead of r, they are computed in the same way as F_T and t_T using the resampled data in place of the original data. The empirical distributions of F_T^* and t_T^* can be accurately estimated using simulations.

3 Finite Sample Performance

In this section we use simulations to compare and contrast the finite sample performance of the standard asymptotic approximation, the fixed-b asymptotic approximation and the naive MBB. We first present results for the simple location model followed by results for a stationary regression model with four regressors. The simple location model simulations illustrate interesting patterns in the relative performance of the different asymptotic approximations. Theoretical results that can explain some of these patterns are provided by the higher order asymptotic theory in Section 5. The simulations for the regression model show that the patterns seen in the simple location model continue to hold in the regression context. A theoretical link between the naive MBB and fixed-b asymptotics suggested by both sets of simulations is formally established in Section 4 for regression models.

Consider the simple location model,

$$y_t = \beta_1 + u_t, \tag{5}$$

where

$$u_t = \rho u_{t-1} + \left(1 - \rho^2\right)^{1/2} \varepsilon_t, \tag{6}$$

$\{\varepsilon_t\} \sim i.i.d.N(0,1)$ and $u_1 = 0$. We consider testing the null hypothesis that $\beta_1 \leq 0$ against the alternative that $\beta_1 > 0$ at a nominal level of 5% using

$$t_{\hat{\beta}_1} = \frac{\hat{\beta}_1}{se\left(\hat{\beta}_1\right)},$$

where $se\left(\hat{\beta}_1\right)$ is the HAC standard error estimate. We focus on a one-sided test because the theoretical results developed in Section 5 apply to one-sided tests. The true parameter, β_1, is set equal to zero and we consider three values for the AR parameter ρ: $0.3, 0.5$ and 0.9. In the simulations, 2000 random samples are generated for the sample sizes $T \in \{25, 50\}$. We consider the Bartlett and the QS kernels (results for other kernels are available from the authors) and report results across 25 different values of the bandwidth: $M = 1, 2, \ldots, 25$, for $T = 25$, and

$M = 2, 4, \ldots, 48, 50$, for $T = 50$. We reject the null hypothesis whenever $t_{\hat{\beta}_1} > t_c$, where t_c is a critical value.

The methods differ in the way in which the critical value is calculated. In particular, $t_c = 1.645$ is used for the standard asymptotic approximation, whereas t_c is the 95% percentile of the fixed-b asymptotic distribution derived by **?** for the fixed-b approximation. For the naive MBB t_c is the 95% bootstrap percentile of the studentized bootstrap t-statistic

$$t_{\hat{\beta}_1^*} = \frac{\hat{\beta}_1^* - \hat{\beta}_1}{se\left(\hat{\beta}_1^*\right)},$$

where $se\left(\hat{\beta}_1^*\right)$ is computed using the same formula as $se\left(\hat{\beta}_1\right)$, but replacing the bootstrap data for the real data. The bootstrap tests are based on 999 replications for each sample. We report results for the block lengths $\ell = \{1, 5\}$. When $\ell = 1$ the MBB reduces to the i.i.d. bootstrap.

We also report rejection probabilities using the Edgeworth approximation for t_c derived by **?**. Using t_z to denote the right-tail standard normal critical value, the Edgeworth critical value is given by

$$t_{edge} = t_z + \frac{1}{2}\delta t_z^2 + f(t_z)\frac{M}{T}, \tag{7}$$

where

$$f(x) = \left(2\int_0^1 k(s)^2 ds + \int_0^1 k(s)ds\right)\frac{x}{2} + \left(\int_0^1 k(s)^2 ds\right)\left(\frac{x^3 - 3x}{4}\right),$$

and $\delta = M^{-1}\Omega^{-1}\sum_{j=-\infty}^{\infty} |j|\,\Gamma_j$ for the Bartlett kernel and $\delta = \frac{18}{125}\pi^2 M^{-2}\Omega^{-1}\sum_{j=-\infty}^{\infty} j^2\Gamma_j$ for the QS kernel. Given the $AR(1)$ structure in the simulations, the formulas for δ simplify to $\frac{2\rho}{M(1-\rho^2)}$ and $\frac{36\pi^2\rho}{M^2(1-\rho)^2}$ respectively for the Bartlett and QS kernels[1].

We implement the Edgeworth approximation in two ways. In the first, we make the unrealistic assumptions (from the perspective of practice) that it is known that the errors are $AR(1)$ and that the value of ρ is known. This provides an infeasible benchmark. In the second, we replace Ω with $\widehat{\Omega}$ and we replace $\sum_{j=-\infty}^{\infty} |j|\,\Gamma_j$ and $\sum_{j=-\infty}^{\infty} j^2\Gamma_j$ with estimators $\sum_{j=-(T-1)}^{T-1} k(\frac{j}{M})\,|j|\,\widehat{\Gamma}_j$ and $\sum_{j=-(T-1)}^{T-1} k(\frac{j}{M}) j^2\widehat{\Gamma}_j$ where $k(x)$ and M are the same as used to construct $\widehat{\Omega}$. This feasible approach preserves the nonparametric nature of the test in that we are not assuming any knowledge about the form of the autocovariance structure.

Figures 1 and 2 contain results for the Bartlett kernel (for $T = 25$ and $T = 50$) whereas figures

[1] The regularity conditions used by **?** appear to exclude the Bartlett and Parzen kernels because their spectral windows do not truncate outside the range $[-\pi, \pi]$ although our simulation results suggest that their results likely hold for the Bartlett kernel. **?** conjecture that their proofs could be modified to allow kernels like the Bartlett and Parzen.

3 and 4 contain results for the Quadratic Spectral (QS) kernel. Each figure contains three panels corresponding to the three values of ρ. Each panel depicts the empirical null rejection probabilities as a function of the bandwidth.

Looking at the figures, several interesting patterns emerge. The naive MBB ($NB1$, $NB5$) is (almost) always more accurate than the standard normal asymptotic approximation (AT) and often substantially so. The improvement of the naive MBB over the standard normal approximation holds for both kernels and the improvement is larger for the QS kernel as compared to the Bartlett kernel. For a given sample size, the larger the bandwidth, the larger the improvement. The i.i.d. bootstrap ($NB1$) tends to closely follow the fixed-b asymptotics (KV), across all DGP's, bandwidths, sample sizes and kernels, despite the presence of autocorrelation. This pattern strongly suggests a systematic relationship between the naive block bootstrap and the fixed-b asymptotic approximation. In addition it is interesting to note that as ρ increases (e.g. $\rho = 0.9$), increasing the block size to 5 helps in further reducing the size distortions. This suggests that the naive block bootstrap may offer an asymptotic refinement over the fixed-b asymptotics with careful choice of the block length.

The performance of the Edgeworth approximations are interesting. The most relevant pattern is that neither Edgeworth approximation closely follows the naive MBB suggesting that theoretical explanations for the patterns displayed by the naive MBB will not be found using Edgeworth arguments. This is not surprising given the theoretical arguments made by **?** and **?** about the naive MBB. When the errors are i.i.d. ($\rho = 0$), the infeasible Edgeworth ($EdgeInf$) is more accurate than the standard normal approximation but is less accurate than the naive MBB or fixed-b approximations. The differences become more apparent as the bandwidth increases. The feasible Edgeworth ($EdgeFeas$) gives rejections between the standard normal and the infeasible Edgeworth. When there is serial correlation in the errors ($\rho = 0.3, 0.9$), the feasible Edgeworth consistently is more accurate than the standard normal but is less accurate than the naive MBB or fixed-b. The infeasible Edgeworth behaves much differently in the presence of serial correlation. When the bandwidth is small, the infeasible Edgeworth tends to under-reject, especially when $\rho = 0.9$. When the bandwidth is large and the Bartlett kernel is used, the infeasible Edgeworth has rejections very similar to the naive MBB but it tends to over-reject more than those tests when the QS kernel is used. Overall it is interesting to note that the feasible and infeasible Edgeworth approximations do not seem systematically linked to each other.

To show that many of the patterns in the simple location model continue to hold in a regression setting, we now report results for a stationary regression model with four regressors using the well-studied setup of **?**. We consider the linear regression model

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \beta_5 x_{t5} + u_t$$

where $\{u_t\}$ is given by (6) and

$$x_{ti} = \rho x_{t-1,i} + \left(1-\rho^2\right)^{1/2} v_{ti}, \quad x_{0i} = 0, \qquad i = 2, \ldots, 5;$$

where v_{ti} are generated as i.i.d. $N(0,1)$ errors that are independent of each other and independent with u_t.

We consider testing the null hypothesis that $\beta_2 = 0$ against the alternative that $\beta_2 \neq 0$ at a nominal level of 5%. The test statistic is the t-test

$$t_{\hat{\beta}_2} = \frac{\hat{\beta}_2}{se\left(\hat{\beta}_2\right)},$$

where $se\left(\hat{\beta}_2\right)$ is a HAC standard error estimate. We set $\beta_2 = 0$ and because $t_{\hat{\beta}_2}$ is exactly invariant to the values of the other regression parameters, we set them to zero without loss of generality. We use the same setup as in the simple location model except that we report results for the naive MBB using block lengths $\ell \in \{1, 5, 10, 20\}$. We not report Edgeworth results because formal Edgeworth expansions do not appear available for regression models. We consider a two-sided test in the regression case to show that many of the patterns seen for the one-sided test in the simple location model carry over to two-sided tests.

The results are depicted in Figures 5 - 8. As in the simple location model, we see that the i.i.d. bootstrap ($NB1$) closely follows the rejections of the fixed-b (KV) approximation and both are usually more accurate than the standard normal (AT) approximation. When the serial correlation in the errors is strong ($\rho = 0.9$) increasing the block length to 5 ($NB5$) and 10 ($NB10$) can further increase the accuracy of the approximation. Making the block length too big relative to the sample size as in Figure 7 ($\ell = 20$ and $T = 25$) can result in substantial over-rejections and the MBB can be a worse approximation than the standard normal when the bandwidth is small to medium in size. When $T = 50$, the MBB performs quite well with $\ell = 20$ although not quite as well as when $\ell = 10$. These simulations confirm the patterns seen in the simple location model and further suggest that careful choice of the block length can further improve the approximation provided by fixed-b asymptotics.

The patterns in the simulations in both the simple location and regression models suggest that the fixed-b approximation and the naive MBB are systematically related and that they may provide an improvement over the standard normal approximation. Careful choice of the block length may provide an improvement over the fixed-b approximation. Obtaining theoretical results that explain all of these patterns is a very challenging research program. In the next section we establish the asymptotic equivalence of the fixed-b approximation and the naive MBB in stationary regression models. We then focus on the simple location model and show that the naive MBB theoretically is more accurate than the standard normal approximation when the Bartlett kernel is used. The very difficult question as to whether careful choice of the block length can improve upon the fixed-b approximation is beyond the scope of this paper.

4 Fixed-b Bootstrap Asymptotics

In this section we derive the asymptotic distribution of naive block bootstrap HAC robust tests under the fixed-b asymptotics. In particular, for linear regression models we show that t_T^* and F_T^* have the same limiting fixed-b distribution as t_T and F_T. Define $S_{[rT]} = \sum_{t=1}^{[rT]} v_t$, where $[rT]$ denotes the integer part of rT with $r \in [0,1]$. Let $X_T(r) = T^{-1/2} S_{[rT]}$ be the corresponding partial sum process. Similarly, define $Q_T(r) = T^{-1} \sum_{t=1}^{[rT]} x_t x_t'$. Following **?** and **?** we make the following two high level assumptions:

A1. $X_T(r) \Rightarrow \Lambda W_s(r)$, with $\Omega = \Lambda\Lambda' = \lim_{T\to\infty} Var\left(T^{-1/2}\sum_{t=1}^{T} v_t\right)$.

A2. $\sup_{r\in[0,1]} |Q_T(r) - rQ| \to 0$ in probability.

Here, we assume in addition that the statistic of interest, A_T, is such that:

A3. A_T can be written as

$$A_T = g\left(X_T(r), Q_T(r), D_T(r)\right),$$

where g is a continuous functional of $(X_T(r), Q_T(r), D_T(r))$, and $D_T(r)$ is a vector of deterministic functions of T and r such that $D_T(r) \to D(r)$ as $T \to \infty$, uniformly in r.

Condition A3 is a general way of expressing statistics that includes t_T and F_T. The function $D_T(r)$ reflects the choice of kernel. Using the arguments of **?**, it follows that $\hat{\Omega}$ is a continuous functional of the processes $X_T(r)$, $Q_T(r)$, and $D_T(r)$, where $D_T(r)$ is a function of $k(r)$. If $k''(r)$ exists, then we can show that $\lim_{T\to\infty} D_T(r) = b^{-2}k''(r/b)$, in which case $D(r) = b^{-2}k''(r/b)$. For kernels that truncate to zero for $|x| \geq 1$, $D_T(r)$ is a 2×1 vector and $D(r)$ has elements given by $b^{-2}k''(r/b)$ for $|r| \leq b$ and $b^{-1}k'_-(1)$, where $k'_-(1)$ is the first derivative of $k(x)$ from the left evaluated at $x = 1$. For the Bartlett kernel we have $D_T(r) = (2b^{-1}, -b^{-1})'$ and $D(r) = (2b^{-1}, -b^{-1})'$. Thus, A3 holds for a wide class of kernels including the Bartlett kernel. See **?** for additional details on how $D_T(r)$ is constructed.

Under conditions A1 through A3, an application of the continuous mapping theorem (CMT) implies that as $T \to \infty$,

$$A_T \Rightarrow g\left(\Lambda W_s(r), rQ, D(r)\right) \equiv G.$$

Suppose that the random variable G has a distribution that is pivotal, i.e. invariant to Λ and Q. For example, this is the case for F_T and t_T as indicated by (3). The goal in this section is to provide a set of primitive conditions on $\{x_t\}$ and $\{v_t\}$ under which the naive block bootstrap test,

F_T^*, weakly converges to G, in which case the naive bootstrap and the fixed-b approximation will be equivalent in a first order sense. Note that results for t_T^* follow as an obvious corollary.

We now need to introduce some additional notation. Given a bootstrap resample $w_t^* = (y_t^*, x_t^{*\prime})'$, let $v_{0t}^* = x_t^* (y_t^* - x_t^{*\prime}\beta) \equiv x_t^* u_{0t}^*$, and let $v_t^* = x_t^* \left(y_t^* - x_t^{*\prime}\hat{\beta}\right)' \equiv x_t^* u_t^*$. In order to simplify the notation, we omit T in the definition of the bootstrap variables, e.g., we write v_t^* instead of v_{Tt}^*. Notice that v_t^* (and not v_{0t}^*) is the bootstrap analogue of v_t as it replaces β with $\hat{\beta}$. Let $S_{[rT]}^* = \sum_{t=1}^{[rT]} v_t^*$ and define the bootstrap partial sum process $X_T^*(r) = T^{-1/2} S_{[rT]}^*$. Similarly, define $Q_T^*(r) = T^{-1} \sum_{t=1}^{[rT]} x_t^* x_t^{*\prime}$. As usual in the bootstrap literature, P^* denotes the probability measure induced by the bootstrap resampling, conditional on a realization of the original time series. We use the following notation for the bootstrap asymptotics (see **?** for similar notation and for several useful bootstrap asymptotic properties): Let Z_T^* be a sequence of bootstrap statistics. We write $Z_T^* = o_{P^*}(1)$ in probability, or $Z_T^* \to^{P^*} 0$ in probability, if for any $\varepsilon > 0$, $\delta > 0$, $\lim_{T\to\infty} P\left[P^*\left(|Z_T^*| > \delta\right) > \varepsilon\right] = 0$. Similarly, we write $Z_T^* = O_{P^*}(1)$ in probability if for all $\varepsilon > 0$ there exists a $M_\varepsilon < \infty$ such that $\lim_{T\to\infty} P\left[P^*\left(|Z_T^*| > M_\varepsilon\right) > \varepsilon\right] = 0$. Finally, we write $Z_T^* \Rightarrow^{P^*} Z$ in probability if, conditional on the sample, Z_T^* weakly converges to Z under P^*, for all samples contained in a set with probability converging to one.

Suppose the bootstrap processes $X_T^*(r)$ and $Q_T^*(r)$ satisfy the following assumptions, in probability:

A1*. $X_T^*(r) \Rightarrow^{P^*} \Lambda^* W_s(r)$, for some Λ^*.

A2*. $\sup_{r\in[0,1]} |Q_T^*(r) - rQ^*| \to^{P^*} 0$ for some Q^*.

In this section we study the asymptotic behavior of *naive* bootstrap statistics, i.e., we suppose that

A3*. The bootstrap statistic A_T^* can be written as

$$A_T^* = g\left(X_T^*(r), Q_T^*(r), D_T(r)\right),$$

where g and $D_T(r)$ are as defined in A3.

According to condition A3*, the bootstrap statistic is equal to the exact same function as the original statistic, but replaces the bootstrap data for the real data. This is the sense in which the bootstrap statistic is naive. It is a very straightforward algebraic calculation to show that t_T^* and F_T^* satisfy condition A3*. It is clear that under Assumptions A1*-A3*, by an application of the CMT, we have that

$$A_T^* \Rightarrow^{P^*} g\left(\Lambda^* W_s(r), rQ^*, D(r)\right),$$

in probability. Because the distribution of the random variable $g(\cdot,\cdot,\cdot)$ is pivotal (as in the case of t and F tests), the limiting distribution of A_T^* coincides with the limiting distribution of A_T, independently of Λ^* and Q^*. Thus, the asymptotic equivalence between A_T^* and A_T depends crucially on the conditions A1* and A2*. Next, we provide primitive conditions on $\{x_t\}$ and $\{v_t\}$ that are sufficient for A1* and A2*.

We derive results under the assumptions that $\{x_t\}$ and $\{v_t\}$ are near epoch dependent (NED) on an underlying mixing process $\{\varepsilon_t\}$. NED processes allow for very general forms of dependence and contain mixing processes as a special case. For a general time series $\{w_t\}$, we view each coordinate w_t as a measurable function of the potentially infinite history of another underlying process $\{\varepsilon_t\}$, i.e. $w_t(\ldots,\varepsilon_{t-1},\varepsilon_t,\varepsilon_{t+1},\ldots)$. Let $\mathcal{F}_s^t \equiv \sigma(\varepsilon_s,\ldots,\varepsilon_t)$ for any $s \le t$ be the sigma-field generated by $\varepsilon_s,\ldots,\varepsilon_t$, and let E_s^t denote the expectation conditional on $\mathcal{F}_s^t$. We say $\{w_t\}$ is L_q-NED on $\{\varepsilon_t\}$, $q \geq 1$, if $\|w_t\|_q < \infty$ and $\nu_{k_0} = \sup_t \left\| w_t - E_{t-k_0}^{t+k_0} w_t \right\|_q \to 0$ as $k_0 \to \infty$. Here and in what follows, $\|w\|_q = \left(\sum_i E|w_i|^q\right)^{1/q}$ denotes the L_q-norm of a random vector w. Similarly, we let $|\cdot|$ denote the Euclidean norm of the corresponding vector or matrix. If the NED coefficients ν_{k_0} are such that $\nu_{k_0} = O\left(k_0^{-a-\delta}\right)$ for some $\delta > 0$, we say $\{w_t\}$ is L_q-NED of size $-a$. We assume $\{\varepsilon_t\}$ is strong mixing. The strong mixing coefficients are $\alpha_{k_0} = \sup_m \sup_{\{A \in \mathcal{F}_{-\infty}^m, B \in \mathcal{F}_{m+k_0}^\infty\}} |P(A \cap B) - P(A)P(B)|$; we require $\alpha_{k_0} \to 0$ as $k_0 \to \infty$ suitably fast.

We impose the following assumptions on $\{x_t\}$ and $\{v_t\}$:

Assumption 1

1a. For some $p > 2$, $\|x_t\|_{2p} \le \Delta < \infty$ for all $t = 1, 2, \ldots$.

1b. $\{x_t\}$ is a weakly stationary sequence L_2-NED on $\{\varepsilon_t\}$ with NED coefficients of size $-\frac{2(p-1)}{p-2}$.

1c. $\|v_t\|_p \le \Delta < \infty$, and $E(v_t) = 0$ for all $t = 1, 2, \ldots$.

1d. $\{v_t\}$ is a weakly stationary sequence L_2-NED on $\{\varepsilon_t\}$ with NED coefficients of size $-\frac{1}{2}$.

1e. $\{\varepsilon_t\}$ is an α-mixing sequence with α_{k_0} of size $-\frac{2p}{p-2}$.

1f. $\Omega = \lim_{T\to\infty} Var\left(T^{-1/2}\sum_{t=1}^T v_t\right)$ is positive definite.

We can show that Assumption 1 is a sufficient assumption for the high level conditions A1 and A2. Note that under Assumption 1, $\Omega = \lim_{T\to\infty} Var\left(T^{-1/2}S_T\right)$ exists. We further assume Ω is positive definite, which ensures the existence of a matrix Λ such that $\Omega = \Lambda\Lambda'$. Next, we show that the following strengthened version of Assumption 1 is sufficient to ensure that conditions A1* and A2* hold.

Assumption 1′

1c′. For some $p > 2$ and $\delta > 0$, $\|v_t\|_{p+\delta} \leq \Delta < \infty$, and $E(v_t) = 0$ for all $t = 1, 2, \ldots$.

1d′. $\{v_t\}$ is a weakly stationary sequence $L_{2+\delta}$-NED on $\{\varepsilon_t\}$ with NED coefficients of size -1.

1e′. $\{\varepsilon_t\}$ is an α-mixing sequence of size $-\frac{(2+\delta)(p+\delta)}{p-2}$.

Lemma 4.1 *Under Assumption 1 strengthened by Assumption 1′, it follows that,*

a) *For any fixed ℓ such that $1 \leq \ell < T$, as $T \to \infty$,*

$$X_T^*(r) \Rightarrow^{P^*} \Lambda_\ell W_s(r), \tag{8}$$

in probability, where Λ_ℓ is the square root matrix of $\Omega_\ell \equiv \Gamma_0 + \sum_{j=1}^{\ell} \left(1 - \frac{j}{\ell}\right) \left(\Gamma_j + \Gamma_j'\right)$.

b) *Let $\ell = \ell_T \to \infty$ as $T \to \infty$ such that $\ell^2/T \to 0$. Then*

$$X_T^*(r) \Rightarrow^{P^*} \Lambda W_s(r), \tag{9}$$

in probability, where Λ is the square root matrix of Ω.

c) *Under both sets of assumptions on ℓ, it follows that*

$$\sup_{r\in[0,1]} |Q_T^*(r) - rQ| \to^{P^*} 0,$$

in probability.

Parts a) and b) of Lemma 4.1 provide functional central limit theorems (FCLT) for the bootstrap partial sum process of the bootstrap scores $X_T^*(r) = T^{-1/2} \sum_{t=1}^{[rT]} v_t^*$. To prove these results, we apply a bootstrap FCLT (Lemma A.3 given in the Appendix) for $Z_T^*(r) = T^{-1/2} \sum_{t=1}^{[rT]} (X_t^* - E^*(X_t^*))$, when $\{X_t^*\}$ is a MBB resample of $\{X_t\}$, a NED process on a mixing process. Lemma A.3 is a multivariate extension of an univariate bootstrap FCLT given in **?** for stationary mixing processes to the NED case.

We consider two cases: a) one where ℓ is fixed as $T \to \infty$, and b) another where $\ell \to \infty$ as $T \to \infty$. Note that the first case includes the i.i.d. bootstrap as a special case. According to Lemma 4.1, the bootstrap partial sum process $X_T^*(r)$ weakly converges to a Brownian motion with the "right" covariance matrix Ω only if the block size ℓ increases with the sample size at an appropriate rate. When ℓ is fixed, the limiting covariance matrix is Ω_ℓ, which is different from Ω under general autocorrelation. This reflects the well-known fact that the MBB with fixed block size (and therefore the i.i.d. bootstrap) achieves only partial correction of dependence (cf. **?**).

Our first formal theoretical result is as follows.

Theorem 4.1 *Let $b \in (0, 1]$ be a constant and suppose $M = bT$. Let Assumption 1 strengthened by Assumption 1′ hold, and let $k(x)$ be the Bartlett kernel or let $k(x)$ be such that $k''(x)$ exists and* *is continuous everywhere with the possible exception of* $|x| = 1$. *Suppose the block size* ℓ *is either fixed as* $T \to \infty$, *or* $\ell \to \infty$ *as* $T \to \infty$ *such that* $\ell^2/T \to 0$. *Then, under* $H_0 : R\beta = r$, *as* $T \to \infty$,

$$F_T^* \Rightarrow^{P^*} W_q(1)' Q_q(b)^{-1} W_q(1)/q,$$

in probability, where $Q_q(b)$ *is a random matrix defined in Definition 1 of* **?**.

Theorem 4.1 shows that the naive bootstrap F test statistic has asymptotically the same distribution of F_T derived under the fixed-b asymptotics nesting of **?**. A similar result holds for t_T^*. The first implication of Theorem 4.1 is that a naive bootstrap is as accurate as the new first order fixed-b asymptotics of **?**. The second implication is that a simple i.i.d. bootstrap is asymptotically valid (and equivalent to the fixed-b asymptotics), even in the presence of serial correlation. This result is a consequence of the asymptotic pivotalness of the F statistic.

5 Higher-order results

In this section we prove that the naive i.i.d. bootstrap is capable of providing an asymptotic refinement over the standard normal approximation even for dependent data. We focus on the t-statistic in the simple location model given by (5), i.e. we assume $x_t \equiv 1$ for all t. Here the score vector v_t is equal to the scalar u_t. We derive results for the Bartlett kernel because $\widehat{\Omega}$ can be expressed as a relatively simple function of $X_T(r)$ in this case. We expect our results to naturally extent to other kernels although the details are likely to be very tedious.

Specifically, we show that the error of the naive i.i.d. bootstrap approximation to the finite sample distribution of a fixed-bandwidth HAC based statistic is of order $o\left(T^{-1/2+3/2p+\epsilon}\right)$ for any $\epsilon > 0$, where p is the number of finite moments of u_t. We also show that the error of the fixed-b asymptotic distribution derived by **?** is of the same magnitude as the error of the naive i.i.d. bootstrap. In contrast, the error of the normal approximation to the distribution of a HAC statistic computed with the optimal MSE bandwidth is of order $O\left(T^{-1/3}\right)$ for the Bartlett kernel. Thus, the naive i.i.d. bootstrap and the fixed-b asymptotics provide a smaller error than the normal approximation whenever $p > 9$. The i.i.d. bootstrap is capable of providing an asymptotic refinement over the standard normal approximation even for dependent data because it mimics the fixed-b asymptotic distribution, which itself improves upon the normal approximation.

In this section we assume u_t is a linear process. This is a more restrictive dependence assumption than our previous NED Assumption 1. To prove our results, we will rely on the method of strong approximations (see below for more details on this method), available for linear processes, and this is the main reason why we restrict attention to the special class of linear processes. We are unaware of such results for NED sequences. Thus, we let

$$u_t = \pi(L)\varepsilon_t = \sum_{j=0}^{\infty} \pi_j \varepsilon_{t-j},$$

with $\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j$, and make the following additional assumptions.

Assumption 2

a) ε_t are i.i.d. with $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma^2$ and $E|\varepsilon_t|^p < \infty$ for some $p > 2$.

b) $\pi(z) \neq 0$ for all $|z| \leq 1$ and $\sum_{i=0}^{\infty} |i| |\pi_i| < \infty$.

Under Assumption 2, the FCLT for linear processes (cf. Theorem 3.4, Phillips and Solo, 1992)**?** implies that

$$W_T^0(r) \equiv T^{-1/2} \sum_{t=1}^{[Tr]} u_t \Rightarrow \Lambda W_1(r),$$

where $\Omega = \Lambda^2 \equiv \pi^2(1)\sigma^2$ is the long run variance under Assumption 2. To establish our results we need a result stronger than this invariance principle. In particular, we need specific rates of convergence of the partial sum process for its limiting process. This can be achieved through the method of strong approximations. Recently, **?** uses strong approximations to show that the bootstrap provides an asymptotic refinement for unit root tests. Similarly, **?** relies on this method to show asymptotic refinements of the bootstrap in the context of weakly integrated processes. Our methods of proof will closely follow those of **?**.

Consider the following probabilistic embedding of the partial sum process of u_t:

$$W_T(r) =_d T^{-1/2} \sum_{t=1}^{[Tr]} u_t,$$

where $=_d$ denotes equality in distribution. W_T is a Brownian motion in $D[0,1]$ having the same distribution as W_T^0. In what follows, we will not make a distinction between W_T and its distributionally equivalent copy W_T^0. Therefore we will interpret the distributional equality $=_d$ as the usual equality. The Skohorod representation theorem guarantees that there exists a probability space $(\Omega, \mathcal{F}, P)$ supporting W_T and W_1 such that $W_T \to \Lambda W_1$ a.s. uniformly in $[0,1]$. Moreover, we can state the following result, which follows from a strong approximation result due to **?** (Theorem 3, p. 74).

Lemma 5.1 *Under Assumption 2, we have that*

a) $\sup_{r \in [0,1]} |W_T(r) - \Lambda W_1(r)| = O_P\left(T^{-1/2+1/p}\right)$.

b) *For any* $\epsilon > 0$, $P\left(\sup_{r\in[0,1]} |W_T(r) - \Lambda W_1(r)| \geq T^{-1/2+3/2p}\right) = o\left(T^{-\frac{1}{2}+\frac{3}{2p}+\epsilon}\right)$.

Part a) of Lemma 5.1 shows that the stochastic order of $\sup_{r\in[0,1]} |W_T(r) - \Lambda W_1(r)|$ is equal to $O_P\left(T^{-1/2+1/p}\right)$. As we will show next, the t-statistic can be written as a functional of $W_T(r)$ (or of its distributionally equivalent copy $T^{-1/2}\sum_{t=1}^{[Tr]} u_t$). Thus, we can use part a) to determine the stochastic order of the error term in the stochastic expansion of the t-statistic. Part b) shows that W_T can be approximated by ΛW_1 with an error that is *distributionally*[2] of order $O\left(T^{-1/2+3/2p}\right)$. Thus, although the approximation error of W_T with ΛW_1 is of order $O_P\left(T^{-1/2+1/p}\right)$, its effect is distributionally of a larger order of magnitude, namely $O\left(T^{-1/2+3/2p}\right)$. We will rely on this result to derive the error of the fixed-b asymptotic approximation.

5.1 Asymptotic expansion of the t-statistic

We first provide an asymptotic expansion for the t-statistic. The t-statistic can be written as follows:

$$t_{\hat{\beta}_1} = \frac{\sqrt{T}\left(\hat{\beta}_1 - \beta_1\right)}{\sqrt{\hat{\Omega}}},$$

where $\widehat{\beta}_1 = \overline{y}$ and

$$\hat{\Omega} = \hat{\Gamma}_0 + 2\sum_{j=1}^{M}\left(1 - \frac{j}{M}\right)\hat{\Gamma}_j, \text{ with } \hat{\Gamma}_j = T^{-1}\sum_{t=j+1}^{T}\hat{u}_t\hat{u}_{t-j}.$$

Thus, $\hat{\Omega}$ is the Bartlett kernel variance estimator of $\Omega = \lim_{T\to\infty} Var\left(T^{-1/2}\sum_{t=1}^{T} u_t\right) = \sigma^2\pi^2(1)$. The bandwidth is equal to $M = bT$, where b is a fixed constant. Following **?**, we can write

$$\hat{\Omega} = 2b^{-1}T^{-2}\sum_{t=1}^{T-1}\hat{S}_t^2 - 2b^{-1}T^{-2}\sum_{t=1}^{T-[bT]-1}\hat{S}_t\hat{S}_{t+[bT]},$$

where $\hat{S}_t = \sum_{i=1}^{t}\hat{u}_i$ and $\hat{S}_t = S_t - \left(\frac{t}{T}\right)S_T$, with $S_t = \sum_{i=1}^{t} u_i$.

Lemma 5.2 *Under Assumption 2, and for any fixed* $b \in (0,1]$, *we have*

$$\hat{\Omega} = \Omega Q_1(b) + O_P\left(T^{-1/2+1/p}\right),$$

with $Q_1(b)$ *given by (4).*

Lemma 5.2 provides an asymptotic expansion for $\hat{\Omega}$ with remainder $O_P\left(T^{-1/2+1/p}\right)$. The leading term of this expansion is the distribution derived by Kiefer and Vogelsang (2005). The rate of convergence of $\hat{\Omega}$ increases with p, the number of finite moments of ε. If all moments of ε exist, we

[2] We follow Park (2003) and say that a random sequence R_T is distributionally of order $o\left(T^{-a+\epsilon}\right) = O\left(T^{-a}\right)$ if $P\left(|R_T| > T^{-a}\right) = O\left(T^{-a}\right)$ for some $\epsilon > 0$.

can set $p = \infty$ and get the parametric convergence rate of $O_P\left(T^{-1/2}\right)$. Our next result provides the asymptotic expansion for the t-statistic.

Theorem 5.1 *Under Assumption 2, and for any fixed $b \in (0,1]$, we have*

$$t_{\hat{\beta}_1} = \frac{W_1(1)}{\sqrt{Q_1(b)}} + O_P\left(T^{-1/2+1/p}\right),$$

where $t_{\hat{\beta}_1}$ and $Q_1(b)$ are defined as above.

The leading term of the expansion for $t_{\hat{\beta}_1}$ is the fixed-b first-order asymptotic distribution derived by **?**. Using Lemma 5.1. b) and following Park (2003, Corollary 3.8) we can prove the following corollary to Theorem 5.1.

Corollary 5.1 *Under Assumption 2, and for any fixed $b \in (0,1]$, we have*

$$P\left(t_{\hat{\beta}_1} \leq x\right) = P\left(\frac{W_1(1)}{\sqrt{Q_1(b)}} \leq x\right) + o\left(T^{-1/2+3/2p+\epsilon}\right),$$

uniformly in $x \in \mathbb{R}$, for any $\epsilon > 0$.

Corollary 5.1 gives the rate at which the fixed-b asymptotic approximation converges to the true sampling distribution of $t_{\hat{\beta}_1}$. When all moments exist (as in the Gaussian case considered in our simulations), $p = \infty$, and the error of the fixed-b asymptotic approximation is of order $o\left(T^{-1/2+\epsilon}\right)$ for any $\epsilon > 0$. For Gaussian stationary time series, **?** show that the error made by the asymptotic normal approximation of a t-statistic studentized with a HAC estimator with bandwidth M is of order $O\left(\left(\frac{M}{T}\right)^{1/2}\right)$ (cf. their equation (6)) when MSE optimal bandwidths are used. Thus, when $p = \infty$, the error of the normal approximation is larger than the $o\left(T^{-1/2+\epsilon}\right)$ associated with the fixed-b asymptotics. This explains why the fixed-b asymptotics outperforms the normal approximation in our Monte Carlo simulations (where errors are Gaussian and therefore $p = \infty$). If we set $M = const.T^{1/3}$ (the optimal rate of M for the BT kernel), the error incurred by the first-order standard normal approximation is $O\left(T^{-1/3}\right) = o\left(T^{-1/3+\epsilon}\right)$ for any $\epsilon > 0$, which is larger than the $o\left(T^{-1/2+3/2p+\epsilon}\right)$ error incurred by the fixed-b asymptotics for $p > 9$. Thus, whenever $p > 9$, the fixed-b asymptotics outperforms the standard normal approximation when the bandwidth is set to $M = const.T^{1/3}$.

We should point out that stronger results than Corollary 5.1 have been obtained in some recent work if it assumed that u_t is Gaussian. **?** has established that the error of the fixed-b asymptotic approximation is $O\left(\frac{\log T}{T}\right)$ for the case of the Bartlett kernel with $b = 1$. This result has been refined to $O(T^{-1})$ and extended to a general class of kernels and wider range of b by **?**. While these error rate results are stronger than ours, it is not known whether they continue to hold without the Gaussian assumption. Because the Gaussian assumption cannot hold for the bootstrap, the methods of proof used by **?** and **?** cannot be directly applied to the bootstrap.

5.2 Asymptotic expansion of the naive i.i.d. bootstrap t-statistic

Next, we provide an asymptotic expansion for the naive i.i.d. bootstrap statistic. Let

$$u_t^* \sim \text{ i.i.d. } \{\hat{u}_t = y_t - \bar{y} : t = 1, \cdots, T\}$$

be an i.i.d. bootstrap sample. Note that $u_t^* = y_t^* - \bar{y}$, where y_t^* is an i.i.d. bootstrap observation drawn from $\{y_t\}$. The naive i.i.d. bootstrap $t-$statistic is defined as

$$t_{\hat{\beta}_1^*} = \frac{\sqrt{T}\left(\hat{\beta}_1^* - \hat{\beta}_1\right)}{\sqrt{\hat{\Omega}^*}},$$

where $\hat{\beta}_1^* = \overline{y^*}$ and $\hat{\Omega}^*$ is of the same form as $\hat{\Omega}$ but evaluated with the bootstrap data:

$$\hat{\Omega}^* = 2b^{-1}T^{-2}\sum_{t=1}^{T-1}\hat{S}_t^{*2} - 2b^{-1}T^{-2}\sum_{t=1}^{T-[bT]-1}\hat{S}_t^*\hat{S}_{t+[bT]}^*,$$

where $\hat{S}_t^* = S_t^* - \left(\frac{t}{T}\right)S_T^*, \quad S_t^* \equiv \sum_{i=1}^t u_i^*$.

Let $\Omega_T^* = Var^*\left(T^{-1/2}\sum_{t=1}^T u_t^*\right)$. We can show that $\Omega_T^* = T^{-1}\sum_{t=1}^T Var^*\left(u_t^*\right) = T^{-1}\sum_{t=1}^T \hat{u}_t^2$, and

$$\Omega^* \equiv p\lim \Omega_T^* = E\left(u_t^2\right) = \sigma^2\sum_{i=1}^{\infty}\pi_i^2 \neq \sigma^2\pi^2\left(1\right) = \Omega,$$

so the i.i.d. bootstrap does not consistently estimate the long run variance of $\hat{\beta}_1$. However, we will show that the i.i.d. bootstrap can still provide an asymptotic refinement over the standard normal approximation.

By a bootstrap FCLT,

$$W_T^{0*}\left(r\right) = T^{-1/2}\sum_{t=1}^{[Tr]}u_t^* \Rightarrow^{d_*} \Omega^{*1/2}W_1\left(r\right),$$

in probability, where W_1 denotes a standard Brownian motion independent of the realization of u_t. As above, we can find a Brownian motion W_T^* that has the same distribution as W_T^{0*}, conditional on the original sample, and such that the following result follows. We write

$$W_T^*\left(r\right) = T^{-1/2}\sum_{t=1}^{[Tr]}u_t^*,$$

in probability, where the equality is to be interpreted as an equality in distribution under the bootstrap measure. The following result is a strong approximation for the bootstrap partial sum process.

Lemma 5.3 *Under Assumption 2, we have*

a) $\sup_{r\in[0,1]} \left| W_T^*(r) - \Omega^{*1/2} W_1(r) \right| = O_{P^*}\left(T^{-1/2+1/p}\right)$, *in probability.*

b) *For any* $\epsilon > 0$, $P^*\left(\sup_{r\in[0,1]} \left| W_T^*(r) - \Omega^{*1/2} W_1(r) \right| \geq T^{-1/2+3/2p}\right) = o_P\left(T^{-1/2+3/2p+\epsilon}\right)$.

The next result gives an expansion for $\hat{\Omega}^*$ and is the bootstrap analogue of Lemma 5.2.

Lemma 5.4 *Under Assumption 2, we have*

$$\hat{\Omega}^* = \Omega^* Q_1(b) + O_{P^*}\left(T^{-1/2+1/p}\right) \ ,$$

in probability, where $Q_1(b)$ *is as defined previously.*

Given Lemma 5.4, we can derive the following asymptotic expansion for the naive i.i.d. bootstrap t-statistic.

Theorem 5.2 *Under Assumption 2, we have*

$$t_{\hat{\beta}_1^*} = \frac{W_1(1)}{(Q_1(b))^{1/2}} + O_{P^*}\left(T^{-1/2+1/p}\right),$$

in probability.

The following corollary to Theorem 5.2 shows that the effect of the remainder term in the asymptotic expansion of $t_{\hat{\beta}_1^*}$ is distributionally of order $O\left(T^{-1/2+3/2p}\right)$.

Corollary 5.2 *Under Assumption 2, we have*

$$P^*\left(t_{\hat{\beta}_1^*} \leq x\right) = P\left(\frac{W(1)}{(Q(b))^{1/2}} \leq x\right) + o_P\left(T^{-1/2+3/2p+\epsilon}\right),$$

uniformly in $x \in \mathbb{R}$, *for any* $\epsilon > 0$.

It then follows from Corollaries 5.1 and 5.2 that

$$\sup_{x\in\mathbb{R}} \left| P^*\left(t_{\hat{\beta}_1^*} \leq x\right) - P\left(t_{\hat{\beta}_1} \leq x\right) \right| = o_P\left(T^{-1/2+3/2p+\epsilon}\right), \qquad (10)$$

uniformly in $x \in \mathbb{R}$, for any $\epsilon > 0$.

The result in (10) shows that the i.i.d. bootstrap error is of the same order of magnitude as the error implied by the fixed-b asymptotic approximation. In particular, if $p = \infty$ the i.i.d. bootstrap error is arbitrarily close to $o_P\left(T^{-1/2+\epsilon}\right)$, smaller than the error implied by the normal approximation. The i.i.d. bootstrap error is smaller than the error associated with the normal approximation when the optimal bandwidth is used to compute the HAC Bartlett kernel estimator whenever $p > 9$. The reason why the i.i.d. bootstrap provides a refinement in this context is that it replicates the the fixed-b distribution. This is true even when the data are dependent, as we showed more generally before.

6 Heuristic Comparisons of Edgeworth and Fixed-b

While rigorous comparisons of the Edgeworth approximations with fixed-b approximations are well beyond the scope of this paper, some heuristic comparisons can be instructive for guiding future work. In deriving formal Edgeworth approximations, **?** approximate the bias and variance of the HAC variance estimator under the traditional assumption that M/T shrinks to zero. In the simple location model we have from **?** for the QS kernel

$$\begin{aligned} bias(\frac{\widehat{\Omega}}{\Omega}) &= E(\frac{\widehat{\Omega}}{\Omega}) - 1 \approx -\frac{18}{125}\pi^2 M^{-2}\Omega^{-1} \sum_{j=-\infty}^{\infty} j^2\Gamma_j - \frac{M}{T}\int_0^1 k(s)ds \\ &= -\frac{18}{125}\pi^2 M^{-2}\Omega^{-1} \sum_{j=-\infty}^{\infty} j^2\Gamma_j - \frac{M}{T}\frac{5}{4}, \\ Var(\frac{\widehat{\Omega}}{\Omega}) &\approx \frac{M}{T}2\int_0^1 k(s)^2 ds = \frac{M}{T}2. \end{aligned}$$

Although the Bartlett kernel does not satisfy the assumptions used by **?**, existing results in the spectral analysis literature give for the Bartlett kernel

$$\begin{aligned} bias(\frac{\widehat{\Omega}}{\Omega}) &\approx -M^{-1}\Omega^{-1} \sum_{j=-\infty}^{\infty} j\Gamma_j - \frac{M}{T}\int_0^1 k(s)ds = -M^{-1}\Omega^{-1} \sum_{j=-\infty}^{\infty} j\Gamma_j - \frac{M}{T}, \\ Var(\frac{\widehat{\Omega}}{\Omega}) &\approx \frac{M}{T}2\int_0^1 k(s)^2 ds = \frac{M}{T}\frac{4}{3}. \end{aligned}$$

Notice that the Edgeworth approximation (7) is a function of these moments.

Alternatively, the fixed-b approximation approximates the entire distribution of $\widehat{\Omega}$:

$$\frac{\widehat{\Omega}}{\Omega} \approx Q_1(b).$$

The bias of $\widehat{\Omega}/\Omega$ can be approximated by $E(Q_1(b) - 1)$. It is interesting to compare $E(Q_1(b) - 1)$ and $Var(Q_1(b))$ with the traditional bias and variance formulas. For the Bartlett kernel (see **?**)

$$\begin{aligned} E(Q_1(b) - 1) &= -b + \frac{1}{3}b^2 \\ Var(Q_1(b)) &= \frac{4}{3}b - \frac{7}{3}b^2 + \frac{14}{15}b^3 + \frac{2}{9}b^4 \text{ for } b \leq \frac{1}{2}. \end{aligned}$$

Recalling that $b = M/T$ we see that the moments of the fixed-b asymptotic distribution match the traditional moments for terms of order M/T. The differences between the two approximations are that the fixed-b approximation does not include the $-M^{-1}\Omega^{-1}\sum_{j=-\infty}^{\infty} j\Gamma_j$ bias term because it is $o(1)$ under fixed-b asymptotics, but the fixed-b approximation includes terms of order $(M/T)^2$ and higher.

This heuristic observations can shed some light on some of the patterns observed in Figures 1-4. In the case of i.i.d. errors, the bias term of order M^{-1} is exactly zero (because $\sum_{j=-\infty}^{\infty} |j| \Gamma_j = \sum_{j=-\infty}^{\infty} j^2 \Gamma_j = 0$) and the main difference between the fixed-b approximation and the Edgeworth approximation are the higher order terms in the bias and variance formulas for the fixed-b approximation. When M is small, the difference between the Edgeworth and fixed-b approximations are negligible whereas for large M, the fixed-b approximation is slightly more accurate. When $\rho = 0.3$, the Edgeworth approximation is more accurate when M is small because it picks up the M^{-1} term whereas for larger M, the fixed-b approximation is more accurate. These differences become more apparent when $\rho = 0.9$.

An intriguing possibility is apparent. Because of the asymptotic equivalence between fixed-b asymptotics and the naive i.i.d. bootstrap, it appears the i.i.d. bootstrap captures the influence of the bias and variance of $\widehat{\Omega}$ to higher orders than the Edgeworth approximation with respect to terms that depend on powers of M/T, but the i.i.d. bootstrap does not capture the bias term that depends on $\sum_{j=-\infty}^{\infty} |j| \Gamma_j$ or $\sum_{j=-\infty}^{\infty} j^2 \Gamma_j$. With careful choice of block length, the naive block bootstrap could capture these bias term while continuing to capture the M/T and higher order terms terms. The simulations reported in Figures 1-4 show that when there is serial correlation in the data, increasing the block length from 1 to 5 does improve the approximation. It is possible this improvement is coming at least in part through the first bias term.

7 Conclusion

In this paper, we theoretically analyze the performance of the naive MBB applied to HAC robust tests based on nonparametric kernel estimators of the long run variance. In simulations reported here and in **?** it was found that the naive MBB outperforms the standard normal approximation in finite samples. This improvement holds for many kernels, including the Bartlett kernel, and holds even for the i.i.d. bootstrap, despite the dependence in the data. These simulations suggest that the performance of the naive MBB is tightly linked to the finite sample performance of the recently developed fixed-b (i.e. fixed bandwidth) asymptotics. We provide a theoretical explanation for this result: we prove that the bootstrap distribution of the naive MBB is asymptotically the same as the fixed-b asymptotic distribution. In addition, for a simple location model we show that a naive i.i.d. bootstrap can reduce the magnitude of the error in estimating one-sided distribution functions of robust $t-$ statistics compared to the standard normal approximation error for statistics studentized with a Bartlett kernel variance estimator based on optimal MSE bandwidths. Our simulations also suggest that the naive MBB can be more accurate than the fixed-b asymptotic approximation when the block size is chosen appropriately. Providing a theoretical explanation for this finding is a challenging topic of for future research.

Appendix A

This Appendix contains the proofs of the results in Section 4. Throughout this Appendix K denotes a generic constant that may change from one usage to the next. We first state four lemmas that are auxiliary in proving Lemma 4.1 and Theorem 4.1 in Section 4. We then provide the proofs of our main results followed by the proofs of the auxiliary lemmas.

The following result is a maximal inequality for mixingales (see e.g. **?** for a definition of mixingale) due to **??**Hansen (1991, 1992). Zero mean NED processes on a mixing process are mixingales and we will repeatedly use this result in our proofs.

Lemma A.1 *For some nondecreasing sequence of σ-fields $\{\mathcal{F}^t\}$ and for some $p > 1$, let $\{X_t, \mathcal{F}^t\}$ be an L_p-mixingale with mixingale coefficients ψ_m and mixingale constants c_t. Then, letting $S_j = \sum_{t=1}^{j} X_t$ and $\Psi = \sum_{m=1}^{\infty} \psi_m$ it follows that*

a) *If $1 < p \leq 2$, $\left\|\max_{j\leq T} |S_j|\right\|_p \leq K\Psi \left(\sum_{t=1}^{T} c_t^p\right)^{1/p}$.*

b) *For $p \geq 2$, $\left\|\max_{j\leq T} |S_j|\right\|_p \leq K\Psi \left(\sum_{t=1}^{T} c_t^2\right)^{1/2}$.*

The following result gives the probability limits of the MBB variance of a scaled bootstrap sample mean under two different assumptions on the block size ℓ: (a) when ℓ is fixed as $T \to \infty$; and (b) when $\ell \to \infty$ as $T \to \infty$ at an appropriate rate. We state the result for a general time series $\{X_t\}$ satisfying the following assumptions:

Assumption A Let $\{X_t\}$ be a weakly stationary sequence of $s \times 1$ random vectors such that the following hold:

(i) For some $p > 2$, $\|X_t\|_p \leq \Delta < \infty$ for all $t = 1, 2, \ldots$.

(ii) $\{X_t\}$ is L_2-NED on $\{V_t\}$ with NED coefficients of size $-1/2$.

(iii) $\{V_t\}$ is α-mixing of size $-\frac{p}{p-2}$.

Let $\{X_t^* : t = 1, 2, \ldots, T\}$ denote a MBB resample obtained from $\{X_t : t = 1, 2, \ldots, T\}$ using block size ℓ. Let $\Omega_T^* = Var^* \left(T^{-1/2} \sum_{t=1}^{T} X_t^*\right)$ denote the bootstrap variance of $\sqrt{T}\bar{X}_T^*$.

Lemma A.2 *Suppose $\{X_t\}$ satisfies Assumption A. Then,*

a) *For any fixed ℓ such that $1 \leq \ell < T$, as $T \to \infty$,*

$$p \lim_{T\to\infty} \Omega_T^* = \Gamma_0 + \sum_{j=1}^{\ell} \left(1 - \frac{j}{\ell}\right) \left(\Gamma_j + \Gamma_j'\right) \equiv \Omega_\ell,$$

where $\Gamma_j = E\left(\left(X_t - \mu\right)\left(X_{t-j} - \mu\right)'\right)$, $\mu = E\left(X_t\right)$.

b) *Let $\ell = \ell_T \to \infty$ as $T \to \infty$ such that $\ell^2/T \to 0$. Then*

$$p \lim_{T\to\infty} \Omega_T^* = \Gamma_0 + \sum_{j=1}^{\infty} \left(\Gamma_j + \Gamma_j'\right) \equiv \Omega.$$

Our next result establishes a FCLT for the bootstrap partial sum process $Z_T^*(r) = T^{-1/2} \sum_{t=1}^{[rT]} (X_t^* - E^*(X_t^*))$. We need to strengthen Assumption A as follows.

Assumption A′ Let $\{X_t\}$ be a weakly stationary sequence of $s \times 1$ random vectors such that the following hold:

a) For some $p > 2$ and some $\delta > 0$, $\|X_t\|_{p+\delta} \leq \Delta < \infty$ for all $t = 1, 2, \ldots$.

b) $\{X_t\}$ is $L_{2+\delta}$-NED on $\{V_t\}$ of size -1.

c) $\{V_t\}$ is α-mixing of size $-\frac{(2+\delta)(p+\delta)}{p-2}$.

Lemma A.3 *Suppose Assumption A′ holds and let Ω_ℓ and Ω as defined in Lemma A.2 be positive definite matrices. It follows that*

a) *For any fixed ℓ such that $1 \leq \ell < T$, as $T \to \infty$,*

$$Z_T^*(r) \Rightarrow^{P^*} \Lambda_\ell W_s(r), \tag{11}$$

in probability, where Λ_ℓ is the square root matrix of Ω_ℓ.

b) *Let $\ell = \ell_T \to \infty$ as $T \to \infty$ such that $\ell^2/T \to 0$. Then*

$$Z_T^*(r) \Rightarrow^{P^*} \Lambda W_s(r), \tag{12}$$

in probability, where Λ is the square root matrix of Ω.

The following result will be used in the proof of Lemma 4.1.

Lemma A.4 *Suppose $\{X_t - E(X_t)\}$ is a weakly stationary L_2-mixingale with $\|X_t\|_p \leq \Delta < \infty$ for some $p > 2$ such that its mixingale coefficients ψ_m satisfy $\sum_{m=1}^{\infty} \psi_m < \infty$ and its mixingale constants are uniformly bounded. Let $\{X_t^* : t = 1, \ldots, T\}$ denote a MBB resample of $\{X_t : t = 1, \ldots, T\}$ with block size ℓ satisfying either of the two following conditions: a) ℓ is fixed as $T \to \infty$, or b) $\ell \to \infty$ as $T \to \infty$ with $\ell = o(T)$. Then, for any $\eta > 0$, as $T \to \infty$,*

$$P^*\left(\sup_{0\leq r\leq 1} \left| T^{-1} \sum_{t=1}^{[rT]} (X_t^* - E^*(X_t^*)) \right| > \eta\right) = o_P(1).$$

Proof of Lemma 4.1. We start with the proof of a) and b), which can be treated simultaneously. Given our definitions of v_{0t}^* and v_t^*, we can write $v_t^* = v_{0t}^* - x_t^* x_t^{*\prime}\left(\hat{\beta} - \beta\right)$, which implies that

$$\begin{aligned} X_T^*(r) &= T^{-1/2}\sum_{t=1}^{[rT]}\left(v_{0t}^* - E^* v_{0t}^*\right) + T^{-1/2}\sum_{t=1}^{[rT]} E^*\left(v_{0t}^*\right) - T^{-1/2}\sum_{t=1}^{[rT]} x_t^* x_t^{*\prime}\left(\hat{\beta} - \beta\right) \\ &\equiv Z_T^*(r) + A_{1T}^*(r) - A_{2T}^*(r). \end{aligned}$$

An application of Lemma A.3 implies that under Assumption 1′, $Z_T^*(r) \Rightarrow^{P^*} \Lambda^* W_s(r)$, in probability, where Λ^* is the square root matrix of $\Omega^* = p\lim Var^*\left(T^{-1/2}\sum_{t=1}^T v_{0t}^*\right)$. In particular, by Lemma A.2, $\Omega^* = \Omega_\ell$ in a) and $\Omega^* = \Omega$ in b). Thus, to prove that $X_T^*(r) \Rightarrow^{P^*} \Lambda^* W_s(r)$, in probability, it suffices to show that $\sup_r \left|A_{1T}^*(r) - A_{2T}^*(r)\right| = o_{P^*}(1)$ in probability. Adding and subtracting $T^{-1/2}\sum_{t=1}^{[rT]} E^*\left(x_t^* x_t^{*\prime}\right)\left(\hat{\beta} - \beta\right)$, and rearranging terms yields

$$\begin{aligned} A_{1T}^*(r) - A_{2T}^*(r) &= T^{-1/2}\sum_{t=1}^{[rT]} E^*\left(x_t^*\left(y_t^* - x_t^{*\prime}\hat{\beta}\right)\right) - T^{-1/2}\sum_{t=1}^{[rT]}\left(x_t^* x_t^{*\prime} - E^* x_t^* x_t^{*\prime}\right)\left(\hat{\beta} - \beta\right) \\ &\equiv B_{1T}^*(r) - B_{2T}^*(r). \end{aligned}$$

Next, we show that $\sup_{r\in[0,1]}\left|B_{1T}^*(r)\right| = o_{P^*}(1)$ and $\sup_{r\in[0,1]}\left|B_{2T}^*(r)\right| = o_{P^*}(1)$, in probability. Note that

$$\begin{aligned} B_{1T}^*(r) &= E^*\left(T^{-1/2}\sum_{m=1}^{M_r}\sum_{s=1}^{B}\hat{v}_{I_m+s}\right) = E^*\left(T^{-1/2}\sum_{m=1}^{M_r}\sum_{s=1}^{\ell}\hat{v}_{I_m+s}\right) - T^{-1/2}E^*\left(\sum_{s=B+1}^{\ell}\hat{v}_{I_{M_r}+s}\right) \\ &\equiv b_{1T}^*(r) - b_{2T}^*(r), \end{aligned}$$

where M_r, B and I_m are as defined in the proof of Lemma A.3. We can show that $\sup_{r\in[0,1]}\left|b_{2T}^*(r)\right| = O_{P^*}\left(k_0^{-1/2}\right)$ in probability, given in particular the fact $\hat{\beta} - \beta = O_P\left(T^{-1/2}\right)$. Moreover,

$$b_{1T}^*(r) = M_r T^{-1/2} E^*\left(\sum_{s=1}^{\ell}\hat{v}_{I_1+s}\right) = M_r\frac{\ell}{T^{1/2}}E^*\left(\hat{U}_1^*\right),$$

where $\hat{U}_1^* = \ell^{-1}\sum_{s=1}^{\ell}\hat{v}_{I_1+s}$. Defining $\bar{X}_T \equiv T^{-1}\sum_{t=1}^T X_t$ for any random variables X_t, it is well known that $E^*\left(\overline{\hat{v}^*}_T\right) = E^*\left(\hat{U}_1^*\right) = \overline{\hat{v}}_T + O_P\left(\frac{\ell}{T}\right)$ (see e.g. **?**). Since $\overline{\hat{v}_T} = 0$ by the FOC defining the OLS estimator, it follows that $E^*\left(\hat{U}_1^*\right) = 0 + O_P\left(\frac{\ell}{T}\right)$, and noting that $\sup_{r\in[0,1]}\left|M_r\right| = k_0$, we have

$$\sup_{r\in[0,1]}\left|b_{1T}^*(r)\right| = \sup_{r\in[0,1]}\left|M_r\right|\frac{\ell}{T^{1/2}}E^*\left(\hat{U}_1^*\right) = k_0\frac{\ell}{T^{1/2}}\left(0 + O_P\left(\frac{\ell}{T}\right)\right) = O_P\left(\frac{\ell}{\sqrt{T}}\right) = o_P(1),$$

under both conditions a) and b). For $B_{2T}^{*}(r)$, note that

$$\sup_{r\in[0,1]}\left|B_{2T}^{*}(r)\right| = \sup_{r\in[0,1]}\left|T^{-1}\sum_{t=1}^{[rT]}\left(x_t^{*}x_t^{*\prime}-E^{*}x_t^{*}x_t^{*\prime}\right)\right|\left|\sqrt{T}\left(\hat{\beta}-\beta\right)\right|.$$

Since $\left|\sqrt{T}\left(\hat{\beta}-\beta\right)\right| = O_P(1)$, it suffices to show that $\sup_{r\in[0,1]}\left|T^{-1}\sum_{t=1}^{[rT]}\left(x_t^{*}x_t^{*\prime}-E^{*}x_t^{*}x_t^{*\prime}\right)\right| = o_{P^*}(1)$ in probability. This follows by an application of Lemma A.4 since $z_t = x_t x_t' - E(x_t x_t')$ satisfies the assumptions of this lemma under Assumption 1. To prove c), note we can write

$$\sup_{r\in[0,1]}\left|T^{-1}\sum_{t=1}^{[rT]}x_t^{*}x_t^{*\prime}-rQ\right| \le I_1+I_2+I_3,$$

where $I_1 \equiv \sup_{r\in[0,1]}\left|T^{-1}\sum_{t=1}^{[rT]}\left(x_t^{*}x_t^{*\prime}-E^{*}\left(x_t^{*}x_t^{*\prime}\right)\right)\right|$, $I_2 \equiv \sup_{r\in[0,1]}\left|T^{-1}\sum_{t=1}^{[rT]}\left(E^{*}\left(x_t^{*}x_t^{*\prime}\right)-x_t x_t'\right)\right|$, and $I_3 \equiv \sup_{r\in[0,1]}\left|T^{-1}\sum_{t=1}^{[rT]}x_t x_t' - rQ\right|$. As just proven, $I_1 = o_{P^*}(1)$, in probability, and $I_3 = o_P(1)$ given Assumption 1. Next, we will show that $I_2 = o_P(1)$. Adding and subtracting $\mu \equiv E(x_t x_t')$ yields

$$I_2 \le T^{-1}\sup_{r\in[0,1]}\left|\sum_{t=1}^{[rT]}E^{*}\left(x_t^{*}x_t^{*\prime}-\mu\right)\right| + T^{-1}\sup_{r\in[0,1]}\left|\sum_{t=1}^{[rT]}\left(x_t x_t'-\mu\right)\right| \equiv i_1+i_2.$$

Under Assumption 1.a)-b) and 1.e), we can show that $\{x_t x_t' - \mu\}$ is an L_2-NED of size -1 on $\{\varepsilon_t\}$ (cf. Davidson, 1994, Example 17.17). It then follows by Davidson's (1994) Theorem 17.5 that $\{x_t x_t' - \mu\}$ is an L_2-mixingale of size -1 with uniformly bounded mixingale constants. Thus, $i_2 = O_P\left(T^{-1/2}\right)$ by an application of Lemma A.1. Similarly, we can show that $i_1 = O_P\left(T^{-1/2}\right) + O_P\left(\frac{\ell}{T}\right)$, which is $o_P(1)$ under our assumptions. Indeed, we can write

$$i_1 = k_0^{-1}\sup_{r\in[0,1]}\left|\sum_{m=1}^{M_r}E^{*}\left(U_m^{*}\right)\right| + O_P\left(\frac{\ell^{1/2}}{T}\right),$$

where $U_m^{*} = \ell^{-1}\sum_{s=1}^{\ell}\left(x_{I_m+s}x_{I_m+s}'-\mu\right)$. It follows that

$$k_0^{-1}\sup_{r\in[0,1]}\left|\sum_{m=1}^{M_r}E^{*}\left(U_m^{*}\right)\right| \le k_0^{-1}k_0\left|E^{*}\left(U_1^{*}\right)\right| = \left|T^{-1}\sum_{t=1}^{T}\left(x_t x_t'-\mu\right)\right| + O_P\left(\frac{\ell}{T}\right),$$

because

$$E^{*}\left(U_1^{*}\right) = E^{*}\left(T^{-1}\sum_{t=1}^{T}\left(x_t^{*}x_t^{*}-\mu\right)\right) = T^{-1}\sum_{t=1}^{T}\left(x_t x_t'-\mu\right) + O_P\left(\frac{\ell}{T}\right).$$

This completes the proof because $T^{-1}\sum_{t=1}^{T}\left(x_t x_t'-\mu\right) = O_P\left(T^{-1/2}\right)$.

Proof of Theorem 4.1. The proof follows from Lemma 4.1, using the same arguments as in **?**.

Proof of Lemma A.2. As is well known, the MBB variance Ω_T^{*} is equal to the Bartlett kernel variance estimator of $\sqrt{T}\bar{X}_T$, up to a term of order $O_P\left(\frac{\ell^2}{T}\right)$ (see e.g. Fitzenberger (1997, p. 252)). Note that this term vanishes in probability under both sets of conditions on ℓ. Result b) follows by **?** Theorem 2.1. Result a) follows by an argument similar to **?** and **?**, under our more general dependence Assumption A.

Proof of Lemma A.3. These results are multivariate versions of a univariate FCLT given in **?** (henceforth PP (2003)). Whereas PP (2003) assume a mixing condition on $\{X_t\}$, here we allow for the more general NED condition. Note that we assume throughout that $E(X_t) = 0$ for all t without loss of generality given that $\{X_t\}$ is stationary. Since Assumption A$'$ implies Assumption A, it follows by Lemma A.2 that $p\lim_{T\to\infty}\Omega_T^* = \Omega^*$, where Ω^* is equal to Ω_ℓ in a) and equal to Ω in b). Since by assumption both Ω_ℓ and Ω are positive definite, $\Omega_T^{*-1/2}$ exists in probability for all T sufficiently large. By the functional Cramer-Wold device, it suffices to show that $\lambda'\Omega_T^{*-1/2}Z_T^*(r) \Rightarrow^{P^*} \lambda' W_s(r)$ in probability for any λ such that $\lambda'\lambda = 1$. Following PP (2003), for any $r \in [0,1]$, we can write

$$W_T^*(r) \equiv \lambda'\Omega_T^{*-1/2}Z_T^*(r) = \lambda'\Omega_T^{*-1/2}T^{-1/2}\sum_{m=1}^{M_r}\sum_{s=1}^{B}\left(X_{I_m+s} - E^*\left(X_{I_{m+s}}\right)\right),$$

where $M_r = [([rT]-1)/\ell] + 1$ and $B = \min\{\ell, [rT] - (m-1)\ell\}$. Here, $I_1, \ldots, I_{k_0}$ are i.i.d. uniformly distributed on $\{0, 1, \ldots, T-\ell\}$. Notice that for $r \in [0,1]$, $M_r \in \{1, \ldots, k_0\}$ and $B \in \{1, \ldots, \ell\}$. As in PP (2003), we can write

$$W_T^*(r) = \lambda'\Omega_T^{*-1/2}T^{-1/2}\sum_{m=1}^{M_r}\sum_{s=1}^{\ell}\left(X_{I_m+s} - E^*\left(X_{I_{m+s}}\right)\right)$$

$$-\lambda'\Omega_T^{*-1/2}T^{-1/2}\sum_{s=B+1}^{\ell}\left(X_{I_{M_r}+s} - E^*\left(X_{I_{M_r+s}}\right)\right) \equiv W_{1T}^*(r) - W_{2T}^*(r).$$

The proof consists of two steps: (1) Show that $\sup_{r\in[0,1]}|W_{2T}^*(r)| = O_{P^*}\left(k_0^{-1/2}\right)$ in probability; and (2) Show that $W_{1T}^*(r) \Rightarrow^{P^*} W_1(r)$ in probability.

We start with (1). Since $\Omega_T^* \to \Omega^*$ in probability, and Ω^* is p.d., it follows that $\Omega_T^{*-1/2} = O_P(1)$. Thus, it suffices to show that

$$E^*\left(\sup_{r\in[0,1]}\left|T^{-1/2}\sum_{s=B+1}^{\ell}\left(X_{I_{M_r}+s} - E^* X_{I_{M_r+s}}\right)\right|\right) = O_P\left(k_0^{-1/2}\right), \tag{13}$$

by Markov's inequality. Since $k_0 = T/\ell$, $k_0 \to \infty$ as $T \to \infty$ under both set of conditions on ℓ, which implies that $\sup_r |W_{2T}^*(r)| = o_{P^*}(1)$ in probability. An application of triangle's inequality and Jensen's inequality implies that

$$E^*\left(\sup_{r\in[0,1]}\left|T^{-1/2}\sum_{s=B+1}^{\ell}\left(X_{I_{M_r}+s} - E^* X_{I_{M_r+s}}\right)\right|\right) \le 2E^*\left(\sup_{r\in[0,1]}\left|T^{-1/2}\sum_{s=B+1}^{\ell} X_{I_{M_r}+s}\right|\right). \tag{14}$$

Since $I_{M_r} \sim$ i.i.d. Uniform $\{0, \ldots, T-\ell\}$, we have that

$$E^*\left(\sup_{r\in[0,1]}\left|T^{-1/2}\sum_{s=B+1}^{\ell} X_{I_{M_r}+s}\right|\right) = \frac{1}{T-\ell+1}\sum_{j=0}^{T-\ell}\sup_{r\in[0,1]}\left|T^{-1/2}\sum_{s=B+1}^{\ell} X_{j+s}\right|. \tag{15}$$

By Markov's inequality, (13) follows from $E\left(k_0^{1/2}E^*\left(\sup_{r\in[0,1]}\left|T^{-1/2}\sum_{s=B+1}^{\ell} X_{I_{M_r}+s}\right|\right)\right) = O(1)$. Recall that for $r \in [0,1]$, $B \in \{1, \ldots, \ell\}$. Thus,

$$k_0^{1/2}E\left(\sup_{r\in[0,1]}\left|T^{-1/2}\sum_{s=B+1}^{\ell} X_{j+s}\right|\right) \leq k_0^{1/2}T^{-1/2}E\left(\max_{1\leq i\leq \ell}\left|\sum_{s=j+i}^{j+\ell} X_s\right|\right) \leq k_0^{1/2}T^{-1/2}\left\|\max_{1\leq i\leq \ell}\left|\sum_{s=j+i}^{j+\ell} X_s\right|\right\|_2. \tag{16}$$

We now apply Lemma A.1. Under Assumption A$'$, $\{X_t\}$ is an $L_{2+\delta}$-mixingale (hence an L_2-mixingale) with mixingale coefficients ψ_m of size -1, hence $\Psi = \sum_{m=1}^{\infty}\psi_m < \infty$. In particular, we apply Theorem 17.5 of **?**, with $r = p+\delta$, $p = 2+\delta$, $b = 1$ and $a = \frac{(2+\delta)(p+\delta)}{p-2}$. Under our assumptions, the NED constants d_t can be set equal to 1, which implies that the mixingale constants $c_t \leq \max\left(\|X_t\|_{p+\delta}, 1\right) < \Delta < \infty$ for all t. Thus, $\left\|\max_{1\leq i\leq \ell}\left|\sum_{s=j+i}^{j+\ell} X_s\right|\right\|_2 \leq K\ell^{1/2}$, and from (16) we have that $k_0^{1/2}E\left(\sup_{r\in[0,1]}\left|T^{-1/2}\sum_{s=B+1}^{\ell} X_{j+s}\right|\right) \leq K$ uniformly in j, given that $k_0 = T/\ell$. This completes the proof of (1).

Next we show step (2). As in PP (2003), we consider the asymptotically equivalent statistic

$$k_0^{-1/2}\sum_{m=1}^{[rk_0]+1}\lambda'\Omega_T^{*-1/2}\left(\ell^{-1/2}\sum_{s=1}^{\ell}\left(X_{I_{m+s}} - E^*X_{I_{m+s}}\right)\right) \equiv k_0^{-1/2}\sum_{m=1}^{[rk_0]+1} V_m^*,$$

where $V_m^* = \lambda'\Omega_T^{*-1/2}U_m^*$, with $U_m^* = \ell^{-1/2}\sum_{s=1}^{\ell}\left(X_{I_{m+s}} - E^*X_{I_{m+s}}\right)$. Note that $\{V_m^* : m = 1, \ldots, [rk_0]+1\}$ is an array of independent variables with $E^*\left(V_m^*\right) = 0$ and $Var^*\left(V_m^*\right) = \lambda'\Omega_T^{*-1/2}Var^*\left(U_m^*\right)\Omega_T^{*1/2}\lambda = 1$, where the last equality holds because we can show that $Var^*\left(U_m^*\right) = \Omega_T^*$. We now apply a FCLT for martingale difference arrays (cf. Billingsley, 1968, p. 194)**?**. In particular, let $\xi_{Tm} = \frac{1}{\sqrt{k_0}}V_m^*$ and note that ξ_{Tm} is a martingale array with respect to the σ-field $\mathcal{F}_{T,m-1}^* = \sigma\left(I_1, \ldots, I_{m-1}\right)$ given the independence of V_m^*. For each $r \in [0,1]$, $\sum_{m=1}^{[rk_0]+1} Var^*\left(\xi_{Tm}\right) = \sum_{m=1}^{[rk_0]+1}\frac{1}{k_0} = \frac{[rk_0]+1}{k_0} \to r$ as $k_0 \to \infty$, which verifies Billingsley's (1968, p. 194) condition (18.3). Next we verify that the Lindeberg condition (cf. Billingsley, 1968, eq. (18.4)) holds in probability. For this, it suffices that $\sum_{m=1}^{[k_0r]+1}E^*\left|\xi_{Tm}\right|^{2+\delta} \to 0$ in probability. Since $\Omega_T^{*-1/2} = O_P(1)$, we need to show that $k_0^{-\frac{2+\delta}{2}}\sum_{m=1}^{[k_0r]+1}E^*\left|U_m^*\right|^{2+\delta} \to 0$ in probability. By definition of U_m^*, we have that

$$E\left(E^*\left|U_m^*\right|^{2+\delta}\right) \leq K\ell^{-\frac{2+\delta}{2}}\frac{1}{T-\ell+1}\sum_{j=0}^{T-\ell}E\left(\left|\sum_{s=1}^{\ell}X_{j+s}\right|^{2+\delta}\right).$$

An application of Lemma A.1 yields $E\left(\left|\sum_{s=1}^{\ell}X_{j+s}\right|^{2+\delta}\right) \leq K\ell^{\frac{2+\delta}{2}}$ uniformly in j, which implies that $E\left(E^{*}\left|U_{m}^{*}\right|^{2+\delta}\right) = O(1)$, showing that $k_0^{-\frac{2+\delta}{2}} \sum_{m=1}^{[k_0 r]+1} E^{*}\left|U_{m}^{*}\right|^{2+\delta} = O_P\left(k_0^{-\delta/2}\right) = o_P(1)$ since $k_0^{-\frac{\delta}{2}} = (\ell/T)^{\frac{\delta}{2}} \to 0$ under both sets of conditions on ℓ.

Proof of Lemma A.4. As in the proof of Lemma A.3, we can write

$$\begin{aligned} T^{-1}\sum_{t=1}^{[rT]}\left(X_t^{*}-E^{*}\left(X_t^{*}\right)\right) &= T^{-1}\sum_{m=1}^{M_r}\sum_{s=1}^{\ell}\left(X_{I_m+s}-E^{*}\left(X_{I_m+s}\right)\right)+T^{-1}\sum_{s=B+1}^{\ell}\left(X_{I_{M_r}+s}-E^{*}\left(X_{I_{M_r}+s}\right)\right) \\ &\equiv A_{1T}^{*}(r)+A_{2T}^{*}(r). \end{aligned}$$

Let $U_m^{*} = \sum_{s=1}^{\ell}\left(X_{I_m+s}-E^{*}X_{I_m+s}\right)$ and note that $S_j^{*} = \sum_{m=1}^{j} U_m^{*}$ is a martingale array with respect to $\mathcal{F}_{T,j}^{*} = \sigma\left(I_1, \ldots, I_j\right)$. Thus, by an application of Markov's inequality first, and of Doob's inequality second, we have that

$$P^{*}\left(\sup_{r\in[0,1]}\left|A_{1T}^{*}(r)\right|>\eta\right) \leq \frac{1}{\eta^2 T^2}E^{*}\left(\sup_{r\in[0,1]}\left|\sum_{m=1}^{M_r}U_m^{*}\right|^2\right) = \frac{1}{\eta^2 T^2}E^{*}\left(\max_{1\leq j\leq k_0}\left|S_j^{*}\right|^2\right) \leq KT^{-2}E^{*}\left(\left|S_{k_0}^{*}\right|^2\right).$$

Adding and subtracting $\mu = E\left(X_t\right)$,

$$E^{*}\left(\left|S_{k_0}^{*}\right|^2\right) = k_0 E^{*}\left(U_1^{*2}\right) = k_0 E^{*}\left[\left(\sum_{s=1}^{\ell}\left(X_{I_1+s}-\mu-E^{*}\left(X_{I_1}-\mu\right)\right)\right)^2\right] \leq K k_0 E^{*}\left(\left|\sum_{s=1}^{\ell}\left(X_{I_1+s}-\mu\right)\right|^2\right),$$

for some constant K. Using the properties of the MBB and Lemma A.1, we can show that $E\left(E^{*}\left(\left|\sum_{s=1}^{\ell}\left(X_{I_1+s}-\mu\right)\right|^2\right)\right) = O(\ell)$, which implies $E^{*}\left(\left|S_{k_0}^{*}\right|^2\right) = O(T)$, and thus $P^{*}\left(\sup_{r\in[0,1]}\left|A_{1T}^{*}(r)\right|>\eta\right) = O_P\left(T^{-1}\right) = o_P(1)$. Similarly, we can show that $E^{*}\left(\sup_{0\leq r\leq 1}\left|A_{2T}^{*}(r)\right|\right) = O_P\left(\frac{\ell^{1/2}}{T}\right) = o_P(1)$ under both a) and b).

Appendix B

This Appendix contains the proofs of the results in Section 5. We first present two useful lemmas. We then present the proofs of the main results, followed by the proofs of the auxiliary lemmas. Throughout this Appendix, we let $\Omega = \pi^2(1)\sigma^2$, $\Lambda = \pi(1)\sigma$, and $\tilde{W}(r) \equiv \Lambda W_1(r)$.

Lemma A.5 *Under Assumption 2, and for any fixed $b \in (0,1]$, we have*

a) $T^{-1/2}\sum_{t=1}^{T} u_t = \Lambda W_1(1) + O_P\left(T^{-1/2+1/p}\right).$

b) $T^{-2}\sum_{t=1}^{T-1} S_t^2 = \Omega \int_0^1 W_1^2(r)\,dr + O_P\left(T^{-1/2+1/p}\right).$

c) $T^{-3/2}\sum_{t=1}^{T-1}\left(\frac{t}{T}\right) S_t = \Lambda \int_0^1 rW_1(r)\,dr + O_P\left(T^{-1/2+1/p}\right).$

d) $T^{-3/2}\sum_{t=1}^{T-[bT]-1}\left(\frac{t}{T}\right) S_t = \Lambda \int_0^{1-b} rW_1(r)\,dr + O_P\left(T^{-1/2+1/p}\right).$

e) $T^{-3/2}\sum_{t=1}^{T-[bT]-1}\left(\frac{t}{T}\right) S_{t+[bT]} = \Lambda \int_0^{1-b} rW_1(r+b)\,dr + O_P\left(T^{-1/2+1/p}\right).$

f) $T^{-3/2}\sum_{t=1}^{T-[bT]-1}\frac{[bT]}{T} S_t = b\Lambda \int_0^{1-b} W_1(r)\,dr + O_P\left(T^{-1/2+1/p}\right).$

g) $T^{-2}\sum_{t=1}^{T-[bT]-1} S_t S_{t+[bT]} = \Omega \int_0^{1-b} W_1(r)\,W_1(r+b)\,dr + O_P\left(T^{-1/2+1/p}\right).$

Lemma A.6 *Under Assumption 2, with probability approaching one, we have that*

a) $T^{-1/2}\sum_{t=1}^{T} u_t^* = \Lambda^* W_1(1) + O_{P^*}\left(T^{-1/2+1/p}\right).$

b) $T^{-2}\sum_{t=1}^{T-1} S_t^{*2} = \Omega^* \int_0^1 W_1^2(r)\,dr + O_{P^*}\left(T^{-1/2+1/p}\right).$

c) $T^{-3/2}\sum_{t=1}^{T-1}\left(\frac{t}{T}\right) S_t^* = \Lambda^* \int_0^1 rW_1(r)\,dr + O_{P^*}\left(T^{-1/2+1/p}\right).$

d) $T^{-3/2}\sum_{t=1}^{T-[bT]-1}\left(\frac{t}{T}\right) S_t^* = \Lambda^* \int_0^{1-b} rW_1(r)\,dr + O_{P^*}\left(T^{-1/2+1/p}\right).$

e) $T^{-3/2}\sum_{t=1}^{T-[bT]-1}\left(\frac{t}{T}\right)S^*_{t+[bT]}=\Lambda^*\int_0^{1-b}rW_1\left(r+b\right)dr+O_{P^*}\left(T^{-1/2+1/p}\right).$

f) $T^{-3/2}\sum_{t=1}^{T-[bT]-1}\frac{[bT]}{T}S^*_t=b\Lambda^*\int_0^{1-b}W_1\left(r\right)dr+O_{P^*}\left(T^{-1/2+1/p}\right).$

g) $T^{-2}\sum_{t=1}^{T-[bT]-1}S^*_tS^*_{t+[bT]}=\Omega^*\int_0^{1-b}W_1\left(r\right)W_1\left(r+b\right)dr+O_{P^*}\left(T^{-1/2+1/p}\right).$

where $\Omega^*=\sigma^2\sum_{i=1}^{\infty}\pi_i^2$ *and* $\Lambda^*=\Omega^{*1/2}$.

Proof of Lemma 5.1. Theorem 3 of Akonom (1993) implies that under our assumptions

$$P\left(\sup_{r\in[0,1]}\left|W_T\left(r\right)-\tilde{W}\left(r\right)\right|>c_T\right)\leq C_2T^{1-p/2}c_T^{-p}E\left|\varepsilon_t\right|^p,$$

for any sequence c_T such that $T^{-1/2+1/p}\leq c_T\leq C_1\left(\log T\right)^{1/2}$, where C_1 and C_2 are constants independent of T. Part a) follows by letting $c_T=cT^{-1/2+1/p}$, for some constant c, whereas part b) follows by setting $c_T=cT^{-1/2+3/2p}$.

Proof of Lemma 5.2. Write $\hat{\Omega}=J_1-J_2$, with $J_1=2b^{-1}T^{-2}\sum_{t=1}^{T-1}\hat{S}_t^2$ and $J_2=2b^{-1}T^{-2}\sum_{t=1}^{T-[bT]-1}\hat{S}_t\hat{S}_{t+[bT]}$. We can write $J_1=2I_1-4I_2+2I_3$, where by Lemma A.5, $I_1=T^{-2}\sum_{t=1}^{T-1}S_t^2=\Omega\int_0^1W_1^2\left(r\right)dr+O_P\left(T^{-1/2+1/p}\right)$;

$$\begin{aligned}I_2&=T^{-2}\sum_{t=1}^{T-1}\left(\frac{t}{T}\right)S_tS_T=\left(T^{-3/2}\sum_{t=1}^{T-1}\left(\frac{t}{T}\right)S_t\right)\left(T^{-1/2}S_T\right)\\&=\left(\Lambda\int_0^1rW_1\left(r\right)dr+O_P\left(T^{-1/2+1/p}\right)\right)\left(\Lambda W_1\left(1\right)+O_P\left(T^{-1/2+1/p}\right)\right)\\&=\Omega\int_0^1rW_1\left(r\right)W_1\left(1\right)dr+O_P\left(T^{-1/2+1/p}\right);\end{aligned}$$

and

$$\begin{aligned}I_3&=T^{-2}\sum_{t=1}^{T-1}\left(\frac{t}{T}\right)^2S_T^2=T^{-1}\sum_{t=1}^{T-1}\left(\frac{t}{T}\right)^2\left(T^{-1/2}S_T\right)^2\\&=\left(\frac{1}{3}+O\left(T^{-1}\right)\right)\left(\Lambda W_1\left(1\right)+O_P\left(T^{-1/2+1/p}\right)\right)^2=\frac{1}{3}\Omega W_1^2\left(1\right)+O_P\left(T^{-1/2+1/p}\right),\end{aligned}$$

since $T^{-1}\sum_{t=1}^{T-1}\left(\frac{t}{T}\right)^2=\frac{1}{6}\frac{2T^2-3T+1}{T^2}=\frac{1}{3}+O\left(T^{-1}\right)$. Thus,

$$J_1=\Omega\left[\frac{2}{b}\int_0^1\left(W_1\left(r\right)-rW_1\left(1\right)\right)^2dr\right]+O_P\left(T^{-1/2+1/p}\right).$$

Next we analyze J_2. Notice that we can write

$$\begin{aligned}\hat{S}_t\hat{S}_{t+[bT]} &= \left(S_t - \frac{t}{T}S_T\right)\left(S_{t+[bT]} - \frac{t+[bT]}{T}S_T\right)\\ &= S_tS_{t+[bT]} - \frac{t+[bT]}{T}S_tS_T - \frac{t}{T}S_{t+[bT]}S_T + \frac{t}{T}\frac{t+[bT]}{T}S_T^2,\end{aligned}$$

implying that

$$\begin{aligned}J_2 &= \frac{2}{b}T^{-2}\sum_{t=1}^{T-[bT]-1}S_tS_{t+[bT]} - \frac{2}{b}T^{-2}\sum_{t=1}^{T-[bT]-1}\frac{t+[bT]}{T}S_tS_T\\ &\quad -\frac{2}{b}T^{-2}\sum_{t=1}^{T-[bT]-1}\frac{t}{T}S_{t+[bT]}S_T + \frac{2}{b}T^{-2}\sum_{t=1}^{T-[bT]-1}\frac{t}{T}\frac{t+[bT]}{T}S_T^2\\ &\equiv A_1 - A_2 - A_3 + A_4.\end{aligned}$$

By Lemma A.5, we have that

$$A_1 = \frac{2}{b}\Omega\int_0^{1-b}W_1(r)\,W_1(r+b)\,dr + O_P\left(T^{-1/2+1/p}\right);$$

$$\begin{aligned}A_2 &= \frac{2}{b}T^{-2}\sum_{t=1}^{T-[bT]-1}\frac{t}{T}S_tS_T + \frac{2}{b}T^{-2}\sum_{t=1}^{T-[bT]-1}\frac{[bT]}{T}S_tS_T\\ &= \frac{2}{b}\left(T^{-1/2}S_T\right)\left(T^{-3/2}\sum_{t=1}^{T-[bT]-1}\left(\frac{t}{T}\right)S_t\right) + \frac{2}{b}\left(T^{-1/2}S_T\right)\left(T^{-3/2}\sum_{t=1}^{T-[bT]-1}\frac{[bT]}{T}S_t\right)\\ &= \frac{2}{b}\Omega\int_0^{1-b}rW_1(r)\,W_1(1)\,dr + \frac{2}{b}b\Omega\int_0^{1-b}W_1(r)\,W_1(1)\,dr + O_P\left(T^{-1/2+1/p}\right)\\ &= \frac{2}{b}\Omega\int_0^{1-b}(r+b)\,W_1(r)\,W_1(1)\,dr + O_P\left(T^{-1/2+1/p}\right);\end{aligned}$$

$$A_3 = \frac{2}{b}\left(T^{-1/2}S_T\right)\left(T^{-3/2}\sum_{t=1}^{T-[bT]-1}\frac{t}{T}S_{t+[bT]}\right) = \frac{2}{b}\Omega\int_0^{1-b}rW_1(r+b)\,W_1(1)\,dr + O_P\left(T^{-1/2+1/p}\right),$$

and

$$\begin{aligned}A_4 &= \frac{2}{b}\left(T^{-1/2}S_T\right)^2\left(T^{-1}\sum_{t=1}^{T-[bT]\ -1}\frac{t}{T}\frac{t+[bT]}{T}\right)\\ &= \frac{2}{b}\Omega W_1^2(1)\left[\frac{1}{3}(1-b)^3 + \frac{1}{2}b(1-b)^2\right] + O_P\left(T^{-1/2+1/p}\right).\end{aligned}$$

The last result uses the fact that

$$T^{-3}\sum_{t=1}^{T-[bT]-1} t\,(t+[bT]) = T^{-3}\sum_{t=1}^{T-[bT]-1} t^2 + T^{-3}\sum_{t=1}^{T-[bT]-1} t\,[bT]$$
$$= \frac{1}{3}(1-b)^3 + \frac{1}{2}b\,(1-b)^2 + O\left(T^{-1}\right).$$

The desired result follows from combining all the previous expansions.

Proof of Theorem 5.1. Write $t_{\hat{\beta}_1} = P_T \Omega_1^{-1/2}\left(\frac{\hat{\Omega}}{\Omega_1}\right)^{-1/2}$, where $P_T = T^{-1/2}\sum_{t=1}^{T} u_t$ and $\Omega_1 = \Omega Q_1(b)$ is the leading term of the expansion of $\hat{\Omega}$. Note that by a Taylor expansion of $f(x) = (1+x)^{-1/2}$ around 0 we can write

$$\left(\frac{\hat{\Omega}}{\Omega_1}\right)^{-1/2} = \left(1+\frac{\hat{\Omega}-\Omega_1}{\Omega_1}\right)^{-1/2} = 1 - \frac{1}{2}\frac{\hat{\Omega}-\Omega_1}{\Omega_1} + O_P\left(\left(\hat{\Omega}-\Omega_1\right)^2\right).$$

Lemma 5.2 implies that $\hat{\Omega}-\Omega_1 = O_P\left(T^{-1/2+1/p}\right)$, and since $\Omega_1 = O_P(1)$, we get that

$$t_{\hat{\beta}_1} = P_T\Omega_1^{-1/2}\left(1+O_P\left(T^{-1/2+1/p}\right)\right) = P_T\Omega_1^{-1/2} + O_P\left(T^{-1/2+1/p}\right).$$

Lemma A.5. a) now implies the result.

Proof of Corollary 5.1. We follow the proof of Corollary 3.8 of Park (2003). In particular, the result follows from Lemma A4 of Park (2003) given that the error terms of the asymptotic expansions in Lemma A.5 are distributionally of order $O(T^{-a})$, with $a = 1/2 - 3/(2p)$, and that the density of Ω_1^{-1} is bounded and all its moments are finite (which follows because Ω_1 is a quadratic form of a Brownian motion with a truncated positive definite kernel). The remainder terms for each statistic are defined in the proof of Lemma A.5. Thus, the remainder term of part a) of Lemma A.5 is equal to $R_{1T} = W_T(1) - \tilde{W}(1)$, which is distributionally of order $O(T^{-a})$ given part b) of Lemma 5.1. For part b), the remainder is $R_{2T} = R_{2T}^{(1)} + R_{2T}^{(2)}$, where $\left|R_{2T}^{(1)}\right| \leq \sup\left|W_T(r) - \tilde{W}(r)\right|^2$ and $\left|R_{2T}^{(2)}\right| \leq 2\left(\sup\left|W_T(r)-\tilde{W}(r)\right|^2\right)^{1/2}\left(\int_0^1 \tilde{W}^2(r)\,dr\right)^{1/2}$. We have that

$$P\left(\sup\left|W_T(r)-\tilde{W}(r)\right|^2 \geq T^{-a}\right) \leq P\left(\sup\left|W_T(r)-\tilde{W}(r)\right|^2 \geq T^{-2a}\right)$$
$$\leq P\left(\sup\left|W_T(r)-\tilde{W}(r)\right| \geq T^{-a}\right) = O\left(T^{-a}\right),$$

showing that $R_{2T}^{(1)}$ is distributionally of order $O(T^{-a})$. Since $\int_0^1 \tilde{W}^2(r)\,dr$ has moments finite up to any order, Lemma A4. b) of Park (2003) implies that $R_{2T}^{(2)}$ is also distributionally of order $O(T^{-a})$. For part c), the remainder is $R_{3T} = R_{3T}^{(1)} + R_{3T}^{(2)}$, where $\left|R_{3T}^{(1)}\right| \leq \sup\left|W_T(r)-\tilde{W}(r)\right|$ and $\left|R_{3T}^{(2)}\right| \leq \sup\left|W_T(r)-\tilde{W}(r)\right| + \frac{1}{T}\int_0^1\left|\tilde{W}(r)\right|dr$. Since $\sup\left|W_T(r)-\tilde{W}(r)\right|$ is distributionally of order $O(T^{-a})$ by Lemma 5.1. b), we only need to show that the same is true for $\frac{1}{T}\int_0^1\left|\tilde{W}(r)\right|dr$. This follows by an application of Markov's inequality, given that $E\left|\int_0^1\left|\tilde{W}(r)\right|dr\right| < \infty$. For part d), note that the remainder is $R_{4T} = R_{4T}^{(1)} + R_{4T}^{(2)}$, where $R_{4T}^{(1)}$ is majorized by the same term as R_{3T}, whereas $\left|R_{4T}^{(2)}\right| \leq \sup\left|W_T(r) - \tilde{W}(r)\right| + \frac{1}{T}\sup\left|\tilde{W}(r)\right|$, which can be handled as $R_{3T}^{(2)}$. The remainder in part e) can be decomposed as $R_{5T}^{(1)} + R_{5T}^{(2)} + R_{5T}^{(3)} + R_{5T}^{(4)} + R_{5T}^{(5)} + R_{5T}^{(6)}$, where $R_{5T}^{(1)}$ and $R_{5T}^{(5)}$ are majorized by $\sup\left|W_T(r) - \tilde{W}(r)\right|$; $R_{5T}^{(2)}$ and $R_{5T}^{(6)}$ are majorized by $\sup\left|W_T(r) - \tilde{W}(r)\right| + \frac{1}{T}\int_0^1\left|\tilde{W}(r)\right|dr$. For $R_{5T}^{(3)}$ and $R_{5T}^{(4)}$ we have that for $i = 3, 4$,

$$T^a P\left(\left|R_{5T}^{(i)}\right| \geq T^{-a}\right) \leq T^a P\left(\left|T^{-3/2}S_{[bT]}\right| \geq T^{-a}\right) \leq T^a \frac{E\left|S_{[bT]}\right|^p}{T^{(3/2-a)p}} \leq KT^{a+p/2-(3/2-a)p} = KT^{-1-\frac{3}{2p}-\frac{p}{2}},$$

which is $o(1)$ for any $p > 0$ and where we have used the fact that $E\left|S_{[bT]}\right|^p \leq T^{p/2}$ by Lemma A.1, part b). The remainders in parts f) and e) can be analyzed using similar arguments and therefore we omit their proofs.

Proof of Lemma 5.3. Since u_t^* are i.i.d. we can apply the strong approximation result of **?** to W_T^*, as in the proof of Lemma 2.4 of **?**. That is, we may choose W_T^* in the same probability space as the Brownian motion $\tilde{W}^*(r) = \Omega^{*1/2}W_1(r)$ such that W_T^* has the same conditional distribution as W_T^{0*} and verifies the following condition:

$$P^*\left(\sup_{0\leq r\leq 1}\left|W_T^*(r) - \tilde{W}^*(r)\right| > c_T\right) \leq Kc_T^{-p}T^{1-p/2}E^*\left|u_t^*\right|^p,$$

where $\tilde{W}^*(r) = \Omega^{*1/2}W(r)$. If we show that $E^*\left|u_t^*\right|^p = O_P(1)$, the first result follows by letting $c_T = cT^{-1/2+1/p}$ for some large $c > 0$ whereas the second result follows by letting $c_T = cT^{-1/2+3/(2p)}$. Note that

$$E^*\left|u_t^*\right|^p = T^{-1}\sum_{t=1}^T\left|\hat{u}_t\right|^p = T^{-1}\sum_{t=1}^T\left|u_t - \bar{u}\right|^p \leq KT^{-1}\sum_{t=1}^T\left|u_t\right|^p \to KE\left|u_t\right|^p < \infty,$$

for some constant K, in probability. This proves that $E^*\left|u_t^*\right|^p = O_P(1)$.

Proof of Lemma 5.4. Given Lemma 5.3, the proof follows the same reasoning as that of Lemma 5.2.

Proof of Theorem 5.2. Given Lemma 5.3, the proof follows the same reasoning as that of Theorem 5.1.

Proof of Corollary 5.2. Given part b) of Lemma 5.3 and Theorem 5.2, the proof is analogous to that of Corollary 5.2.

Proof of Lemma A.5. We follow closely the proof of Lemma 3.1 of **?**. For a), note that $T^{-1/2}\sum_{t=1}^T u_t = W_T(1) = \tilde{W}(1) + R_{1T}$, where $R_{1T} = W_T(1) - \tilde{W}(1)$. By Lemma 5.1.a), $R_{1T} = O_P\left(T^{-1/2+1/p}\right)$, proving the result. For b), let $S_0 \equiv 0$ and write $T^{-2}\sum_{t=1}^{T-1}S_t^2 = T^{-2}\sum_{t=1}^T S_{t-1}^2$. Note that $T^{-1/2}S_{t-1} = T^{-1/2}\sum_{i=1}^{t-1}u_i = W_T\left(\frac{t-1}{T}\right)$. Thus,

$$T^{-2}\sum_{t=1}^T S_{t-1}^2 = T^{-1}\sum_{t=1}^T W_T^2\left(\frac{t-1}{T}\right) = \int_0^1 W_T^2(r)\,dr = \int_0^1 \tilde{W}^2(r)\,dr + R_{2T},$$

where $R_{2T} = \int_0^1 \left(W_T^2(r) - \tilde{W}^2(r) \right) dr$. We can write

$$R_{2T} = \int_0^1 \left(W_T(r) - \tilde{W}(r) \right)^2 dr + 2\int_0^1 \left(W_T(r) - \tilde{W}(r) \right) \tilde{W}(r)\, dr \equiv R_{2T}^{(1)} + R_{2T}^{(2)}.$$

Then, by Lemma 5.1.a),

$$R_{2T}^{(1)} \leq \int_0^1 \sup_{r\in[0,1]} \left| W_T(r) - \tilde{W}(r) \right|^2 dr \leq \left(\sup_{r\in[0,1]} \left| W_T(r) - \tilde{W}(r) \right| \right)^2 = O_P\left(\left(T^{-1/2+1/p} \right)^2 \right).$$

For $R_{2T}^{(2)}$, by the Cauchy-Schwartz inequality,

$$\left| R_{2T}^{(2)} \right| \leq 2\left(\int_0^1 \left(W_T(r) - \tilde{W}(r) \right)^2 dr \right)^{1/2} \left(\int_0^1 \tilde{W}^2(r)\, dr \right)^{1/2} = 2A_{1T} \cdot A_{2T}.$$

By Lemma 5.1.a), $A_{1T} = O_P\left(T^{-1/2+1/p}\right)$ and $A_{2T} = \int_0^1 \tilde{W}^2(r)\, dr = O_P(1)$. For c), we can write

$$\begin{aligned} T^{-3/2} \sum_{t=1}^{T-1} \left(\frac{t}{T} \right) S_t &= T^{-1} \sum_{t=1}^{T} \frac{t-1}{T} \left(T^{-1/2} S_{t-1} \right) = T^{-1} \sum_{t=1}^{T} \frac{t-1}{T} W_T\left(\frac{t-1}{T} \right) \\ &= \sum_{t=1}^{T} \int_{(t-1)/T}^{t/T} \frac{[Tr]}{T} W_T(r)\, dr = \int_0^1 \frac{[Tr]}{T} W_T(r)\, dr = \int_0^1 r\tilde{W}(r)\, dr + R_{3T}, \end{aligned}$$

where $R_{3T} = R_{3T}^{(1)} + R_{3T}^{(2)}$, with

$$R_{3T}^{(1)} = \int_0^1 r\left(W_T(r) - \tilde{W}(r) \right) dr \quad \text{and} \quad R_{3T}^{(2)} = \int_0^1 \left(\frac{[Tr]}{T} - r \right) W_T(r)\, dr.$$

By Lemma 5.1.a) we have that $R_{3T}^{(1)} = O_P\left(T^{-1/2+1/p}\right)$. For $R_{3T}^{(2)}$, note that

$$\begin{aligned} \left| R_{3T}^{(2)} \right| &\leq \int_0^1 \left| \frac{[Tr]}{T} - r \right| \left| W_T(r) \right| dr \leq \frac{1}{T} \int_0^1 \left| W_T(r) \right| dr \\ &= \frac{1}{T} \int_0^1 \left| W_T(r) - \tilde{W}(r) \right| dr + \frac{1}{T} \int_0^1 \left| \tilde{W}(r) \right| dr \equiv A_{3T} + A_{4T}, \end{aligned}$$

where $A_{3T} = O_P\left(T^{-3/2+1/p}\right)$ and $A_{4T} = O_P\left(T^{-1}\right) = O_P\left(T^{-1/2+1/p}\right)$. For d), we have

$$T^{-3/2} \sum_{t=1}^{T-[bT]-1} \frac{t}{T} S_t = \int_0^{1-b} r\tilde{W}(r)\, dr + R_{4T},$$

where $R_{4T} \equiv R_{4T}^{(1)} + R_{4T}^{(2)}$. $R_{4T}^{(1)}$ is of the same form as R_{3T} but with the $\int_0^1$ replaced by $\int_0^{1-b}$, and $R_{4T}^{(2)} = \int_{1-b}^{1-b+b-[bT]/T} \frac{[Tr]}{T} W_T(r)\, dr$. Following the proof for R_{3T}, we can show that $R_{4T}^{(1)} =$

$O_P\left(T^{-1/2+1/p}\right)$. $R_{4T}^{(2)}$ can be bounded by

$$\int_{1-b}^{1-b+b-[bT]/T}\left|\frac{[Tr]}{T}\right|\left|W_T\left(r\right)\right|dr\leq\int_{1-b}^{1-[bT]/T}\left|r\right|\left|W_T\left(r\right)\right|dr$$
$$\leq\sup_{r\in[0,1]}\left|W_T\left(r\right)\right|\int_{1-b}^{1-[bT]/T}1dr\leq\sup_{r\in[0,1]}\left|W_T\left(r\right)\right|\left|b-\frac{[bT]}{T}\right|=O_P\left(T^{-1}\right),$$

since $\sup_{r\in[0,1]}\left|W_T\left(r\right)\right|=O_P\left(1\right)$ and $\left|b-\frac{[bT]}{T}\right|=O\left(T^{-1}\right)$. For e), write

$$\begin{aligned}T^{-3/2}\sum_{t=1}^{T-[bT]-1}\frac{t}{T}S_{t+[bT]} &= T^{-1}\sum_{t=1}^{T-[bT]}\frac{t-1}{T}\left(T^{-1/2}S_{t+[bT]-1}\right)=T^{-1}\sum_{t=[bT]+1}^{T}\frac{t-1-[bT]}{T}\left(T^{-1/2}S_{t-1}\right)\\ &= T^{-1}\sum_{t=[bT]+1}^{T}\frac{t-1}{T}\left(T^{-1/2}S_{t-1}\right)-T^{-1}\sum_{t=[bT]+1}^{T}\frac{[bT]}{T}\left(T^{-1/2}S_{t-1}\right)\equiv M_{1T}+M_{2T}.\end{aligned}$$

We analyze M_{1T} and M_{2T} separately. For M_{1T}, we can write

$$\begin{aligned}M_{1T} &= \int_{[bT]/T}^{1}\frac{[Tr]}{T}W_T\left(r\right)dr=\int_b^1\frac{[Tr]}{T}W_T\left(r\right)dr+\int_{[bT]/T}^{b}\frac{[Tr]}{T}W_T\left(r\right)dr\\ &= \int_b^1 r\tilde{W}\left(r\right)dr+\int_b^1 r\left(W_T\left(r\right)-\tilde{W}\left(r\right)\right)dr+\int_b^1\left(\frac{[rT]}{T}-r\right)W_T\left(r\right)dr+\int_{[bT]/T}^{b}\frac{[Tr]}{T}W_T\left(r\right)dr\\ &\equiv \int_b^1 r\tilde{W}\left(r\right)dr+R_{5T}^{(1)}+R_{5T}^{(2)}+R_{5T}^{(3)}.\end{aligned}$$

We can majorize $R_{5T}^{(1)}$ and $R_{5T}^{(2)}$ by the same terms that majorize $R_{3T}^{(1)}$ and $R_{3T}^{(2)}$ respectively. For $R_{5T}^{(3)}$ we have that

$$\left|R_{5T}^{(3)}\right|\leq\left|T^{-1/2}S_{[bT]}\right|\left(b-\frac{[bT]}{T}\right)\leq\left|T^{-3/2}S_{[bT]}\right|=O_P\left(T^{-1/2+1/p}\right).$$

For M_{2T}, we can write

$$\begin{aligned}M_{2T} &= -b\int_b^1\tilde{W}\left(r\right)dr-\frac{[bT]}{T}\int_{[bT]/T}^{b}W_T\left(r\right)dr-b\int_b^1\left(W_T\left(r\right)-\tilde{W}\left(r\right)\right)dr-\left(\frac{[bT]}{T}-b\right)\int_b^1 W_T\left(r\right)dr\\ &\equiv -b\int_b^1\tilde{W}\left(r\right)dr-R_{5T}^{(4)}-R_{5T}^{(5)}-R_{5T}^{(6)}.\end{aligned}$$

We can show that $\left|R_{5T}^{(4)}\right|\leq\left|T^{-3/2}S_{[bT]}\right|=O_P\left(T^{-1/2+1/p}\right)$ and $R_{5T}^{(5)}$ and $R_{5T}^{(6)}$ can also be shown to be $O_P\left(T^{-1/2+1/p}\right)$ by arguments similar to those used above. Thus

$$M_{1T}+M_{2T}=\int_b^1\left(r-b\right)\tilde{W}\left(r\right)dr+O_P\left(T^{-1/2+1/p}\right)=\int_0^{1-b}r\tilde{W}\left(r+b\right)dr+O_P\left(T^{-1/2+1/p}\right).$$

For f), letting $S_0 \equiv 0$, we can write

$$\begin{aligned}
T^{-3/2}\sum_{t=1}^{T-[bT]-1}\frac{[bT]}{T}S_t &= T^{-1}\sum_{t=1}^{T-[bT]}\frac{[bT]}{T}\left(T^{-1/2}S_{t-1}\right)=\frac{[bT]}{T}\sum_{t=1}^{T-[bT]}T^{-1}W_T\left(\frac{t-1}{T}\right)\\
&= \frac{[bT]}{T}\int_0^{1-b}W_T(r)\,dr+\frac{[bT]}{T}\int_{1-b}^{(T-[bT])/T}W_T(r)\,dr\\
&= \frac{[bT]}{T}\int_0^{1-b}\tilde{W}(r)\,dr+\frac{[bT]}{T}\int_0^{1-b}\left(W_T(r)-\tilde{W}(r)\right)dr+\frac{[bT]}{T}\int_{1-b}^{(T-[bT])/T}W_T(r)\,dr\\
&\equiv b\int_0^{1-b}\tilde{W}(r)\,dr+R_{6T},
\end{aligned}$$

where $R_{6T}=R_{6T}^{(1)}+R_{6T}^{(2)}+R_{6T}^{(3)}$, with

$$\begin{aligned}
R_{6T}^{(1)} &= \left(\frac{[bT]}{T}-b\right)\int_0^{1-b}\tilde{W}(r)\,dr=O_P\left(T^{-1}\right);\\
R_{6T}^{(2)} &= \frac{[bT]}{T}\int_0^{1-b}\left(W_T(r)-\tilde{W}(r)\right)dr=O_P\left(T^{-1/2+1/p}\right);\\
R_{6T}^{(3)} &= \frac{[bT]}{T}\int_{1-b}^{(T-[bT])/T}W_T(r)\,dr\leq\frac{1}{T}\int_0^1\left|W_T(r)\right|dr=O_P\left(T^{-1}\right).
\end{aligned}$$

Finally, for g) write

$$\begin{aligned}
T^{-2}\sum_{t=1}^{T-[bT]-1}S_tS_{t+[bT]} &= T^{-1}\sum_{t=1}^{T-[bT]}\left(T^{-1/2}S_{t-1}\right)\left(T^{-1/2}S_{t-1+[bT]}\right)\\
&= T^{-1}\sum_{t=1+[bT]}^{T}\left(T^{-1/2}S_{t-1-[bT]}\right)\left(T^{-1/2}S_{t-1}\right)\\
&= \int_b^1\tilde{W}(r-b)\,\tilde{W}(r)\,dr+R_{7T}=\int_0^{1-b}\tilde{W}(r)\,\tilde{W}(r+b)\,dr+R_{7T},
\end{aligned}$$

where $R_{7T}=R_{7T}^{(1)}+R_{7T}^{(2)}+R_{7T}^{(3)}$, with

$$\begin{aligned}
R_{7T}^{(1)} &= \int_b^1\left(W_T(r-b)-\tilde{W}(r-b)\right)\tilde{W}(r)\,dr;\\
R_{7T}^{(2)} &= \int_b^1\left(W_T(r)-\tilde{W}(r)\right)W_T(r-b)\,dr;\\
R_{7T}^{(3)} &= \int_{[bT]/T}^{b}W_T(r-b)\,W_T(r)\,dr.
\end{aligned}$$

Using arguments similar as above, we can show that each of these terms is $O_P\left(T^{-1/2+1/p}\right)$, completing the proof.

Proof of Lemma A.6. The first result follows trivially from part a) of Lemma 5.3. The remaining result follow exactly as in the proof of Lemma A.5, given Lemma 5.3.

Figure 1: Empirical Null Rejection Probabilities, Linear Regression, Bartlett Kernel, $T = 25$

Figure 2: Empirical Null Rejection Probabilities, Linear Regression, Bartlett Kernel, $T = 50$

Figure 3: Empirical Null Rejection Probabilities, Linear Regression, QS Kernel, $T = 25$

Figure 4: Empirical Null Rejection Probabilities, Linear Regression, QS Kernel, $T = 50$

Figure 5: Empirical Null Rejection Probabilities, Simple Location Model, Bartlett Kernel, $T = 25$

Figure 6: Empirical Null Rejection Probabilities, Simple Location Model, Bartlett Kernel, $T = 50$

Figure 7: Empirical Null Rejection Probabilities, Simple Location Model, QS Kernel, $T = 25$

Figure 8: Empirical Null Rejection Probabilities, Simple Location Model, QS Kernel, $T = 50$