

# FIXED-B ASYMPTOTICS FOR THE STUDENTIZED MEAN FROM TIME SERIES WITH SHORT, LONG, OR NEGATIVE MEMORY

TUCKER MCELROY  
*U.S. Census Bureau*

DIMITRIS N. POLITIS  
*University of California at San Diego*

This paper considers the problem of variance estimation for the sample mean in the context of long memory and negative memory time series dynamics, adopting the fixed-bandwidth approach now popular in the econometrics literature. The distribution theory generalizes the short memory results of Kiefer and Vogelsang (2005, *Econometric Theory* 21, 1130–1164). In particular, our results highlight the dependence on the kernel (we include flat-top kernels), whether or not the kernel is nonzero at the boundary, and, most important, whether or not the process is short memory. Simulation studies support the importance of accounting for memory in the construction of confidence intervals for the mean.

## 1. INTRODUCTION

This paper considers the asymptotics of estimates of the variance of the sample mean constructed from a kernel-smoothed sum of sample autocovariances, when the underlying data generating process (DGP) exhibits either long, short, or intermediate memory. As in Kiefer and Vogelsang (2002, 2005), we work out the so-called fixed- $b$  asymptotics, i.e., the case that bandwidth is a fixed proportion of sample size. In Kiefer, Vogelsang, and Bunzel (2000), results are obtained for the Bartlett kernel and show that the limiting numerator and denominator are independent. Although this result is more generally true for all kernels, it does not hold for the case of long/intermediate memory, as is shown in Theorems 1 and 2 below. A deeper analysis of the reasons for this phenomenon is afforded by the Fourier-Laplace transform (FLT) techniques described below.

We study the situation in which we have a sample  $Y = \{Y_1, Y_2, \dots, Y_n\}$  from a strictly stationary time series with mean  $EY_t = \mu$ , autocovariance

Research partially supported by NSF grant DMS-07-06732. We have benefited from the helpful comments of two anonymous referees and the editors. This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau. Address correspondence to Dimitris Politis, Dept. of Mathematics, University of California at San Diego, La Jolla, CA 92093-0112; e-mail: dpolitis@ucsd.edu.

$\gamma_h = \text{Cov}(Y_t, Y_{t+h})$ , and integrable spectral density function  $f(\lambda) = \sum_h \gamma_h e^{-ih\lambda}$ . Memory strength can be parameterized through the partial sums of autocovariances;

$$\sum_{|h| < n} \gamma_h \sim CL(n)n^\beta, \quad (1)$$

where in general  $A_n \sim B_n$  denotes  $A_n/B_n \rightarrow 1$  as  $n \rightarrow \infty$ . In (1),  $C$  is a positive constant, and  $L$  is slowly varying at infinity (Embrechts, Klüppelberg, and Mikosch, 1997), with a limit that can be zero, one, or infinity. Then  $\beta$  and  $L$  parameterize memory as follows:  $1 > \beta > 0$  or  $\beta = 0$  and  $L$  tending to infinity correspond to long memory (LM), in which case  $f(0) = \infty$ ;  $\beta = 0$  and  $L$  tending to unity correspond to the usual short memory (SM) case where  $0 < f(0) < \infty$ ; finally,  $-1 < \beta < 0$  or  $\beta = 0$  and  $L$  tending to zero correspond to the less-studied case where  $f(0) = 0$ , which we will denote by negative memory (NM). In this context, Brockwell and Davis (1991) used the terminology ‘‘intermediate memory,’’ whereas others have used ‘‘antipersistence’’ (Robinson (2005), Lieberman and Phillips (2008)) or ‘‘negative dependence’’ (Samorodnitsky and Taqqu, 1994) due to negative correlations; our choice of terminology follows these latter authors. These definitions encompass autoregressive fractionally integrated moving average (ARFIMA) models (Hosking, 1981), fractional exponential models (see Beran, 1993, 1994), and fractional Gaussian noise models. Some authors prefer to parameterize memory in terms of the rate of explosion of  $f$  or  $1/f$  at frequency zero, but it is more convenient for us to work in the time domain; see Palma (2007) for a recent overview.

We stipulate  $\beta < 1$  to ensure that the sample mean is a consistent estimator of the mean  $\mu$ , and we assume  $\beta > -1$  to ensure that  $Y_t$  is not over-differenced, i.e., equal to the first difference of another stationary process. The SM case was covered in Kiefer and Vogelsang (2005), who used a variety of kernels for smoothing; our results provide extensions to LM and NM DGPs. A related paper is Sun (2004), which treats regression problems for both stationary and nonstationary LM DGPs. In contrast to these papers, we here consider the NM case as well as general flat-top tapers, and study the properties of the limit distribution in Theorem 2. The chief problem of interest is to properly normalize the partial sums  $S_n = \sum_{t=1}^n Y_t$ , which have finite-sample variance  $V_n$ . In general  $V_n$  grows at a rate dependent on  $\beta$  (e.g., see Taqqu, 1975; Taniguchi and Kakizawa, 2000), which makes the problem of normalization more tricky. Supposing that  $V_n^{-1/2}(S_n - n\mu)$  converges weakly to a nondegenerate distribution, it is of interest to develop an estimate of  $V_n$  that can be plugged in. We consider an estimator  $V_{\Lambda, M}$  based on a kernel-smoothed sum of sample autocovariances and bandwidth  $M$ , which grows at the same asymptotic rate as  $V_n$ .

Alternatively, one might consider a small-bandwidth approach to the above problem, which is studied in Robinson (2005). That work estimates  $\sqrt{V_n}$  directly by employing a consistent estimate of  $\beta$ . Our work, in contrast, focuses on the fixed- $b$  asymptotics for generic kernels (including flat-top kernels) and

demonstrates that the limiting distribution depends vitally on the unknown  $\beta$ , but not, interestingly, upon any parameters associated with remaining short-run dynamics; i.e., the limit distribution is pivotal with respect to short-term dependence. One can then obtain practicable estimates of the limiting quantiles via utilizing a plug-in estimator of  $\hat{\beta}$ . This differs from Robinson in that the plug-in estimator here is not used to construct an estimate of  $V_n$ , but rather to estimate the limit quantiles. Our theoretical results, along with discussion, are provided in Section 2. The behavior of simulated quantiles is summarized in Section 3, where we also study the empirical size of a plug-in approach to estimation of the limit quantiles. Technical proofs are deferred to the Appendix.

## 2. ASYMPTOTIC RESULTS

As in Kiefer and Vogelsang (2005), let the bandwidth  $M$  be proportional to sample size  $n$ , i.e.,  $M = bn$  with  $b \in (0, 1]$ . We first introduce the following notation: The sample autocovariance is  $\hat{\gamma}_k = n^{-1} \sum_{t=1}^{n-k} (Y_{t+k} - \bar{Y})(Y_t - \bar{Y})$  for  $0 \leq k < n$ , and  $\bar{Y}$  the sample mean. Also let  $\hat{S}_i = \sum_{t=1}^i (Y_t - \bar{Y})$  (so that  $\hat{S}_n = 0$ ), and define the kernel-smoothed sum of autocovariances by  $V_{\Lambda, M} = \sum_h \Lambda(h/M) \hat{\gamma}_h$ , where  $\Lambda$  is a kernel. We consider kernels  $\Lambda(x)$  from the following general family:

{ $\Lambda$  is continuous and even, with support on  $[-1, 1]$  such that  $\Lambda(x) = 1$  for  $|x| \leq c$ , for some  $c \in [0, 1)$ ; also,  $\Lambda$  is continuous and twice continuously differentiable on  $(c, 1) \cup (-1, -c)$ .}

(2)

The above class of kernels includes the family of flat-top kernels of Politis (2005) where  $c > 0$ , the Bartlett kernel (letting  $c = 0$  and a linear decay of  $\Lambda$ ), as well as other kernels considered in Kiefer and Vogelsang (2005).

A derivative of  $\Lambda$  from the left (with respect to  $x$ ) is denoted  $\dot{\Lambda}_-$ , whereas from the right it is  $\dot{\Lambda}_+$ ; the second derivative is  $\ddot{\Lambda}$ . The greatest integer function is denoted by  $[\cdot]$ . With this notation, the following basic proposition is presented.

**PROPOSITION 1.** *Let  $\Lambda$  be a kernel from family (2), and assume (1) with  $|\beta| < 1$ . Then*

$$\begin{aligned} nV_{\Lambda, M} &= \sum_{i, j=1}^n \hat{S}_i \hat{S}_j \left( 2\Lambda \left( \frac{i-j}{M} \right) - \Lambda \left( \frac{i-j+1}{M} \right) - \Lambda \left( \frac{i-j-1}{M} \right) \right) \\ &= -\frac{2}{bn} \sum_{i=1}^{n-[c bn]} \hat{S}_i \hat{S}_{i+[c bn]} \left( \dot{\Lambda}_+(c) + \frac{1}{2bn} \ddot{\Lambda}(c) + O(n^{-2}) \right) \\ &\quad - \frac{1}{b^2 n^2} \sum_{[c bn] < |i-j| < [bn]} \hat{S}_i \hat{S}_j \left( \ddot{\Lambda} \left( \frac{|i-j|}{bn} \right) + O(n^{-1}) \right) \\ &\quad + \frac{2}{bn} \sum_{i=1}^{n-[bn]} \hat{S}_i \hat{S}_{i+[bn]} \left( \dot{\Lambda}_-(1) + O(n^{-1}) \right). \end{aligned}$$

**Remark 1.** In case the kernel is continuously differentiable at  $c$ ,  $\dot{\Lambda}_+(c) = 0$  and the second derivative becomes dominant in the first term, which can then be recombined with the second term to yield

$$-\frac{1}{b^2 n^2} \sum_{[cbn] \leq |i-j| < [bn]} \widehat{S}_i \widehat{S}_j \left( \ddot{\Lambda} \left( \frac{|i-j|}{bn} \right) + O(n^{-1}) \right).$$

Likewise, if there is no kink at  $|x| = 1$ , then  $\dot{\Lambda}_-(1) = 0$  and the third term vanishes completely.

Since we want the asymptotics of  $nV_{\Lambda, M}$ , we need functional limit theorems for the partial sums, as  $\widehat{S}_i = S_i - i/n S_n$ . To that end we suppose that

$$V_n^{-1/2} (S_{[nr]} - [nr]\mu) \xrightarrow{\mathcal{L}} B(r) \tag{3}$$

in the sense that the corresponding probability measures on  $\mathcal{D}[0, 1]$  converge weakly. Here  $\mathcal{D}[0, 1]$  refers to the space of functions on  $[0, 1]$  that are right continuous with left limits, endowed with the Skorohod topology (Taniguchi and Kakizawa, 2000). Also,  $B(\cdot)$  is a fractional Brownian motion (FBM) process of parameter  $(\beta + 1)/2$  (Samorodnitsky and Taqqu, 1994).

Sufficient conditions for (3) include linearity and a moment condition (see Taniguchi and Kakizawa, 2000, Thm. 5.2.4), as well as supposing that the process is an instantaneous functional of a long memory Gaussian (see Taqqu, 1975, Thm. 5.1, or Taniguchi and Kakizawa). In the interest of brevity we will henceforth assume that (3) holds, from which it follows that  $\widehat{S}_{[rn]}/\sqrt{V_n}$  converges weakly to the process  $\widetilde{B}(r) = B(r) - rB(1)$ , which is a fractional Brownian bridge (FBB). Then we may conclude the following result.

**THEOREM 1.** *Let  $\Lambda$  be a kernel from family (2), and suppose that  $\{Y_i\}$  is a DGP such that (3) holds. Also assume that (1) holds with  $|\beta| < 1$ . Then*

$$\frac{S_n - n\mu}{\sqrt{nV_{\Lambda, M}}} \xrightarrow{\mathcal{L}} \frac{B(1)}{\sqrt{Q(b)}} \tag{4}$$

as  $n \rightarrow \infty$ , where  $Q(b)$  is defined by

$$\begin{aligned} & -\frac{2}{b} \dot{\Lambda}_+(c) \int_0^{1-cb} \widetilde{B}(r) \widetilde{B}(r+cb) dr \\ & -\frac{1}{b^2} \int_{cb < |r-s| < b} \widetilde{B}(r) \widetilde{B}(s) \ddot{\Lambda} \left( \frac{|r-s|}{b} \right) dr ds \\ & +\frac{2}{b} \dot{\Lambda}_-(1) \int_0^{1-b} \widetilde{B}(r) \widetilde{B}(r+b) dr. \end{aligned}$$

**Example 1**

The trapezoidal kernel is the benchmark flat-top kernel whose use was proposed by Politis and Romano (1995); it is defined by

$$\Lambda^{T,c}(x) = \begin{cases} 1 & \text{if } |x| \leq c \\ \frac{|x|-1}{c-1} & \text{if } c < |x| \leq 1 \\ 0 & \text{else.} \end{cases}$$

Hence the second derivative for  $|x| \in (c, 1]$  is zero, and

$$Q(b) = \frac{2}{b(c-1)} \left( \int_0^{1-b} \tilde{B}(r) \tilde{B}(r+b) dr - \int_0^{1-cb} \tilde{B}(r) \tilde{B}(r+cb) dr \right).$$

We can further describe the joint distribution of  $B(1)$  and  $Q(b)$  through their joint Fourier-Laplace transform (FLT); see Fitzsimmons and McElroy (2010). In order to describe the result given below, we need to introduce some concepts from Tziritas (1987). Define the kernel as a tempered distribution,

$$\begin{aligned} K(r,s) = & -\frac{1}{b^2} \left\{ \ddot{\Lambda} \left( \frac{r-s}{b} \right) - \Delta_1(s) \int_0^1 x \ddot{\Lambda} \left( \frac{r-x}{b} \right) dx \right. \\ & - \Delta_1(r) \int_0^1 x \ddot{\Lambda} \left( \frac{x-s}{b} \right) dx \\ & \left. + \Delta_1(r) \Delta_1(s) \int_0^1 \int_0^1 xy \ddot{\Lambda} \left( \frac{x-y}{b} \right) dx dy \right\} \\ & - \frac{2}{b} \dot{\Lambda}_+(c) \Omega_c(r,s) + \frac{2}{b} \dot{\Lambda}_-(1) \Omega_1(r,s), \end{aligned}$$

$$\begin{aligned} \Omega_a(r,s) = & \Delta_{ab}(r-s) - \Delta_1(s) \int_0^{1-ab} x \Delta_{ab}(r-x) dx - \Delta_1(r) \int_0^{1-ab} x \Delta_{ab}(x-s) dx \\ & + \Delta_1(r) \Delta_1(s) \int_0^{1-ab} x(x+ab) dx, \end{aligned}$$

for  $r, s \in [0, 1]$ . Here  $\Delta_a$  for  $a \in [0, 1]$  denotes the Dirac delta function. We need to consider the space of square integrable (real-valued) tempered distributions on  $[0, 1]$ , endowed with the inner product  $\langle u, v \rangle = \int_0^1 u(s)v(s) ds$  for elements  $u, v$ . This is a separable Hilbert space and will be called  $H$ . Linear operators on  $H$  map elements to the space of functions on  $[0, 1]$  as follows: An operator  $O$  maps  $u \in H$  to  $Ou$ , which has value  $[Ou](s) = \int_0^1 O(s, x)u(x) dx$ . Let  $\iota \in H$  denote the function that is identically one; then the trace of an operator is  $\text{tr}[O] = \langle O\iota, \iota \rangle$ . Let  $\zeta \in H$  denote the line given by  $\zeta(s) = s$ ; then some algebra shows that  $K\zeta$  is equal to the zero element (this result is important later). Next, let the kernel function of FBM be denoted  $T(r, s) = \frac{1}{2} (r^{\beta+1} + s^{\beta+1} - |r-s|^{\beta+1})$ , which is nonnegative definite (Samorodnitsky and Taqqu, 1994, p. 318); moreover it is self-adjoint (symmetric), linear, and bounded on  $H$ . The identity operator is denoted by 1. We can now state our main result.

**THEOREM 2.** *For any  $\beta \in (-1, 1)$  the joint Fourier-Laplace transform of  $B(1)$ ,  $Q(b)$  is*

$$\mathbb{E} \exp\{i\theta B(1) - \phi Q(b)\} = \exp \left\{ -\frac{\theta^2}{2} \langle T(1 + 2\phi KT)^{-1} \Delta_1, \Delta_1 \rangle \right\} \\ [\det(1 + 2\phi KT)]^{-1/2}.$$

*It follows that  $B(1)/\sqrt{Q(b)}$  is symmetric about zero and has a cumulative distribution function that is continuous in  $\beta$ .*

Setting  $\theta = 0$  produces the Laplace transform of  $Q(b)$ , which is analogous to the  $\chi^2$ ; sufficient conditions for  $Q(b)$  to be  $\chi^2$  or a sum of gamma variables are given by Propositions 2 and 3 of Tziritas (1987). Setting  $\phi = 0$  produces  $\exp\{-\theta^2/2\}$ , corresponding to the standard normal, since  $\langle T \Delta_1, \Delta_1 \rangle = 1$  for all  $\beta \in (-1, 1)$ . Since the FLT does not in general factor into a product of functions only involving  $\theta$  and  $\phi$ , we conclude that  $B(1)$  and  $Q(b)$  are dependent. However, when  $\beta = 0$  we find that  $T$  simplifies to the pairwise minimum function; hence  $T \Delta_1 = \zeta$ . Expanding  $(1 + 2\phi KT)^{-1} = \sum_{n \geq 0} (-2\phi)^n (KT)^n$  in the first term of the FLT and using the fact that  $K\zeta = 0$ , all the terms in the infinite sum actually vanish except the first; as a result we are left with just  $\exp\{-\theta^2/2\}$  times the determinantal term. Thus we obtain independence of  $B(1)$  and  $Q(b)$  when  $\beta = 0$  (a known result). When  $\beta \neq 0$  dependence remains, essentially because  $T \Delta_1$  is nonlinear and hence not in the null space of  $K$ . The fact that  $K$  annihilates line stems directly from the sample mean centering that is used to construct sample autocovariances in  $V_{\Lambda, M}$ .

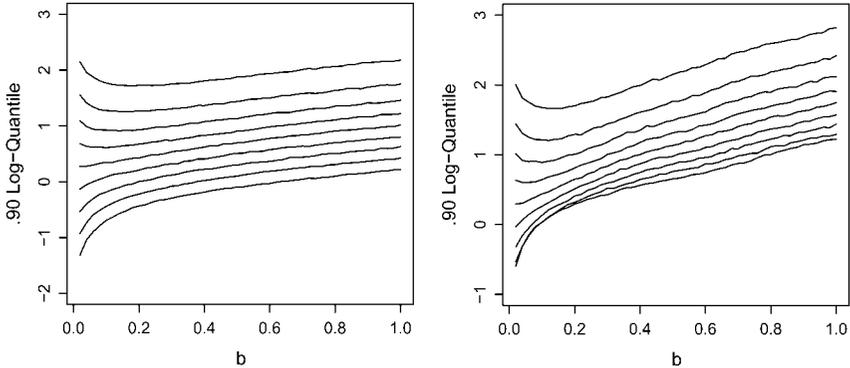
In principle we can set  $\theta = 0$  and differentiate the FLT to get all moments of  $Q(b)$ ; using the formulas from the proof of Theorem 2 we find that

$$\mathbb{E}[Q(b)]^k = (k-1)! 2^{k-1} \text{tr}[(KT)^k].$$

The trace calculation is algebraically challenging when  $k > 1$ , and hence mitigates the utility of the above formula. Finally, we note that the FLT does not provide a practical method for computing the cumulative distribution function of  $B(1)/\sqrt{Q(b)}$ , as the inversion of Laplace transforms is a difficult mathematical problem.

### 3. NUMERICAL RESULTS

In this section we investigate the distribution  $B(1)/\sqrt{Q(b)}$  of equation (4) for various choices of  $\beta$ ,  $b$ , and kernel. Following Kiefer and Vogelsang (2005), we simulated the upper quantiles of this distribution using the device of regressing on a convenient function of  $b$  for fixed kernel,  $\alpha$ -level, and  $\beta$ . (Since the distribution of  $B(1)/\sqrt{Q(b)}$  is symmetric for all  $|\beta| < 1$ , it is sufficient to consider the upper quantiles.) The simulation was conducted by generating 50,000 sample-paths of an FBB of length 1,000, and computing  $B(1)/\sqrt{Q(b)}$  for a given choice of kernel,



**FIGURE 1.** Log quantiles of the limiting studentized statistic based on 50,000 simulations of fractional Brownian bridge, for the upper .90 quantile, plotted against bandwidth proportion  $b$ . Multiple curves represent increasing values of  $\beta \in \{-.8, -.6, -.4, -.2, 0, .2, .4, .6, .8\}$  in ascending order (curves corresponding to higher values of  $\beta$  lie above those for lower values of  $\beta$ ). The left panel is for the Bartlett taper, and the right panel for the trapezoidal taper with  $c = .5$ .

50 values of  $b$  (evenly spaced between .02 and 1.00), and nine values of  $\beta \in \mathbb{B} = \{-.8, -.6, -.4, -.2, 0, .2, .4, .6, .8\}$ . Full tables summarizing our regression on to an exponential quintic function, as well as details concerning our method of simulation, can be found in the technical report of McElroy and Politis (2009).

Now, although results varied slightly by taper and percentile (we also considered the Parzen, Daniell, quadratic spectral, Tukey-Hanning, and Bohman tapers), qualitatively the behavior of the quantiles as a function of  $b$  and  $\beta$  can be demonstrated through a single figure. Figure 1 displays the results for the Bartlett (left) and trapezoidal (right) with  $c = .5$ , with  $b$  between .02 and 1.00 in increments of .02 along the x-axis, and  $\beta \in \mathbb{B}$  in successive curves overlaid. The y-axis displays the logarithm of the .90 upper one-sided quantile (logs are used to make the patterns more discernible). Note that the curves are ordered from lowest to highest values according to the values of  $\beta$ , their ordering being monotonic.

The middle curve corresponds to  $\beta = 0$ , and is identical (up to Monte Carlo error) to the results of Kiefer and Vogelsang (2005). The effect of both  $\beta$  and  $b$  on the limiting distribution is apparent from these graphs. The trumpet shape at the left of each plot is indicated by a divergence in the behavior of the mean of  $Q(b)$  when  $b$  is small, depending on whether  $\beta$  is positive or negative; this phenomenon is further discussed in McElroy and Politis (2009). Because these log quantile curves generally tend to move in parallel for moderate to high values of  $b$ , the consequences of using the  $\beta = 0$  distribution when in fact  $\beta \neq 0$  can be considerable at all bandwidths.

Consider now the practical procedure of using a plug-in estimator of  $\beta$  to construct a confidence interval. We utilize some consistent estimate  $\hat{\beta}$ , find the nearest

**TABLE 1.** Empirical  $\alpha$ -level (i.e., proportion of simulations for which true mean lay outside the constructed interval) for a nominal 95% two-sided confidence interval for the mean, using the plug-in quantile method described in Section 3

$b$	Memory $\beta$				
	-.8	-.4	0	.4	.8
Bartlett					
.2	.008 (.000)	.039 (.004)	.075 (.051)	.105 (.245)	.294 (.529)
.4	.016 (.000)	.039 (.003)	.078 (.041)	.108 (.186)	.256 (.463)
.6	.025 (.000)	.046 (.005)	.094 (.050)	.116 (.182)	.275 (.435)
.8	.041 (.000)	.054 (.006)	.093 (.053)	.137 (.192)	.287 (.432)
1.0	.049 (.000)	.067 (.005)	.086 (.057)	.135 (.184)	.325 (.438)
Trap ( $c = .5$ )					
.2	.048 (.019)	.054 (.033)	.068 (.050)	.079 (.113)	.202 (.323)
.4	.055 (.021)	.063 (.031)	.081 (.043)	.106 (.112)	.174 (.250)
.6	.061 (.012)	.084 (.027)	.100 (.063)	.116 (.117)	.201 (.248)
.8	.039 (.010)	.073 (.024)	.089 (.050)	.126 (.118)	.241 (.267)
1.0	.041 (.010)	.063 (.019)	.109 (.042)	.134 (.117)	.239 (.273)

Note: True  $\beta$  is given by column, and each row describes taper and bandwidth proportion. Numbers in parentheses correspond to the default method wherein the quantiles for the  $\beta = 0$  case are implicitly used no matter what the true  $\beta$  is.

$\beta \in \mathbb{B}$ , and compute the corresponding quantile for the specified taper and value of  $b$ . This procedure is not optimal, due to the discrete grid of  $\beta$ s that are available, and yet it may have superior size to the naïve procedure wherein the  $\beta = 0$  quantiles are utilized.

There is a large literature on the estimation of the memory parameter  $\beta$ ; available methods are either parametric (e.g., Giraitis and Taqqu, 1999), semiparametric (Giraitis and Surgailis, 1990; Hurvich, 2002), or even non-parametric (McElroy and Politis, 2007). We here implement an estimator suggested by the rate estimation framework of the latter reference, namely  $\hat{\beta} = \log V_{\Lambda, M} / \log n$ . Supposing that  $V_n \sim \tilde{C} L(n) n^{\beta+1}$  (cf. Thm. 5.2.1 of Taniguchi and Kakizawa, 2000) follows from (1), we have  $\log(n^{-1} V_n) / \log n \sim \beta$  since  $\log L(n) / \log n \rightarrow 0$  for any slowly varying function (Embrechts et al. 1997). Then, with  $n V_{\Lambda, M} / V_n \xrightarrow{\mathcal{L}} Q(b)$ , we have

$$\hat{\beta} = \frac{\log(n V_{\Lambda, M} / V_n)}{\log n} + \frac{\log(V_n / n)}{\log n} = \beta + O_P(1 / \log n) + o(1).$$

We implemented this procedure in simulation, examining Gaussian time series with  $\beta \in \{-.8, -.4, 0, .4, .8\}$ , for various tapers and sample sizes. Table 1 summarizes the empirical performance for a nominal 95% two-sided interval, for sample size  $n = 1,000$  and the Bartlett or trapezoidal ( $c = .5$ ) tapers, with  $b \in \{.2, .4, .6, .8, 1.0\}$ . Even this crude procedure produces some advantages over

the default method, which assumes that  $\beta = 0$ , except in the case where  $\beta$  is truly zero. Our proposed method is less effective for positive  $\beta$  than negative  $\beta$ , which seems to be due to the inferior performance of  $\hat{\beta}$  and the heavier tails of the limit distribution in the LM case.

#### 4. CONCLUSION

This paper investigates the distribution of the studentized sample mean in the context of NM and LM time series dynamics, adopting the fixed-bandwidth approach now popular in the econometrics literature. We derive the limiting distribution in Theorems 1 and 2, thus generalizing the results of Sun (2004) and Kiefer and Vogelsang (2005) to antipersistence and flat-top tapers. Our results highlight the influence of the kernel—e.g., whether or not the kernel is nonzero at the boundary of its support—and the influence of the DGP's type of memory. A main finding is that utilization of the SM fixed-bandwidth quantiles when NM or LM is present will imperil inference, whether  $b$  is small, moderate, or large. This assertion is further supported by our modest empirical studies; improved coverage can be obtained by using a plug-in estimator of  $\beta$ .

In practice, then, it is essential to have some estimate of  $\beta$ , whether one takes the approach of Robinson (2005) or the fixed-bandwidth approach discussed here. Once an estimate  $\hat{\beta}$  is obtained, one needs to compute the corresponding quantiles (via simulation, or by using a look-up table). However, this distribution will be distorted by the variability in  $\hat{\beta}$  (see Table 1 for an example); although the FLT of Theorem 2 is continuous in  $\beta$ , and hence a plug-in approach has theoretical merit, common estimators of  $\beta$  are notorious for having high variance, and this will likely have a large impact on inference. Future work should further study the finite-sample performance of a plug-in approach, and make comparisons with the alternative approach of Robinson (2005).

#### REFERENCES

- Beran, J. (1993) Fitting long-memory models by generalized linear regression. *Biometrika* 80, 817–822.
- Beran, J. (1994) *Statistics for Long Memory Processes*. Chapman and Hall.
- Brockwell, P. & R. Davis (1991) *Time Series: Theory and Methods*, 2nd ed. Springer.
- Embrechts, P., C. Klüppelberg, & T. Mikosch (1997) *Modeling Extremal Events for Insurance and Finance*. Springer-Verlag.
- Fitzsimmons, P. & T. McElroy (2010) On joint Fourier-Laplace transforms. *Communications in Statistics* 39, 1883–1885.
- Giraitis, L. & D. Surgailis (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate. *Probability Theory and Related Fields* 86, 87–104.
- Giraitis, L. & M. Taqqu (1999) Whittle estimator for finite-variance non-Gaussian time series with long memory. *Annals of Statistics* 27, 178–203.
- Hosking, J.R.M. (1981) Fractional differencing. *Biometrika* 68, 165–176.

- Hurvich, C. (2002) Multistep forecasting of long memory series using fractional exponential models. *International Journal of Forecasting* 18, 167–179.
- Kiefer, N. & T. Vogelsang (2002) Heteroscedastic-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica* 70, 2093–2095.
- Kiefer, N. & T. Vogelsang (2005) A new asymptotic theory for heteroscedasticity-autocorrelation robust tests. *Econometric Theory* 21, 1130–1164.
- Kiefer, N., T. Vogelsang, & H. Bunzel (2000) Simple robust testing of regression hypotheses. *Econometrica* 68, 695–714.
- Lieberman, O. & P. Phillips (2008) A complete asymptotic series for the autocovariance function of a long memory process. *Journal of Econometrics* 147, 99–103.
- McElroy, T. & D.N. Politis (2007) Computer-intensive rate estimation, diverging statistics, and scanning. *Annals of Statistics* 35, 1827–1848.
- McElroy, T. & D. Politis (2009) Fixed- $b$  Asymptotics for the Studentized Mean from Time Series with Short, Long or Negative Memory. Technical report, University of California San Diego.
- Palma, W. (2007) *Long-Memory Time Series*, Wiley.
- Politis, D.N. (2005) Higher-Order Accurate, Positive Semi-Definite Estimation of Large-Sample Covariance and Spectral Density Matrices, Discussion Paper 2005-03, University of California San Diego.
- Politis, D.N. & J.P. Romano (1995) Bias-corrected nonparametric spectral estimation. *Journal of Time Series Analysis* 16(1), 67–104.
- Robinson, P. (2005) Robust covariance matrix estimation: HAC estimates with long memory/ antipersistence correction. *Econometric Theory* 21, 171–180.
- Samorodnitsky, G. & M. Taquq (1994) *Stable Non-Gaussian Random Processes*. Chapman & Hall.
- Sun, Y. (2004) A convergent  $t$ -statistic in spurious regressions. *Econometric Theory* 20, 943–962.
- Taniguchi, M. & Y. Kakizawa (2000) *Asymptotic Theory of Statistical Inference for Time Series*. Springer-Verlag.
- Taqqu, M. (1975) Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 31, 287–302.
- Tziritas, G. (1987) On the distribution of positive-definite Gaussian quadratic forms. *IEEE Transactions on Information Theory* 33, 895–906.

## APPENDIX

**Proof of Proposition 1.** For shorthand let  $W_t = Y_t - \bar{Y}$ . Then, using summation by parts as in Kiefer and Vogelsang (2002, 2005),

$$\begin{aligned}
 nV_{\Lambda, M} &= \sum_{|h| < n} \Lambda(h/M) \sum_{t=1}^{n-|h|} W_t W_{t+|h|} \\
 &= \sum_{i,j=1}^n W_i W_j \Lambda\left(\frac{|i-j|}{bn}\right) \\
 &= \sum_{i=1}^n W_i \left[ \sum_{j=1}^{n-1} \left( \Lambda\left(\frac{i-j}{bn}\right) - \Lambda\left(\frac{i-j-1}{bn}\right) \right) \widehat{S}_j \right] \\
 &= \sum_{i,j=1}^n \widehat{S}_i \widehat{S}_j \left( 2\Lambda\left(\frac{i-j}{bn}\right) - \Lambda\left(\frac{i-j+1}{bn}\right) - \Lambda\left(\frac{i-j-1}{bn}\right) \right).
 \end{aligned}$$

Consider  $2\Lambda\left(\frac{h}{bn}\right) - \Lambda\left(\frac{h+1}{bn}\right) - \Lambda\left(\frac{h-1}{bn}\right)$ . If  $[cbn] < h < [bn]$ , then the approximation  $-b^{-2}n^{-2}\ddot{\Lambda}\left(\frac{h}{bn}\right)$  holds. If  $h = [cbn]$ , we obtain  $2\Lambda(c) - \Lambda(c+1/bn) - 1 + o(1) =$

$-\hat{\Lambda}_+(c)/bn - \check{\Lambda}(c)/(2b^2n^2) + O(n^{-3})$ . Finally, if  $h = [bn]$  we obtain  $-\Lambda(1 - 1/bn) + o(1) = \hat{\Lambda}_-(1)/bn + O(n^{-2})$ . This completes the proof of the proposition. ■

**Proof of Theorem 1.** This follows at once from Proposition 1 and (3), noting that  $V_n^{-1/2}(S_n - n\mu) \xrightarrow{\mathcal{L}} B(1)$  jointly with  $V_n^{-1/2}\hat{S}_i$  tending to  $\tilde{B}(i/n)$ . The second-order terms in  $nV_{\Lambda, M}$  in Proposition 1 drop out, and the summations become integrals. In the case that  $\hat{\Lambda}_+(c) = 0$ , we can apply Remark 1 and extend the integral to  $|r - s| = cb$ . Since this set has measure zero, it has no impact on the final limit  $Q(b)$ . ■

**Proof of Theorem 2.** First we consider the FBM process  $B(s)$ , which is a random function with values in  $H$ , such that it induces a probability measure; it is in this sense that we write  $B \in H$ . Then it is a tedious exercise to show that  $Q(b) = \langle KB, B \rangle$ . Let  $W$  be another Gaussian process that is mean zero with covariance kernel  $K$  and that is independent of  $B$ . Then the FLT is

$$\mathbb{E} \exp\{i\theta B(1) - \phi Q(b)\} = \mathbb{E} \exp\{i\theta B(1) + i\sqrt{2\phi} \langle B, W \rangle\},$$

which follows by use of conditional expectations, i.e., the characteristic function of  $\langle B, W \rangle$  conditional on  $B$  is given by the scaled Laplace transform of  $\langle KB, B \rangle$ . Next, let  $Z = \theta \Delta_1 + \sqrt{2\phi}W$ , which is another Gaussian process with mean element  $\theta \Delta_1$  and covariance kernel  $2\phi K$ . So our FLT is just  $\mathbb{E} \exp\{-\frac{1}{2} \langle TZ, Z \rangle\}$ , obtained this time by conditioning on  $Z$  (i.e., on  $W$ ). Note that  $T$  satisfies the conditions of Proposition 1 of Tziritas (1987), which we apply to get the stated result; also, we use formula (8) of that paper, which applies to the case of real-valued Gaussian processes. The  $n$ th cumulant of  $\langle TZ, Z \rangle$  is

$$\kappa_n = 2^{n-1} \left( (n-1)! \text{tr}[(2\phi KT)^n] + \theta^2 n! \langle T(2\phi KT)^{n-1} \Delta_1, \Delta_1 \rangle \right).$$

The FLT is then just  $\exp\{\sum_{n \geq 1} (-1/2)^n \kappa_n / n!\}$ , which upon manipulation easily produces the stated formula. The cumulant formula can also be used to obtain the moments of  $Q(b)$ . The symmetry assertion is proved as follows: The FLT is even in its first argument, so that  $(B(1), Q(b))$  is equal in distribution to  $(-B(1), Q(b))$ , since the FLT characterizes the bivariate distribution (Fitzsimmons and McElroy, 2010, Thm. 1(i)). From the symmetry of the joint density function in its first argument follows the symmetry of the density of  $B(1)/\sqrt{Q(b)}$ . For continuity in  $\beta \in (-1, 1)$ , we first show weak convergence of any sequence of pairs  $(B(1), Q(b))$  depending on a sequence of  $\beta_n$ s using Theorem 1 (ii) of Fitzsimmons and McElroy (2010); then continuity of the cumulative distribution functions follows. ■