Title

Distribution Theory for the Studentized Mean for Long, Short, and Negative Memory Time Series

Authors

McElroy, Tucker S
Politis, D N

Publication Date

2011-09-01

Distribution Theory for the Studentized Mean for Long, Short, and Negative Memory Time Series

Tucker McElroy
Statistical Research Division
U.S. Census Bureau
4600 Silver Hill Road
Washington, D.C. 20233-9100
tucker.s.mcelroy@census.gov

Dimitris N. Politis*
Department of Mathematics
University of California at San Diego
La Jolla, CA 92093-0112
dpolitis@ucsd.edu

Abstract

We consider the problem of estimating the variance of the partial sums of a stationary time series that has either long memory, short memory, negative/intermediate memory, or is the first-difference of such a process. The rate of growth of this variance depends crucially on the type of memory, and we present results on the behavior of tapered sums of sample autocovariances in this context when the bandwidth vanishes asymptotically. We also present asymptotic results for the case that the bandwidth is a fixed proportion of sample size, extending known results to the case of flat-top tapers. We adopt the fixed proportion bandwidth perspective in our empirical section, presenting two methods for estimating the limiting critical values – both the subsampling method and a plug-in approach. Extensive simulation studies compare the size and power of both approaches as applied to hypothesis testing for the mean. Both methods perform well – although the subsampling method appears to be better sized – and provide a viable framework for conducting inference for the mean. In summary, we supply a unified asymptotic theory that covers all different types of memory under a single umbrella.

Keywords. Kernel, Lag-windows, Overdifferencing, Spectral estimation, Subsampling, Tapers, Unit-root problem.

*Research partially supported by NSF grants DMS-07-06732 and DMS-10-07513.

1 Introduction

Consider a sample $Y = \{Y_1, Y_2, \cdots, Y_n\}$ from a strictly stationary time series with mean $EY_t = \mu$, autocovariance $\gamma_h = Cov(Y_t, Y_{t+h})$, and integrable spectral density function $f(w) = \sum_h \gamma_h e^{-ihw}$. We are interested in studying the distribution of the studentized sample mean where the normalization involves the summation of sample autocovariances weighted by an arbitrary taper, and when the stochastic process exhibits either short or long memory or even when the process is overdifferenced. The latter case is especially tricky—and not well-studied in the literature—since in this case the studentization is achieved by dividing with a quantity that tends to zero. The objective in studentizing/self-normalizing the mean is the generation of a pivotal asymptotic distribution that can serve as the basis for the construction of confidence intervals and hypothesis tests for the unknown mean μ.

In the case that the autocovariances γ_h are (absolutely) summable with $\sum_h \gamma_h > 0$, it is well-known—under regularity conditions—that the sample mean $\overline{Y} = n^{-1} \sum_{t=1}^n Y_t$ is asymptotically normal with variance $f(0) = \sum_h \gamma_h$, i.e., $\sqrt{n}(\overline{Y} - \mu) \overset{\mathcal{L}}{\Longrightarrow} N(0, f(0))$. A consistent estimate of $f(0)$ is given by

$$W_{\Lambda,M} = \sum_{|h| \leq M} \Lambda_M(h) \widetilde{\gamma}_h, \tag{1}$$

where Λ_M is an arbitrary taper (described in Section 4), and $\widetilde{\gamma}_h$ are the sample autocovariance estimators defined by

$$\widetilde{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-|k|} (Y_{t+|k|} - \overline{Y})(Y_t - \overline{Y}) \qquad \text{for} \quad |k| < n. \tag{2}$$

As usual, $M = M(n)$ is a bandwidth parameter tending to ∞ as $n \to \infty$ but in such a way that $M << n$. Define the *bandwidth-fraction* to be $b = b(n) = M(n)/n$. The literature on such taper-based, "lag-window" spectral estimators is extensive, going back over fifty years; see e.g. Hannan (1970), Brillinger (1981), Priestley (1981), Rosenblatt (1985), Brockwell and Davis (1991), and the references therein. Also see Grenander and Rosenblatt (1957), Blackman and Tukey (1959), and Percival and Walden (1993). Much is already known about the classical case where $f(0)$ is bounded above and bounded away from zero, but also about the long memory case where $f(0)$ is infinite – see e.g. Beran (1994), Robinson (1994), and Palma (2007). Other recent literature includes Sun (2004) and Robinson (2005). However, little is yet known in the case that $f(0) = 0$, although this possibility was brought to the forefront early on by Rosenblatt (1961). We attempt to remedy this situation in the paper at hand. The case $f(0) = 0$ will be referred to henceforth as the superefficient (SE) case, since it implies $\overline{Y} = \mu + o_P(n^{-1/2})$, i.e., superefficient estimation of the mean.

The SE case has received little attention in the literature despite the fact that it is of some importance in applied econometrics. We illustrate this through a brief discussion of the unit root problem that arises from over-differencing. Although most economic time series exhibit obvious trends, much debate rages over whether processes are stationary or $I(1)$; witness the extensive literature on unit-root testing, starting with Dickey and Fuller (1979, 1981). Hamilton (1994) gives an overview; also see Phillips and Perron (1988) and Parker, Paparoditis, and Politis (2006) and the references therein. It is commonly felt that an economic time series is rarely $I(2)$, and yet many such $I(2)$ models are selected by automatic model identification software, such as X-12-ARIMA and TRAMO-SEATS (both of which often select the Box-Jenkins airline model, which is $I(2)$); see the discussion in Findley, Monsell, Bell, Otto, and Chen (1998), and Maravall and Caparello (2004).

Over-differencing relates to over-specification of the order of differencing, i.e., modeling a process as $I(1)$ when it is stationary, or as $I(2)$ when it is only $I(1)$. Naturally, estimation of parameters (such as the mean and other regression effects, but also maximum likelihood estimation of ARMA parameters) is performed on the differenced series, where the nonstationarity has been removed. But if the differencing order has been over-specified, then there will be over-differencing; this results in the spectrum of the differenced data being zero at frequency zero. For example, if Y_t is stationary with spectral density f but is viewed as $I(1)$, then the spectrum of the data's first difference is $|1 - e^{-i\lambda}|^2 f(\lambda)$. We mention in passing that similar issues exist, at least in theory, for seasonal frequencies, i.e., taking a seasonal difference when the seasonal component of the time series is actually stationary. In this case, the zero in the spectrum takes place at the seasonal frequencies corresponding to the angular portions of the zeroes of the seasonal differencing operator. Such zeroes offer no impediment to the estimation of the sample mean, but may generate other problems in model estimation; we do not pursue this point further here.

One approach to this problem is to do a pre-test of a possible unit root before differencing the data (Dickey and Fuller, 1979). Also see the treatment in Tanaka (1990, 1996). However, given that Type II errors will occur some of the time, we advocate in this paper the use of robust studentized sample mean estimates, where the robustness is with respect to the three basic cases for $f(0)$: infinite (long memory), finite and positive (short memory), or zero (the SE case, distinguished in what follows as either negative memory or differential memory). To that end, we study the finite-sample and asymptotic distributional properties of studentized sample means, where the normalization is of the form (1) and the stochastic process satisfies some very general conditions. Note that $W_{\Lambda,M}$ must mirror the properties of the variance of the sample mean under the three scenarios: for long memory, it must diverge at the appropriate rate; for short memory it should converge to the same constant; and for negative/differential memory it should tend to zero at the same rate. Our work will determine what conditions on a taper are needed in order to ensure such a robustness against different alternative memory scenarios.

Our main goal is to understand the joint distributional properties of $\overline{Y}$ and $W_{\Lambda,M}$ for various stochastic processes and various tapers, distinguishing between the case that the bandwidth-fraction b is vanishing and the case that it is a constant proportion. In the former case, we obtain central limit theorems for the sample mean, while the variance estimate tends in probability to a constant when appropriately normalized (Section 3). But in the latter case, the variance estimate tends to a random limit (Section 4). In Section 2 we give precise definitions of the four types of memory (LM, SM, NM, and DM), and also discuss some basic properties. Given the limit results, it is not obvious how to proceed with testing for the mean. This is addressed in Section 5 through two proposed methods, one for the vanishing bandwidth-fraction case and one for the fixed bandwidth-fraction case. These procedures are extensively evaluated through size and power simulations, covering many stochastic processes, tapers, and bandwidth-fractions. Section 6 summarizes our findings, and proofs are gathered in an Appendix.

2 Types of Memory

We now discuss in detail the different memory scenarios. By Short Memory (SM), we refer to the condition that the autocovariances are absolutely summable and their sum $f(0)$ is a nonzero constant. By Long Memory (LM), we mean that the autocovariances are not summable, and

$$\sum_{|k|\leq n} \gamma_k \sim C\, L(n) n^{\beta}. \tag{3}$$

Throughout $A_n \sim B_n$ denotes $A_n/B_n \to 1$ as $n \to \infty$. In (3) L is slowly varying at infinity (Embrechts, Klüppelberg, and Mikosch, 1997), with a limit that can be zero, one, or infinity. Also C is a positive constant. Most importantly, $\beta > 0$ (and is less than 1). In the SM case (3) applies with $\beta = 0$ and L tending to unity, so that $C = \sum_k \gamma_k$. The case that $\beta = 0$ but L tends to infinity is also LM (e.g., say $\gamma_k = k^{-1}$ for $k \geq 1$).

We will denote by Negative Memory (NM) the case of an absolutely summable autocovariance sequence such that (3) holds, but with $\beta < 0$ (though the case that $\beta = 0$ and L tends to zero is also NM). Some authors have used the term "intermediate memory" for this concept (Brockwell and Davis, 1991). Our nomenclature is due to the negative memory exponent, and also the result that most of the autocovariances are negative in this case (see Remark 2 below); the same conditions on L apply here. When the autocovariances are zero past a certain threshold, we obtain an example of Differential Memory (DM). For example, consider an $MA(q)$ data process that is over-differenced; the resulting autocovariances are identically zero for lags exceeding q, and $\sum_{|k|<n} \gamma_k = 0$ for $n > q + 1$. These definitions encompass ARFIMA models (Hosking, 1981), FEXP models (Beran (1993, 1994)), and fractional Gaussian Noise models, as well as the case of over-differenced processes. Some authors prefer to parametrize 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 an overview.

To distinguish between the LM, SM, NM, and DM memory cases, the key determinant in the limit theorems for $S_n = n\overline{Y}$ is the rate of growth of $V_n = Var(S_n)$, which in turn is related to

$\sum_{|k|\leq n}\gamma_k$. In the rest of this paper we study $S_n/\sqrt{\widehat{V_n}}$, where $\widehat{V_n}$ is some estimate of the variance V_n such that $\sqrt{\widehat{V_n}} = O_P(S_n)$. Let $W_n = \sum_{|k|\leq n}\gamma_k$ by definition, which in turn has asymptotic behavior given by (3); then we have the following identity:

$$V_n = \sum_{k=0}^{n-1} W_k \tag{4}$$

that is proved using summation by parts. Now for LM processes, W_n diverges, whereas in the SM case W_n tends to a nonzero constant. The superefficient case (SE) where $f(0) = 0$ is characterized by W_n tending to zero; however, we distinguish the case that W_n is summable (DM case) vs. not (NM case).

Definition 1 *Define the four types of memory as follows:*

- *(i) LM: W_n has asymptotics given by (3) with either $\beta \in (0,1)$ or $\beta = 0$ and L is tending to infinity.*
- *(ii) SM: $W_n \to C$ with $C > 0$.*
- *(iii) NM: W_n has asymptotics given by (3) with either $\beta < 0$ or $\beta = 0$ and L is tending to zero. When $\beta = -1$, suppose that W_n is not summable.*
- *(iv) DM: W_n is summable.*

The condition (3) on W_n in case (i) is denoted $LM(\beta)$, and the same condition on W_n in case (iii) is denoted $NM(\beta)$.

Remark 1 Note that cases (i), (ii), and (iii) are mutually exclusive, since case (ii) essentially corresponds to $\beta = 0$ in (3) with L tending to unity. When $\beta = -1$ in (3), it is possible for W_n to be either summable or not summable; the former case belongs to case (iv). If $\beta < -1$ in (3), then necessarily this is case (iv). Proposition 1 below shows that most of the autocovariances are negative in case (iii), and this is the reason for the term "negative memory." In case (iv) $\{Y_t\}$ can be represented as the difference of another stationary process (Proposition 2), and hence the nomenclature "differential memory." The case of an over-differenced series, as discussed in Section 1, is always included in case (iv). Both NM and DM have $f(0) = 0$, so both are part of the SE case.

Remark 2 These four cases are mutually exclusive but do not cover all possibilities, since we can consider a function L in (3) that is not slowly-varying (e.g., $L(n) = 1 + \sin(n)$). However, the four cases *are* exhaustive for processes satisfying (3) with L slowly-varying and $\beta < 1$.

In addition, we will consider the following condition on L, which imposes an eventual monotonicity on the function. By the representation for slowly-varying functions (Theorem $A3.3$ of Embrechts, Klüppelberg, and Mikosch, 1997), we can write L in the form

$$L(n) = c(n) \exp \int_z^n \eta(u)/u \, du, \tag{5}$$

with $c(n) \to c$ a positive constant, z some fixed positive constant, and $\eta(u)$ tending to zero as $u \to \infty$. Note that in (5) we can take the variable argument of L to be a continuous variable x, so that we can discuss the derivative of $L(x)$ – this will be denoted $\dot{L}(x)$. In general the function c needs only be measurable, but if it were smooth then clearly $\dot{c}(n) \to 0$; the required condition below is slightly stronger than that.

Condition M : c in (5) is continuously differentiable with $\dot{c}(n) = o(1/n)$.

From Condition M it follows that $n\dot{L}(n)/L(n) = o(1)$. We remark that the ARFIMA process satisfies this condition (actually, the slowly varying function can be taken to be a constant) by Theorem 13.2.2 of Brockwell and Davis (1991). The following result gives the behavior of V_n in each of the four cases described above, and also discusses the implied asymptotic behavior of γ_k.

Proposition 1 *Under condition M, we have*

$$V_n \sim \frac{1}{\beta+1} nW_n \sim \frac{C\, L(n) n^{\beta+1}}{\beta+1} \tag{6}$$

for cases (i) and (ii), and for $\beta > -1$ in case (iii). When $\beta = -1$ we have $V_n/(nW_n) \to \infty$ (so in a sense (6) holds true). In case (ii), we let $\beta = 0$ in the formula. In case (iv), $V_n \to \sum_{k \geq 0} W_k$. If $\beta \neq 0$ in cases (i) and (iii), we also have

$$\gamma_n \sim C \, \frac{\beta}{2} L(n) n^{\beta-1}. \tag{7}$$

In cases (i) and (iii) with $\beta = 0$ we have $\gamma_n \sim C \, \dot{L}(n)/2 = o(n^{-1})$.

This result shows that $V_n \to \infty$ in cases (i), (ii), and (iii), as long as $\beta > -1$. This will facilitate a fairly standard limit theorem for $\overline{Y}$ under some additional conditions. Case (iv) produces a very different sort of limit theorem; these results are discussed in Section 3 below.

Remark 3 Note that $L(n)$ must be non-negative for all n larger than some n_0, say; this follows from (6) and the fact that $V_n > 0$ for all n. Hence, for large n all the γ_n are *negative* in the NM case and *positive* in the LM case by (7). This is obvious from (7), but holds for the $\beta = 0$ case as well; in cases (i) or (iii) respectively we must have L tending to infinity or zero respectively, and hence asymptotically $\dot{L}(n)$ is positive or negative in the respective cases, and this determines the sign of γ_n. Essentially, NM is due to heavy negative correlation and LM to heavy positive correlation; this justifies the name "negative memory" for the NM case.

Example 1 Let $\beta > 0$ with $\gamma_h = (h+1)^{-\beta}$ for $h \geq 0$. This corresponds to a LM process. If we (temporally) difference the process, then the resulting ACF is

$$2(h+1)^{-\beta} - (h+2)^{-\beta} - h^{-\beta} = \left(-\frac{\beta(\beta+1)}{2}\right) h^{-(\beta+2)} + o(1).$$

Although this appears at first to be NM (comparing to (7)), in fact it can be shown that W_n is summable so that the differenced process is DM (see Corollary 1 below).

Example 2 Let $\beta < 0$ and $\gamma_h = -h^{\beta-1}$ for $h > 0$, and $\gamma_0 = 2\sum_{h\geq 1} h^{\beta-1}$. By (7) this seems to have the form of a NM ACF, but we must check the summability of W_n. Direct calculation shows that

$$W_n = 2\sum_{h>n} h^{\beta-1} = O(n^{\beta}).$$

Thus if $\beta \geq -1$ the process is NM, but otherwise is DM.

In case (ii) there is no result of the form (7), but in case (iv) the autocovariances have a particular structure if we suppose in addition that $\{Y_t\}$ is purely nondeterministic. This is discussed in the proposition below, whose result is similar to Theorem 8.6 of Bradley (2007).

Proposition 2 *Suppose that W_k is summable and $\{Y_t\}$ is purely nondeterministic. Then there exists a strictly stationary process $\{Z_t\}$ with autocovariance sequence r_k such that $Y_t = Z_t - Z_{t-1} + \mu$ and $\gamma_k = 2r_k - r_{k+1} - r_{k-1}$. Also $W_k = 2(r_k - r_{k+1})$ and $r_0 = \sum_{k\geq 0} W_k/2$.*

We note that the process $\{Z_t\}$ might be LM, SM or NM, even DM. The following corollary gives a converse statement.

Corollary 1 *Given any stationary process $\{Z_t\}$ with autocovariance function r_k tending to zero, the differenced process $Z_t - Z_{t-1}$ is DM, since $W_k = 2(r_k - r_{k+1})$ is summable.*

Example 3 Let $Z_t \sim$ i.i.d. and let $Y_t = Z_t - Z_{t-1}$. Then clearly Y_t is DM with autocorrelation of $-.5$ at lag one and zero at higher lags. If the innovation variance is unity, then $W_0 = 2$ and $W_k = 0$ for $k \geq 1$; this is clearly a summable sequence.

Example 4 A more interesting example goes as follows: let $Y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ where ϵ_t is white noise and $\psi_j = (-1)^j/j^p$ for $p > 0$ and $j \geq 1$, with $\psi_0 = -\sum_{j=1}^{\infty} \psi_j$ (which clearly exists by the alternating series test). Hence $\sum_j \psi_j = 0$, which implies that $\sum_k \gamma_k = 0$. Notice that the variance of Y_t only exists if $p > 1/2$, since

$$\sum_{j=0}^{\infty} \psi_j^2 = \psi_0^2 + \sum_{j=1}^{\infty} j^{-2p} = F^2(1/2, p) + F(1, 2p),$$

where $F(x, s)$ is the periodic-zeta function given by $F(x, s) = \sum_{n\geq 1} e^{2\pi i x n} n^{-s}$. So from now on suppose that $p > 1/2$, as we are not concerned with infinite variance time series in this paper. As in the proof of Proposition 2, define

$$\theta_j = \sum_{k=0}^{j} \psi_k = -\sum_{k=j+1}^{\infty} \frac{(-1)^k}{k^p}.$$

The sum of any two consecutive terms of this series, up to a minus sign, can be written as $k^{-p} - (k+1)^{-p} = ((1+1/k)^p - 1)/(k+1)^{-p}$, which by Taylor series expansion about zero yields an approximation of $p\,k^{-1}(k+1)^{-p}$ plus terms that decay at order k^{-2-p}. Thus asymptotically, the sum of consecutive terms in θ_j is $p\,k^{-p-1}$, plus other terms that decay even faster. So such a sequence is summable, and we find that $\theta_j = O(j^{-p})$ as $j \to \infty$. Since $p > 1/2$, this sequence is square summable. This implies that the time series $Z_t = \sum_{j\geq 0} \theta_j \epsilon_{t-j}$ is well-defined, i.e., is finite almost surely, since it has finite variance. Clearly $Y_t = Z_t - Z_{t-1}$, and the other assertions of Proposition 2 apply; in particular, Y_t is DM.

Example 5 Suppose that an observed time series X_t satisfies $(1-B)^d X_t = \epsilon_t$ for $d \in [0, 1]$, i.e., it is an ARFIMA $(0, d, 0)$. (This example can be easily generalized from ARFIMA, but it is easier to state things in this context.) If $d = 1$ this is just a random walk, and if $d \in [.5, 1)$ the process is said to have nonstationary long memory. If $d < .5$ the process is stationary, but with long memory if $d > 0$. Of course $d = 0$ corresponds to short memory (white noise). If the observed process is differenced once to produce $Y_t = X_t - X_{t-1}$, it is easy to see that the result is stationary with memory parameter $\beta = 2d - 2$. That is, if $d = 1$ we obtain short memory; if $d \in [.5, 1)$ we obtain a negative memory process of parameter $\beta \in [-1, 0)$; if $d < .5$ then we obtain a process with differential memory. The borderline case $d = .5$ is interesting: we don't get a differential memory process, since the original process is nonstationary – instead we get a negative memory process with $\beta = -1$ and nonsummable W_k sequence.

3 Limit Theory for the Case of Vanishing Bandwidth-Fraction

In the case that $b(n) \to 0$ as $n \to \infty$, we can treat the asymptotics of S_n and $W_{\Lambda,M}$ separately (recall that $W_{\Lambda,M}$ was defined in (1)), since the variance estimate always collapses to a constant, when appropriately normalized. Let us then consider the partial sums first; it is necessary to impose some additional assumptions. Typical assumptions for limit theorems involve either moment and mixing conditions, or linearity of the process involved. But limit theorems have also been derived under the assumption that the given process is a direct function of an underlying Gaussian process. The assumptions on our process $\{Y_t\}$ that we will consider are given below:

- **Process** P1. $\{Y_t\}$ is strongly mixing[1] with finite fourth moments, and is either (i) LM(β) with $\beta \in [0,1)$; (ii) SM; or (iii) NM(β) with $\beta \in (-1,0)$. Moreover, $\mathbb{E}[S_n^4] = O(V_n^2)$.

- **Process** P2. $\{Y_t\}$ is a linear process with square integrable i.i.d. inputs, and is either (i) LM(β) with $\beta \in [0,1)$; (ii) SM; or (iii) NM(β) with $\beta \in [-1,0]$.

- **Process** P3. $Y_t = g(X_t)$ for each t, where g is a function in $\mathbb{L}^2(\mathbb{R}, e^{-x^2/2})$ of Hermite rank[2] τ, and $\{X_t\}$ is a Gaussian process with autocovariance r_k. Assume that $\{Y_t\}$ is either (i) LM(β) with $\beta \in [0,1)$; (ii) SM; or (iii) is NM(β) with $\beta \in (-1,0)$. In case (i) we also assume that $(1-\beta)\tau < 1$, but in cases (ii) and (iii) there is no restriction on τ.

Each of the above three assumptions provides sufficient conditions for a limit theorem for S_n, as shown below. Note that these cover only cases of LM, SM, and NM; the DM case must be handled separately in what follows.

Remark 4 Central Limit Theorems can be established for processes P1 with weaker moment conditions – as shown in Rosenblatt (1956), only $2+\delta$ moments (for some arbitrarily small $\delta > 0$) are really required. We formulate P1 in terms of fourth moments since it gives a condition on S_n that is often straight-forward to verify. In particular, in the LM case, we might assume that there exists some $\epsilon \geq 1/2$ such that $\sum_{|k|<n} \alpha_k^{1-\epsilon} = O(n^\beta L(n))$ – then this relation holds for all smaller values of ϵ as well. This implies $\mathbb{E}[S_n^4] = O(n^{2+\beta} L(n))$ by Davydov's (1970) mixing inequality—see Lemma A.0.1 of Politis, Romano, and Wolf (1999). Thus $\mathbb{E}[S_n^4] = O(V_n^2)$. In the SM case, we only need require that the mixing coefficients are summable. The NM case might be handled similarly to the LM case, though now $\beta \in (-1,0)$. However, since mixing coefficients cannot be computed from data, these types of conditions are impossible to verify in practice. Note that if the fourth order cumulant (Taniguchi and Kakizawa, 2000) function is zero then $\mathbb{E}[S_n^4] = 3V_n^2$ by equation (18) of Rosenblatt (1961).

Now in the above, the $\beta = -1$ situation is only handled under P2. In the case that the autocovariances are summable (SM or NM(β) with $\beta \leq 0$), we have P2 $\Rightarrow$ P1 under some additional conditions (Pham and Tran, 1985); but when $\beta > 0$, the coefficients in the linear representation will not be summable, so that we cannot conclude the above implication. Also, it is known that Gaussian processes are strongly mixing iff their spectrum is bounded away from zero and infinity (Kolmogorov and Rozanov, 1960); since P3 involves fairly general functions of a Gaussian that can have unbounded spectrum, we cannot conclude that P3 $\Rightarrow$ P1. It is also clear that P2 and P3 cover distinct cases, since in the latter the process can be nonlinear, whereas in the former the marginal

[1] A stationary process $\{Y_t\}$ is strongly mixing (Rosenblatt, 1956) if $\alpha_k \to 0$ as $k \to \infty$ where $\alpha_k = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$, and $\mathcal{F}_j^m$ is the σ-algebra generated by $\{Y_k, j \leq k \leq m\}$.

[2] The definition of Hermite rank can be found in Taqqu (1975).

distribution can be quite general. The restriction on τ in the LM case of P3 is for convenience – limit results are also available for the case that $(1-\beta)\tau > 1$, though the boundary case $(1-\beta)\tau = 1$ cannot be handled without additional knowledge about $\{\gamma_k\}$.

Theorem 1 *Suppose that $\{Y_t\}$ is a strictly stationary process satisfying Condition M and one of P1, P2, or P3. Then $\frac{S_n - n\mu}{\sqrt{V_n}} \stackrel{\mathcal{L}}{\Longrightarrow} Q$ as $n \to \infty$, where Q is an absolutely continuous random variable that is standard normal except when $\tau > 1$ in the LM case of P3.*

The above theorem covers the cases of LM, SM, and NM under the three different conditions P1, P2, and P3. We turn next to the DM process, and consider only purely non-deterministic processes so that Proposition 2 can be applied. We consider two assumptions on the $\{Z_t\}$ process: either it is linear (like the P2 case) or is strong mixing (like the P1 case). Then we have the following result.

Theorem 2 *Suppose that $\{Y_t\}$ is a strictly stationary purely non-deterministic process with summable W_k (so the second moments are assumed to exist). If the process $\{Z_t\}$ of Proposition 2 is either linear (like the P2 case) or strongly mixing (like the P1 case), then $\frac{S_n - n\mu}{\sqrt{V_n}} \stackrel{\mathcal{L}}{\Longrightarrow} (Z_* - Z_0)/\sqrt{2r_0}$ as $n \to \infty$, where Z_* is a random variable equal in distribution to Z_0, but independent of it.*

So much for the asymptotics of S_n for vanishing bandwidth-fraction. We now turn to the behavior of $W_{\Lambda,M}$, which depends on the type of taper as well as the type of memory of the process. The tapers that we consider are very general: Λ_M is a piecewise smooth (i.e., piecewise differentiable), even function on the integers such that $\Lambda_M(h) = 0$ for $|h| \geq M$. Letting U_M denote the maximum value of $\Lambda_M(h)$ for all h, we suppose that U_M does not grow too fast as $M \to \infty$. Classical tapers are bounded, in which case U_M can be taken constant. The triangular (Bartlett) kernel, the trapezoidal (Politis and Romano, 1995), and the more general flat-top kernels (Politis 2001, 2005) all satisfy these conditions. We begin by decomposing $W_{\Lambda,M}$:

$$\begin{aligned} W_{\Lambda,M} &= \widetilde{W}_M + E_M^{(1)} + E_M^{(2)} \\ \widetilde{W}_M &= \sum_h \Lambda_M(h) \left(1 - \frac{|h|}{n}\right) \gamma_h \\ E_M^{(1)} &= \sum_h \Lambda_M(h) \left(\widetilde{\gamma}_h - \overline{\gamma}_h\right) \\ E_M^{(2)} &= \sum_h \Lambda_M(h) \left(\overline{\gamma}_h - \gamma_h\right), \end{aligned} \tag{8}$$

where $\overline{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-|k|}(Y_t - \mu)(Y_{t+k} - \mu)$ for $k \geq 0$, and recall $\widetilde{\gamma}_k$ is given by (2). The terms $E_M^{(1)}$ and $E_M^{(2)}$ are stochastic errors, whereas $\widetilde{W}_M$ is deterministic and serves as an approximation to W_M. Proposition 3 below shows that the stochastic errors $E_M^{(1)}$ and $E_M^{(2)}$ are negligible under some mild hypotheses. One assumption we utilize parallels Assumption A of Andrews (1991):

Assumption B: $\mathbb{E}|Y_t^4| < \infty$ and the fourth order cumulant $cum(Y_0, Y_j, Y_k, Y_h)$ is absolutely summable, i.e. $\sum_{j,k,h} |cum(Y_0, Y_j, Y_k, Y_h)| < \infty$.

The definition of the cumulant can be found in Taniguchi and Kakizawa (2000). The above assumption is compatible with the process conditions P1, P2, and P3 for the LM, SM, and NM cases. Assumption B is also compatible with the DM case. As discussed in Andrews (1991), linear processes with absolutely summable coefficients and finite fourth moments satisfy Assumption B, even if the process is LM. As Lemma 1 of Andrews (1991) shows, Assumption B is also implied by a strong mixing plus moments condition.

Proposition 3 *Suppose that one of cases (i) – (iv) of Definition 1 in Section* 2 *hold, as well as Assumption B and Condition M. Also suppose that* $b(n) + 1/M(n) \to 0$ *as* $n \to \infty$*. Then, the error terms* $E_M^{(1)}$ *and* $E_M^{(2)}$ *of (8) satisfy* $E_M^{(1)} = O_P(U_M\, M V_n n^{-2})$ *and* $E_M^{(2)} = O_P(U_M\, M n^{-\eta} K(n))$ *as* $n \to \infty$*, where* K *is slowly-varying and* $\eta = 1/2$ *if* $\beta < 1/2$ *and* $\eta = 1 - \beta$ *if* $\beta \geq 1/2$ *(in the LM case).*

Next we consider the asymptotics of the deterministic term $\widetilde{W}_M = \sum_h \Lambda_M(h)(1 - |h|/n)\gamma_h$, which depend upon the memory assumptions. The following result summarizes the various cases that can occur for tapers of the form $\Lambda_M(h) = \Lambda(h/M)$ for a fixed function $\Lambda(x)$. We assume Λ is an even, piecewise smooth function, that is real-analytic on every such interval; by $\dot{\Lambda}_+(x)$ for $x \geq 0$, we denote the derivative from the right (and $\Lambda_+^{(j)}(x)$ for higher order derivatives). An example is a "flat-top" taper where there exists an interval $[0, c]$ (with $c > 0$) for which Λ equals unity – see Politis (2001).

Theorem 3 *Let* $\Lambda(x)$ *be an even, piecewise differentiable function supported on* $[-1, 1]$*, with* $\Lambda_M(h) = \Lambda(h/M)$*. Suppose that* $b + 1/M \to 0$ *as* $n \to \infty$ *and Condition M. For cases (i), (ii), and (iii) of Definition 1 for any* $\beta \in (-1, 1)$*, we have*

$$\frac{\sum_h \Lambda_M(h) \left(1 - \frac{|h|}{n}\right) \gamma_h}{C M^\beta L(M)} \sim \frac{\widetilde{W}_M}{W_M} \to \zeta \equiv - \int_0^1 \dot{\Lambda}_+(x) x^\beta \, dx \;\; as \; M \to \infty. \tag{9}$$

With the truncation taper $\Lambda = 1_{[-1,1]}$*,* ζ *in (9) is instead given by one for all* $\beta \in (-1, 1)$*.*
In the DM case we instead have

$$\widetilde{W}_M \sim -2\, r_0\, \dot{\Lambda}_+(0)\, M^{-1} \left(1_{\{c=0\}} + o(1)\right).$$

In the SM case recall that $C \sim \sum_k \gamma_k$ and $L(M) \equiv 1$, so that (9) with $\beta = 0$ yields $\zeta = \Lambda(0)$; this equals unity as long as $\Lambda(0) = 1$, which is commonly assumed. In case (iv) for non-flat-top kernels, the overall error is controlled by the first derivative $\dot{\Lambda}_+(0)$. This is because $c = 0$; but when $c > 0$, the rate of decay is even faster, and is hard to describe in a general result. The quantity ζ in (9) is equal to one plus the relative bias $(\widetilde{W}_M - W_M)/W_M$, measuring the asymptotic discrepancy between our variance estimate $W_{\Lambda,M}$ and the sequence W_M. We refer to ζ as the *quotient bias* hereafter. Note that ζ is well-defined, since the derivative of Λ exists almost everywhere.

Remark 5 It follows from Proposition 3, Theorem 3, and Proposition 1 – when their respective conditions are satisfied – that in probability

$$
\begin{aligned}
MW_{\Lambda,M} &\sim (\beta+1)\zeta\, V_M \\
MW_{\Lambda,M} &\sim -2\, r_0\, \dot{\Lambda}_+(0)\ \left(1_{\{c=0\}} + o(1)\right)
\end{aligned}
\tag{10}
$$

for the cases of LM/SM/NM and DM respectively. Although $V_M = Var(S_M)$ is of some interest, we really need to obtain a quantity asymptotic to V_n. Unfortunately, since $b(n) = M(n)/n \to 0$, we have $V_M/V_n \to 0$ except in the DM case. So we cannot normalize S_n by $\sqrt{MW_{\Lambda,M}}$ in the LM/SM/NM case, since the normalization rate is not correct (also there is the matter of the unknown $(\beta+1)\zeta$ factor). In the DM case, if we use a non-flat-top taper, we can indeed utilize $S_n\left(-MW_{\Lambda,M}/\dot{\Lambda}_+(0)\right)^{-1/2}$ since $2r_0 = \sum_{k\geq 0} W_k = V_\infty$; by Theorem 2, this studentized statistic converges to $Z_* - Z_0$.

Now if $\beta > 0$, then ζ can be rewritten as $\beta \int_0^1 \Lambda(x)x^{\beta-1}\,dx$ using integration by parts piecewise. In this case, we proceed to calculate ζ for some commonly used tapers. The simplest flat-top taper is the trapezoidal taper of Politis and Romano (1995) given by

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

with $c \in (0,1]$. Then it follows that

$$
\zeta^{T,c} = -\int_0^1 \dot{\Lambda}^{T,c}(x)x^\beta\,dx = \frac{1-c^{\beta+1}}{(1-c)(1+\beta)}.
$$

Also in the DM case, it can be shown rather easily that $M\widetilde{W}_M \sim \frac{2}{1-c}(r_{[cM]} - r_M)$, which tends to zero at a rate that depends upon the autocovariance sequence $\{r_k\}$ and the truncation c. The triangular (Bartlett's) taper is obtained as the limiting case of $\Lambda^{T,c}$ as $c \to 0$; in this case, $\zeta^{Bar} = 1/(\beta+1)$. Interestingly, the factor $(1+\beta)\zeta$ appearing in (10) is then unity for any $\beta \in (-1,1)$. Also since $\dot{\Lambda}_+(0) = -1$ for the Bartlett, we have $MW_{\Lambda,M} \sim V_M$ for all cases, including DM.

The asymptotic quotient bias ζ is also easily computed for the Parzen taper, given by

$$
\Lambda^{Par}(x) = \begin{cases} 1 - 6|x|^2 + 6|x|^3 & \text{if } |x| \leq 1/2 \\ 2(1-|x|)^2 & \text{if } 1/2 < |x| \leq 1 \\ 0 & \text{if } |x| > 1 \end{cases}.
$$

Then $\zeta^{Par} = (2 - (1/2)^\beta)(3/(\beta+3) - 6/(\beta+2) + 3/(\beta+1))$.

4 Limit Theory for the Case of Fixed Bandwidth-Fraction

The vanishing bandwidth-fraction results of Theorem 3 and Remark 5 indicate a difficulty with using tapers when LM or NM is present, since it is difficult to capture the correct rate for all types of memory. In fact, there is an asymptotic distortion equal to the quotient bias ζ that depends on the unknown β parameter. In addition, there is the presence of the slowly-varying function L. These problems can be resolved by using a fixed bandwidth-fraction approach, which is described in this section.

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 stress that in this section b is a fixed number, and does not grow with n as in the previous Section. Let $\widehat{S}_i = \sum_{t=1}^{i}(Y_t - \overline{Y})$ (so that $\widehat{S}_n = 0$). A derivative of Λ from the left is denoted $\dot{\Lambda}_-$, whereas the second derivative is $\ddot{\Lambda}$. The greatest integer function is denoted by $[\cdot]$. We consider tapers from the following family:

$$\{\Lambda \text{ is even and equals unity for } |x| \leq c,\, c \in [0, 1). \quad \text{Furthermore, } \Lambda \text{ is supported on } [-1, 1], \text{ is continuous, and is twice continuously differentiable on } (c, 1) \cup (-1, -c).\} \tag{11}$$

This assumption is slightly less restrictive than the conditions of Theorem 3, since we only require two continuous derivatives. This family of tapers includes the family of 'flat-top' kernels of Politis (2005)where $c > 0$, as well as the Bartlett kernel (letting $c = 0$ and a linear decay of Λ), and other kernels considered in Kiefer and Vogelsang (2005). The following result was proved in McElroy and Politis (2009), restated here for convenience.

Proposition 4 *(McElroy and Politis, 2009) Let Λ be a kernel from family (11). Let $b \in (0, 1]$ be a constant bandwidth-fraction. Then*

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

Remark 6 In case the taper 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

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

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

In order to apply this result, we need functional limit theorems for the partial sums, since $\widehat{S}_i = S_i - i/n\, S_n$. For the LM, SM, and NM cases such limit theorems can be proved which extend Theorem 1 under more restrictive conditions; the DM case is treated separately.

Theorem 4 *Suppose that $\{Y_t\}$ is a strictly stationary process satisfying Condition M as well as either P1, P2, or P3, with $\beta \in (-1, 1)$. In the case of a P3 process with $\beta > 0$, also assume that $\tau = 1$. Then as $n \to \infty$*

$$V_n^{-1/2} \left(S_{[nr]} - [nr]\mu \right) \stackrel{\mathcal{L}}{\Longrightarrow} B(r) \tag{12}$$

in the sense that the corresponding probability measures on $C[0, 1]$ (the space of continuous functions on $[0, 1]$) converge weakly. $B(\cdot)$ is a Fractional Brownian Motion (FBM) process of parameter β.

Remark 7 For the linear case P2, Davydov (1970) and Gorodetskii (1977) provide a proof of this result requiring higher moments. Marinucci and Robinson (2000) relax the requirement to $2 + \delta$ moments for some $\delta > 0$. In that work, the moment condition is used to establish tightness of the stochastic process, but uses a sufficient condition for tightness that is less convenient then the formulation of Problem 4.11 of Karatzas and Shreve (1991). This latter formulation allows us to require only a second moment.

FBM is defined in Samorodnitsky and Taqqu (1994). It follows from Theorem 4 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 putting Proposition 4 and Theorem 4 together with treatment of the DM case yields the following result.

Theorem 5 *Let Λ be a kernel from family (11), and let the bandwidth $M = bn$. Let $b \in (0, 1]$ be a constant bandwidth-fraction. If $\{Y_t\}$ is LM, SM, or NM assume the conditions of Theorem 4. Then*

$$\frac{S_n - n\mu}{\sqrt{nW_{\Lambda,M}}} \stackrel{\mathcal{L}}{\Longrightarrow} \frac{B(1)}{\sqrt{Q(b)}} \tag{13}$$

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

$$\begin{aligned} &- \frac{2}{b}\dot{\Lambda}_+(c) \int_0^{1-cb} \widetilde{B}(r)\widetilde{B}(r + cb)\, dr - \frac{1}{b^2} \int_{cb<|r-s|<b} \widetilde{B}(r)\widetilde{B}(s)\ddot{\Lambda} \left(\frac{|r - s|}{b} \right) drds \\ &+ \frac{2}{b}\dot{\Lambda}_-(1) \int_0^{1-b} \widetilde{B}(r)\widetilde{B}(r + b)\, dr. \end{aligned} \tag{14}$$

If instead the process is DM, assume the conditions of Theorem 2. Let $\widetilde{C}(r) = (rZ_* + (1-r)Z_0)/\sqrt{r_0}$; here Z_* is a random variable equal in distribution to Z_0, but independent of it, and recall that r_0 is the variance of the $\{Z_t\}$ process. Then

$$\frac{S_n - n\mu}{\sqrt{nW_{\Lambda,M}}} \stackrel{\mathcal{L}}{\Longrightarrow} \frac{\widetilde{C}(1) - \widetilde{C}(0)}{\sqrt{P(b)}} \tag{15}$$

as $n \to \infty$, where $P(b)$ is defined as the sum of $-2b^{-1}\dot{\Lambda}_+(0)1_{\{c=0\}}$ and $Q(b)$ given by (14), substituting $\widetilde{C}$ for $\widetilde{B}$.

Remark 8 The first result (13) was derived in a preliminary calculation in McElroy and Politis (2009), and the distribution (14) has been tabulated. The joint distribution of $B(1)$ and $Q(b)$ was also explored in McElroy and Politis (2011) through the device of the joint Fourier-Laplace Transform. The DM case (15) is novel here. Note that the first term of $P(b)$ in the $c = 0$ case, i.e., $-2b^{-1}\dot{\Lambda}_+(0)$ agrees with the first term in the expansion of $\widetilde{W}_M$ in the DM case of Theorem 3, since in the vanishing bandwidth-fraction case we have $nW_{\Lambda,M} \sim -2r_0\frac{n}{M}\dot{\Lambda}_+(0)$ plus higher order terms.

Remark 9 The expression for $P(b)$ in the DM case really reduces to a quadratic in $Z_*/\sqrt{r_0}$ and $Z_0/\sqrt{r_0}$, the coefficients of which can be calculated in terms of the taper's derivatives. However, knowing these quantities is not helpful towards understanding the distribution of $P(b)$, since the distribution of Z_* is unknown.

Unlike the special case studied by Kiefer, Vogelsang, and Bunzel (2000), the numerator $B(1)$ of (13) is not independent of the denominator $Q(b)$ if $\beta \neq 0$. To elaborate, Kiefer et al. (2000) considered the case $b = 1$ and $c = 0$, the kernel is the Bartlett, and $\beta = 0$ (although later work by Kiefer and Vogelsang (2002, 2005) generalizes to $b < 1$). Then $Q(1) = 2\int_0^1 \widetilde{B}^2(r)\,dr$, and the authors note that $B(1)$ is independent of $\widetilde{B}(r)$. As shown in McElroy and Politis (2011), this is true for other kernels as long as $\beta = 0$; however, if $\beta \neq 0$, then $B(1)$ and $Q(b)$ are dependent. Fortunately, it is a simple matter to determine the limiting distribution numerically for any given value of β, and any choice of taper and bandwidth fraction b.

5 Applications and Numerical Studies

The preceding two sections give two different perspectives on the asymptotic behavior of taper-normalized sample means. If we normalize $S_n - n\mu$ by $\sqrt{nW_{\Lambda,M}}$ or by $\sqrt{MW_{\Lambda,M}}$, the studentized statistic does not converge – except in the SM and DM cases, respectively for the two normalizations – when adopting the vanishing bandwidth-fraction perspective (see Remark 5). However, $(S_n - n\mu)/\sqrt{nW_{\Lambda,M}}$ converges to a nondegenerate distribution under all types of memory by Theorem 5, when adopting the fixed bandwidth-fraction approach. Since bandwidth choice is up to the practitioner, it appears that the fixed bandwidth-fraction viewpoint might be preferrable in our attempt towards a unified treatment of inference for the mean that is valid in all kinds of set-ups.

However, the limit distribution of the studentized sample mean will generally depend on β. Either one must estimate the nuisance parameters – including β – or a nonparametric technique such as the bootstrap or subsampling (Politis, Romano, and Wolf (1999)) must be utilized to get the limit quantiles. The parametric bootstrap is not feasible here (since no model is specified for the data in our context) and the block bootstrap tends to perform badly when autocorrelation dies gradually (Lahiri, 2003). However, given that the limit distribution has been tabulated for some values of $\beta \in (-1,1)$ and some popular tapers (cf. McElroy and Politis, 2009), one can utilize a plug-in estimator of β instead; this is similar in spirit to the approach advocated in Robinson (2005).

These two techniques are described in more detail below, along with statistical justification. A new estimator of β, based on the rate estimation ideas of McElroy and Politis (2007), is also discussed. Then in the following subsection, both methods are applied to the study of size and power for testing the null hypothesis that $\mu = 0$. We look at Gaussian processes exhibiting LM, SM, NM, or DM, at a variety of sample sizes and choices of taper.

5.1 Subsampling Methodology for Obtaining Critical Values

Firstly, we consider the subsampling method applied to the statistic $S_n/\sqrt{nW_{\Lambda,M}}$, or equivalently, $\overline{Y}/\sqrt{W_{\Lambda,M}/n}$. Letting $M = bn$ as usual with b fixed and constant, under the assumptions of Theorem 5 we obtain the nondegenerate limit distribution $B(1)/\sqrt{Q(b)}$ for the LM/SM/NM case, and $[\widetilde{C}(1) - \widetilde{C}(0)]/\sqrt{P(b)}$ in the DM case. The subsampling distribution estimator (sde) will be consistent for this limit distribution under a mixing condition, such as strong mixing (see Section 3). Since long memory Gaussian time series cannot be strong mixing, a more appropriate framework is that of weak dependence as formulated in Doukhan and Louhichi (1999); see alo Bardet et al. (2008). In Jach, McElroy, and Politis (2011) the Doukhan—Louhichi dependence condition is shown to be sufficient to establish consistency of subsampling distribution estimators for studentized statistics; Ango-Nze et al. (2003) was an earlier work on subsampling under weak dependence.

As for the sde itself, we first select a subsampling blocksize an, where a is the subsampling-fraction; as usual, this is assumed to be vanishing, i.e., $a = a(n) \to 0$, though $an \to \infty$. Then $n - an + 1 = (1-a)n + 1$ contiguous overlapping blocks of the time series are constructed, and the statistics $\overline{Y}$ and $W_{\Lambda,bn}$ are evaluated on each block. This means a sample mean over an random variables, and the corresponding tapered variance estimate based on this subsample, so that the bandwidth is actually abn rather than bn. As a practical matter, unless b and a are taken fairly large, the bandwidth abn becomes unmanageably small. The subsampled statistics can be collected into a set:

$$\left\{ \frac{\overline{Y}_{an,i} - \overline{Y}_n}{\sqrt{W_{\Lambda,abn,i}/(an)}} \text{ for } i = 1, \ldots, n - an + 1. \right\}$$

Now taking the order statistics on this collection produces the quantiles of the sde. Further details of the construction can be found in Politis, Romano, and Wolf (1999), but we here sketch the remaining theoretical details. First we note a general result that follows from Theorem 3 and

Remark 5: if $M/n \to 0$ as $n \to \infty$, then

$$\frac{V_n/n^2}{\widetilde{W}_M/M} \to 0. \tag{16}$$

This is true for the LM/SM/NM case, since by Proposition 1 the limit is $O([M/n]^{1-\beta})$. (16) is also true for the DM case if a non-flat-top taper (i.e., $c = 0$) is used, since the above limit will be asymptotic to $-M/\dot{\Lambda}_+(0)n$. (The result is not guaranteed for flat-top tapers; this will depend on the rate that $M/n \to 0$, versus the rate of decay of $\{r_k\}$.)

Therefore, the probability that the ith subsample statistic exceeds a given x is

$$\mathbb{P}\left[\sqrt{an}\,\frac{\overline{Y}_{an,i} - \overline{Y}_n}{\sqrt{W_{\Lambda,abn,i}}} > x\right] = \mathbb{P}\left[\frac{S_{an} - an\,\mu}{\sqrt{anW_{\Lambda,abn}}} - b^{-1/2}\,\frac{S_n - n\mu}{\sqrt{V_n}}\sqrt{\frac{V_n/n^2}{W_{\Lambda,abn}/(abn)}} > x\right].$$

Using Theorems 1 and 2, Proposition 3, and (16), the second term in the probability will tend to zero. Hence (in the LM/SM/NM case)

$$\mathbb{P}\left[\frac{S_{an} - an\,\mu}{\sqrt{anW_{\Lambda,abn}}} > x\right] \to \mathbb{P}\left[\frac{B(1)}{\sqrt{Q(b)}} > x\right]$$

as $n \to \infty$, noting that $an \to \infty$ by assumption (and b is fixed). In the DM case the limit is $[\widetilde{C}(1) - \widetilde{C}(0)]/\sqrt{P(b)}$. Now, as long as the ith subsample statistic is approximately independent to the jth one when $|i - j|$ is large – which is implied by the weak dependence condition – the sde is consistent for the target limit distribution, and subsampling is valid. Note that we assume the fixed bandwidth-fraction condition for this result, but end up utilizing some of the vanishing bandwidth-fraction results, since the actual bandwidth-fraction for the subsampled tapered variance estimate is the vanishing quantity ab.

5.2 Plug-in Methodology for Obtaining Critical Values

Alternatively, one can use a plug-in approach as in McElroy and Politis (2011). Adopting the fixed bandwidth-fraction asymptotics (so assume b is constant throughout), and assuming that $\beta \in (-1, 1)$ (so that the DM case is explicitly excluded), we proceed to estimate the quantiles of the limit distribution $B(1)/\sqrt{Q(b)}$ via first estimating β from the data, and then utilizing $x_\alpha(\widehat{\beta})$, where $x_\alpha(\beta)$ is the upper right α quantile of $B(1)/\sqrt{Q(b)}$. That is,

$$\mathbb{P}\left[\frac{B(1)}{\sqrt{Q(b)}} > x_\alpha(\beta)\right] = \alpha.$$

These quantiles have been tabulated for $\beta \in B = \{-.8, -.6, -.4, -.2, 0, .2, .4, .6, .8\}$, three values of α, several commonly used tapers, and all values of b (via regression) – see the tables in McElroy and Politis (2009). Since the distribution is continuous in β, any consistent estimate $\widehat{\beta}$ can be utilized. Then one finds the member of B closest to the given $\widehat{\beta}$ – call this $\widetilde{\beta}$ – and utilizes $x_\alpha(\widetilde{\beta})$.

This will be called the empirical plug-in method. Clearly, a finer mesh of simulation values for B would improve the procedure, but we may yet expect to obtain results superior to just using $\beta = 0$ in ignorance of the true memory. This latter approach, which essentially assumes that only short memory is present, will be referred to as the default plug-in method, and will be utilized as a benchmark for the empirical plug-in method.

Now many nonparametric estimators of β can be utilized. To fix ideas, we now propose a simple estimator that will work in our context, namely:

$$\widehat{\beta} = \frac{\log W_{\Lambda,bn}}{\log n}. \tag{17}$$

This is consistent for β under the assumptions common to this paper.

Proposition 5 *Assume that $b \in (0,1]$ is fixed, as well as the condition M and the hypotheses of Theorem 5. Then $\widehat{\beta}$ defined by (17) converges in probability to β when $\{Y_t\}$ is LM, SM, or NM. When $\{Y_t\}$ is DM, $\widehat{\beta} \xrightarrow{P} -1$.*

This estimator is similar in spirit to the tail index estimator of Meerschaert and Scheffler (1998), since it is based upon a convergence rate. The performance of $\widehat{\beta}$ can be poor in finite sample when long memory is present, but it is quite versatile and simple to implement. Note that $\widehat{\beta} \xrightarrow{P} -1$ in the DM case, but since $x_\alpha(-1)$ does *not* correspond to the quantile of $[\widetilde{C}(1) - \widetilde{C}(0)]/\sqrt{P(b)}$, we must exclude the DM case by assumption to avoid an inconsistent procedure.

5.3 Size and Power of Methods

We next evaluate the two methodologies – subsampling and empirical plug-in – through simulations. We adopt the perspective of testing a null hypothesis of $\mu = \mu_0$, since it is easier to evaluate the finite-sample properties through size and power, as opposed to the confidence interval perspective. In the simulations we take $\mu_0 = 0$. For each method, we consider values of μ between 0 and 1, with $\mu = 0$ corresponding to the null hypothesis. We compute the statistic $S_n/\sqrt{nW_{\Lambda,bn}}$ for a few different values of b, and five different tapers. The critical values of the limit distribution can be approximated using either of the two methods described above (though the plug-in method cannot be used with DM processes, and we cannot justify the use of flat-top tapers in the DM case). Then we record the proportion of times that the statistic exceeds these critical values, using a two-sided test. Note that when $\mu > 0$, this assessment is interpreted as empirical power, but when $\mu = 0$ we obtain the empirical size.

The size of the plug-in approach has been partially addressed in McElroy and Politis (2009, 2011). However, the subsampling method's performance has not been previously studied, so we provide some additional material regarding its size. Tables 1 through 10 display size results for various Type I error rates, for a two-sided test: $\alpha = .10, .05, .01$. The sampling fraction for the subsampling method was selected at values $a = .2, .1, .04$. Also we have the five tapers – Bartlett, Trapezoid with $c = .25$, Trapezoid with $c = .5$, Parzen, and Daniell – with three choices of bandwidth fraction $b = .1, .5, 1$. The sample size is varied from $n = 250, 500, 1000$. Results are for the empirical coverage, and so the target for the columns are the values .90, .95, .99.

For data generation processes we focus on the Gaussian distribution, and consider simple white noise for the SM case (since serially correlated processes have been considered in previous literature, it suffices to take the simple white noise case); we consider four NM processes with $\beta = -.2, -.4, -.6, -.8$, where the autocovariance function is determined from Example 2 of Section 2. Also there are four LM processes, with $\beta = .2, .4, .6, .8$ and autocovariance function given in Example 1 of Section 2. Finally, our tenth DGP is DM, generated by the first difference of a white noise process. Informally, we will refer to this via $\beta = -1$, by an abuse of notation. We generated 1000 simulations of each specification.

The power surfaces are organized a bit differently. We focus on one sample size $n = 250$ and one Type I error rate $\alpha = .05$ (for a two-sided test for the mean, so we use the upper one-sided critical value at .975 from the subsampling distributions and the tabulated values). Restricting to this α value gives the general behavior, and a reasonable sense of the power can be gleaned from the $n = 250$ case – higher sample sizes tend to shift the contours upwards (not dramatically for high β), but the overall shape is the same. We consider the same tapers (Bartlett, Trapezoid (.25), Trapezoid (.5), Parzen, and Daniell), but with bandwidth fractions $b = .2, .5, 1$. The sampling fractions are $a = .04, .12, .2$. These choices are convenient, as it is always guaranteed that abn is an integer. The range of μ was chosen so as to capture the main qualitative features of the power surface across all DGPs: $\mu \in \{j/20\}_{j=0}^{19}$ proved to yield power close to 50 percent for the long memory DGPs, while being small enough to allow visual discrimination of cases.

The six methods are placed in each figure as sub-panels. Moving from top left to bottom right, the first three methods correspond to subsampling with various sampling fractions. Then we have the empirical plug-in method, followed by the default plug-in (which uses $\beta = 0$ critical values throughout) method. Note that these methods are greatly flawed when the DGP is DM, but we present the results anyways for thoroughness. The final panel is an omniscient plug-in, based on knowing the true value of β (so it is not a practicable method, but is helpful for understanding power). The discrepancy in power between this and the empirical plug-in method (middle right panels) is mainly due to error in our estimator of β.

Now we discuss these numerical results. The size results for the SM case are fairly standard, being adequate at sample size $n = 500$ and greater; the higher bandwidth fractions gave better results. At $n = 1000$ all the tapers were roughly comparable in performance. For NM the coverage improved for higher values of β; the flat-top tapers perform well, and the Parzen is also tolerable for higher bandwidths. The Bartlett fared poorly except for the $\beta = -.2$ case. The results for the LM case were much worse, with poor coverage at $\beta = .2$; higher bandwidths seemed to improve things slightly. At $\beta = .4$ the results are best with the $b = .05$ bandwidth, but for $\beta = .6, .8$ the coverage deteriorates. Finally, for the DM case it seems that a smaller bandwidth is preferable. The Bartlett taper performs badly, but the Trapezoidal (.5) and Truncated tapers are adequate (except for $b = .1$). In summary, if there is DM, NM, or SM, then one can use a flat-top (with higher value of c) or Parzen taper with a middle bandwidth value, such as $b = .05$, and obtain adequate coverage for larger samples.

If a statistic rejects too often under the null hypothesis – as is seen to happen in the tables for the subsampling methods – then it is liable to have higher power than otherwise. The fact that all methods tend to be over-sized is evident in the surface plots in Figures 1 through 15 by examination of the $\mu = 0$ cross-sectional curve towards the right side of the surface. But as μ increases, the DM and NM DGPs generate high power relatively quickly, giving a mesa shape to the surfaces. For SM and weaker LM, the rise to full power is slower. An ironic feature is that when μ is quite low, the power for strong LM is better than for weaker LM, essentially due to the methods being over-sized. This is seen in the "ruffle" feature of the curves along the $\beta = .8$ cross-section.[3] Since power approaches 50 percent for all DGPs as μ increases to unity, there is the question of how this is meaningful relative to the variation in the process. All were constructed with $\gamma_0 = 1$ (an alternative way to normalize is to set the innovation variances equal, by in each case dividing through by the square root of the integral of the log spectrum), so the coefficient of variation is $1/\mu$ for all DGPs. Primarily, we view these figures as a way to contrast the power of methods and tapers.

6 Conclusion

This paper sets out a thorough study of self-normalized mean estimation when long memory or negative memory is present. The main statistic of interest is the sample mean of a stationary time series (appropriately differenced beforehand), normalized by a tapered sum of sample autocovariances. The behavior of autocovariances changes greatly depending on whether a time series has long range dependence, anti-persistence, or short memory. This in turn has a large impact on the convergence rates of sample mean and tapered autocovariances. We provide a unified treatment of the various types of memory, including the important Super-Efficient (SE) case, wherein the partial sums are $o_P(\sqrt{n})$. The SE scenario is important, since it can easily arise from over-differencing of a time series suspected of having trend nonstationarity.

Several novel results on the memory of a time series are presented, which – together with examples – furnish some intuition for the qualitative behavior. The properties of Differential Memory (DM) processes are elucidated, and shown to be distinct from the behavior of Negative

[3] Some authors prefer to investigate what the power would be were these statistics to be corrected to be correctly sized; however, in practice such a procedure is impossible to implement. We have chosen to display the power that would occur were a practitioner to utilize any of the methods.

Memory (NM) processes – together, the DM and NM cases partition the important SE case. But our main interest is in the asymptotics of sample mean and tapered autocovariances, and we treat these topics through several theorems. For the asymptotic results we consider both the vanishing bandwidth-fraction case (a more classical approach, going back to Parzen (1957)) and the fixed bandwidth-fraction case (a more recent approach espoused in Kiefer, Vogelsang, and Bunzel (2000)). We both summarize known results, and prove new ones, examining three broad classes of DGP that exhibit the various types of memory described herein.

In order to make use of the asymptotic results, it is still necessary to get the critical values of the limiting distributions, which in the fixed bandwidth-fraction case are functionals of the fractional Brownian Bridge. We propose two methodologies: subsampling, which avoids explicit estimation of the memory parameter β but requires selection of a sampling fraction a; and the plug-in approach, which requires an estimate of β and a look-up table of critical values (computed ahead of time via simulation) for the limiting distributions. These methods are compared through extensive finite-sample size and power simulations, which are succinctly summarized here.[4] While power tends to deteriorate with greater memory, Type II error is comparatively quite small with anti-persistent processes. There are size problems with the plug-in method, whereas in contrast the subsampling method tends to have superior coverage (i.e., empirical size is closer to the nominal level).

In summary, we provide a viable framework for conducting inference for the mean, supplying a unified asymptotic theory that covers all different types of memory under a single umbrella. This framework is robust against different memory specifications, obviating the need to do extensive modeling.

Appendix

Proof of Proposition 1. For cases (i), (ii), and (iii), we apply L'Hopital's rule to V_n/nW_n. So long as $\beta \neq -1$, we have

$$\frac{V_n}{nW_n} \sim \frac{\sum_{k=0}^{n-1} W_k}{Cn^{\beta+1}L(n)} \sim \frac{W_n}{C(\beta+1)n^{\beta}L(n) + Cn^{\beta+1}\dot{L}(n)} \sim \frac{1}{(\beta+1) + n\dot{L}(n)/L(n)}$$

in cases (i) or (iii). By condition M, this tends to $1/(\beta+1)$. But if $\beta = -1$, we have $V_n/(nW_n) \sim \sum_{h<n} h^{-1}L(h)/L(n) \to \infty$ by Theorem A3.6 of Embrechts, Klüppelberg, and Mikosch (1997). In case (ii), $V_n/(nW_n) \to 1$ since it is a Cesaro sum of a convergent sequence, divided by the sequence W_n. Case (iv) is immediate from its definition. For (7), in cases (i) and (iii) with $\beta \neq 0$ L'Hopital's rule yields

$$C \sim \frac{\sum_{|k|<n} \gamma_k}{n^{\beta}L(n)} \sim \frac{2\gamma_n}{\beta n^{\beta-1}L(n) + n^{\beta}\dot{L}(n)} = \frac{2\gamma_n}{\beta n^{\beta-1}L(n)\left(1 + n\dot{L}(n)/\beta L(n)\right)},$$

[4]All code and results are available from the first author.

and $n\dot{L}(n)/L(n) = n\dot{c}(n)/c(n) + \eta(n) = o(1)$ by condition M; the result follows. Now if $\beta = 0$ in case (i) or (iii) we can use L'Hopital's rule again to obtain $C \sim 2\gamma_n/\dot{L}(n)$. $\square$

Proof of Proposition 2. Without loss of generality, suppose that $\mu = 0$. By the Wold decomposition (see Brockwell and Davis, 1991) there exists an uncorrelated sequence $\{\epsilon_t\}$ and square summable coefficients $\{\psi_j\}$ such that $Y_t = \sum_{j\geq 0} \psi_j \epsilon_{t-j}$. Let $\sigma^2 = Var(\epsilon_t)$ be equal to one for simplicity. We have $0 = \lim_{k\to\infty} W_k = \sum_h \gamma_h = (\sum_j \psi_j)^2 \sigma^2$, so $\sum_j \psi_j = 0$. It follows that $1 - z$ should divide $\Psi(z) = \sum_{j\geq 0} \psi_j z^j$, though we must establish that $\Theta(z) = \Psi(z)/(1-z)$ converges. Extending ψ_j to be zero if $j < 0$, we define $\theta_j = \sum_{k=-\infty}^{j} \psi_k$ for any integer j. Note that $\theta_j \to 0$ as $j \to \infty$ and is zero if $j < 0$. We will show that the θ_j sequence is square summable, so that $\sum_j \theta_j \epsilon_{t-j}$ is finite with probability one; then this will define Z_t, from which $Y_t = Z_t - Z_{t-1}$ follows at once. Note that $\theta_j = \sum_{k\geq 0} \psi_{j-k}$; let $\theta_{j,m} = \sum_{k=0}^{m} \psi_{j-k}$. Then $\sum_j \theta_{j,m}^2 = \sum_{i,k=0}^{m} \gamma_{i-k} = \sum_{k=0}^{m} W_k$ for each m. Then by Fatou's Lemma

$$\sum_j \theta_j^2 \leq \underline{\lim}_{n\to\infty} \sum_j \theta_{j,m}^2 = \sum_{k\geq 0} W_k < \infty. \quad \square$$

Proof of Theorem 1. Since Condition M is assumed, Proposition 1 holds, and in particular we can utilize (6). Under P1, the central limit theorem follows from $V_n \to \infty$, the stated hypotheses, and Rosenblatt (1956) – also see Rosenblatt (1961, 1984). Note that we must exclude $\beta = -1$ in the NM case to ensure $V_n \to \infty$.

If P2 holds, then $Y_t = \sum_j \psi_j \epsilon_{t-j}$ for an *iid* sequence $\{\epsilon_t\}$, and by assumption $\gamma_0 \propto \sum_j \psi_j^2$ exists. So Theorem 5.2.3 of Taniguchi and Kakizawa (2000) (which is due to Ibragimov and Linnik (1971, p. 359)) gives the result. Note that by Hosoya (1996), we could relax the independence assumption on $\{\epsilon_t\}$ to a type of mixing condition.

Under P3 first consider the LM case. Since $\beta > 0$, we have $r_k \sim \frac{\beta}{2} k^{\beta-1} L(k)$. Then we apply Theorem 5.2.2 of Taniguchi and Kakizawa (2000) (which is due to Taqqu, 1975) when $(1-\beta)\tau < 1$, obtaining a non-central limit theorem if $\tau > 1$. In this case $\{Y_t\}$ is $LM(1-(1-\beta)\tau)$. The distribution of Q is described in the above references, and is non-Gaussian if $\tau > 1$. If the $\{X_t\}$ process is SM, then a central limit theorem holds by applying Breuer and Major (1983).

Under P3 and NM, write $g(x) = \sum_{q\geq 0} H_q(x) J_q/q!$, with H_q the qth Hermite polynomial and $J_q = \mathbb{E}[g(Z)H_q(Z)]$ where Z is standard normal. Then we have $\gamma_k = \mathbb{E}[Y_t Y_{t+k}] = \sum_q r_k^q J_q^2/q!$ (see 5.2.10 of Taniguchi and Kakizawa (2000)). Since $\{Y_t\}$ is NM(β), we have $0 = \sum_k \gamma_k = \sum_q J_q^2 \sum_k r_k^q/q!$. Since $J_q^2/q! \geq 0$ and $\sum_k r_k^q \geq 0$ (since it is the limit of $n^{-1} Var(\sum_{t=1}^n H_q(X_t))$), for each q, we have $J_q = 0$ or $\sum_k r_k^q = 0$, or both. Now

$$1 = \frac{\sum_{|k|\leq n} \gamma_k}{W_n} = \sum_q \frac{J_q^2}{q!} \frac{\sum_{|k|\leq n} r_k^q}{W_n}.$$

Since $\sum_{|k|\leq n} r_k^q / W_n$ is eventually positive, we can use the Dominated Convergence Theorem to get $1 = \sum_q J_q^2 D_q / q!$, where $D_q = \lim_{n\to\infty} \sum_{|k|\leq n} r_k^q / W_n$. It follows that there exists a q such that $J_q \neq 0$ and $D_q \neq 0$. For any such q, $D_q \neq 0$ implies $\lim_{n\to\infty} \sum_{|k|\leq n} r_k^q = 0$, and L'Hopital's rule yields $r_k^q \sim \frac{\beta}{2} C D_q k^{\beta-1} L(k)$. This is only summable if $q = 1$, so $1 = J_1^2 D_1$. Also for all $q > 1$, either $J_q = 0$ or $D_q = 0$. In the latter case, we have $\sum_{|k|\leq n} r_k^q = o(n^\beta L(n))$. Thus

$$\sum_{t=1}^n Y_t = J_1 \sum_{t=1}^n X_t + \sum_{q>1} \frac{J_q}{q!} \sum_{t=1}^n H_q(X_t)$$

$$Var(\sum_{t=1}^n H_q(X_t)) = \sum_{|h|<n} (n - |h|) r_h^q q! = q! \sum_{h=0}^{n-1} \sum_{|k|\leq h} r_k^q = o(n^{\beta+1} L(n))$$

using equation 5.2.9 of Taniguchi and Kakizawa (2000) and Theorem A3.6 of Embrechts et al. (1997). Hence

$$V_n^{-1/2} \sum_{t=1}^n Y_t = o_P(1) + J_1 \sum_{t=1}^n X_t / \sqrt{V_n}. \tag{A.1}$$

This latter term is Gaussian with asymptotic variance $J_1^2 D_1 = 1$, i.e., the partial sum is asymptotic to the first (Gaussian) term of the Hermite expansion in the NM case, the nonlinear portions being negligible. This proves the result. □

Proof of Theorem 2. Without loss of generality suppose that $\mu = 0$. Applying Proposition 2, we have $S_n = Z_n - Z_0$. If $\{Z_t\}$ is strongly mixing, let $\tilde{Z}_n$ be equal in distribution to Z_n for every n, but independent of Z_0. Then

$$\left| \mathbb{E} \exp\{i\nu(Z_n - Z_0)\} - \mathbb{E} \exp\{i\nu(\tilde{Z}_n - Z_0)\} \right| \leq 16 \alpha_k$$

by Lemma B.0.6 of Politis, Romano, and Wolf (1999), which is due to Ibragimov (1962). Now $\phi_{\tilde{Z}_n - Z_0}(\nu) = \phi_Z(\nu) \cdot \phi_Z(-\nu)$ where ϕ denotes the characteristic function, and Z has the common distribution of the $\{Z_t\}$ process.

If $\{Z_t\}$ is linear, then $\{\epsilon_t\}$ are *iid* in the representation $Z_t = \sum_{j\geq 0} \theta_j \epsilon_{t-j}$, with the convention that $\theta_j = 0$ if $j < 0$. Then

$$S_n = \sum_j (\theta_{n-j} - \theta_{-j}) \epsilon_j = \sum_{j=1}^n \theta_{n-j} \epsilon_j + \sum_{j\leq 0} (\theta_{n-j} - \theta_{-j}) \epsilon_j.$$

These two terms are independent, and the first is equal in distribution to $\sum_{j=0}^{n-1} \theta_j \epsilon_j$, which tends in probability (and thus weakly) to $\sum_{j\geq 0} \theta_j \epsilon_j$ by Theorem 22.6 of Billingsley (1995). This limit is equal in distribution to Z, so $\sum_{j=1}^n \theta_{n-j} \epsilon_j \stackrel{\mathcal{L}}{\Longrightarrow} Z$. For the second term, we have $\sum_{j\leq 0} \theta_{n-j} \epsilon_j \stackrel{P}{\longrightarrow} 0$, since its variance is equal to $\sum_{j\geq n} \theta_j^2 \sigma^2$. What's left is $-\sum_{j\leq 0} \theta_{-j} \epsilon_j$, which is equal to $-Z$ in distribution. Since the two terms are independent, we have that S_n converges weakly to the difference of two independent random variables, each with the distribution of Z. □

Proof of Proposition 3. The expression $\widetilde{\gamma}_h - \gamma_h$ is unchanged if we replace Y_t by $Y_t - \mu$, so without loss of generality suppose that $\mu = 0$. Then for $h \geq 0$

$$\begin{aligned} \widetilde{\gamma}_h &= \gamma_h + (\overline{\gamma}_h - \gamma_h) + (\tilde{\gamma}_h - \overline{\gamma}_h) \\ &= \left(1 - \frac{h}{n}\right)\gamma_h + \frac{1}{n}\sum_{t=1}^{n-h}(Y_t Y_{t+h} - \gamma_h) + \overline{Y}\left(\overline{Y} - \overline{Y}_{1:n-h} - \overline{Y}_{h+1:n}\right), \end{aligned}$$

where $\overline{Y}_{1:n-h} = \sum_{t=1}^{n-h} Y_t/n$ and $\overline{Y}_{h+1:n} = \sum_{t=h+1}^{n} Y_t/n$. Let $c_n = V_n n^{-2}$. Then $\overline{Y} = O_P(n^{-1}V_n^{-1/2})$, so the third term above is $O_P(c_n)$ uniformly in h (this is true since $|h| \leq M = o(n)$). Now $c_n = n^{\beta-1}L(n)$ in the LM and NM cases, with $\beta \in [-1, 1)$, but $c_n = n^{-1}$ for the SM case and $c_n = n^{-2}$ for the DM case. Using the U_M bound on $\Lambda_M(h)$, we obtain $E_M^{(1)} = O_P(U_M\, M c_n)$ as claimed. For the error term $E_M^{(2)}$, we compute

$$\begin{aligned} Var(E_M^{(2)}) &\leq n^{-2}\sum_{h,k}\Lambda_M(h)\Lambda_M(k)\sum_{t,s=1}^{n} Cov(Y_t Y_{t+h}, Y_s Y_{s+k}) \\ &\leq n^{-2}\sum_{h,k}\Lambda_M(h)\Lambda_M(k)\sum_{t,s=1}^{n}\left(cum(Y_t, Y_{t+h}, Y_s, Y_{s+k}) + \gamma_{t-s}\gamma_{t-s+h-k} + \gamma_{t-s-k}\gamma_{t+h-s}\right) \\ &\leq n^{-1}\sum_{h,k}\Lambda_M(h)\Lambda_M(k)\sum_{|l|\leq n}(1 - |l|/n)\left(cum(Y_0, Y_h, Y_l, Y_{l+k}) + \gamma_l\gamma_{l-h+k} + \gamma_{l+k}\gamma_{h-l}\right). \end{aligned}$$

The inequality used here only concerns $|h|$ terms, which is negligible with respect to n. The sum of the cumulant function is bounded using Assumption B, resulting in an overall bounds of U_M^2/n. For the sum over $\gamma_l\gamma_{l-h+k}$, we can use the Cauchy-Schwarz inequality to obtain the bound $O(M^2 U_M^2 n^{\xi-1}K(n))$, where K is a slowly-varying function and $n^{\xi}K(n)$ represents the order of $\sum_{|l|\leq n}\gamma_l^2$; using Proposition 1 we have $\xi = 0$ if $\beta < 1/2$ (as this results in a square summable sequence) or $\xi = 2\beta - 1$ if $\beta > 1/2$. When $\beta = 1/2$ and $L(n) \equiv 1$, we take $\xi = 0$ and $K(n) = \log n$. The analysis of the sum over $\gamma_{l+k}\gamma_{h-l}$ yields a similar order. Taking square roots, we learn that $E_M^{(2)} = O_P(M U_M n^{-\eta} K^{1/2}(n))$ as stated. □

Proof of Theorem 3. First note that we can remove the term $(1 - |h|/n)$ since $M/n \to 0$. The case of the truncation filter is trivial. Case (iv) is treated differently, so we first consider cases (i), (ii), and (iii). We proceed to break the sum over h up according to the intervals of smoothness for Λ. The first such interval is $[-c, c]$, which corresponds to the flat-top interval; if there is no flat-top region, then $c = 0$. In general, consider an interval $(r, s]$ such that the restriction of Λ is continuously differentiable there. Then

$$\sum_{[rM]<|h|\leq[sM]}\Lambda\left(\frac{h}{M}\right)\gamma_h = \Lambda\left(\frac{[sM]}{M}\right)W_{[sM]} - \Lambda\left(\frac{[rM]}{M}\right)W_{[rM]} + \sum_{h=[rM]}^{[sM]-1}\left[\Lambda\left(\frac{h}{M}\right) - \Lambda\left(\frac{h+1}{M}\right)\right]W_h \tag{A.2}$$

via summation by parts. The first two terms on the right hand side will cancel with other like terms for the other intervals, leaving the last term on the right hand side. The terms in the square brackets consist of values of Λ restricted to $(r,s]$, excepting only the first term $h = [rM]$; however, since $\Lambda([rM]/M) - \Lambda(([rM]+1)/M) \approxeq -M^{-1}\dot{\Lambda}_+(r)$ by continuity, the first term's analysis is the same as the others. In general, $\Lambda(h/M) - \Lambda(h+1/M) = -\dot{\Lambda}(h/M)M^{-1} + O(M^{-2})$. The case of $r > 0$ for an interval $(r,s]$ has a different analysis from the $r = 0$ case, which we consider later. So long as $r > 0$, we have $h \to \infty$ as $M \to \infty$ in the above summation. Thus in case (ii), $W_h \to W_\infty$ and this convergence occurs uniformly in $h \in (rM, sM)$ as $M \to \infty$. Using the boundedness of $\dot{\Lambda}(x)$ and the limit of a Cesaro sum, we obtain a limit of $-W_\infty \int_r^s \dot{\Lambda}(x)\,dx$. The argument can be extended to cases (i) and (iii) as follows:

$$\frac{\sum_{h=[rM]}^{[sM]-1} \left[\Lambda\left(\frac{h}{M}\right) - \Lambda\left(\frac{h+1}{M}\right)\right] W_h}{CM^\beta L(M)} = \sum_{h=[rM]}^{[sM]-1} \left\{\Lambda\left(\frac{h}{M}\right) - \Lambda\left(\frac{h+1}{M}\right)\right\} (h/M)^\beta \left[W_h h^{-\beta}/CL(M)\right],$$

and the expression in square brackets equals one plus error tending to zero as $M \to \infty$, uniformly in $h \in ([rM],[sM]]$. This is because

$$\left|\frac{W_h}{Ch^\beta L(M)} - 1\right| \leq \left|\frac{W_h}{Ch^\beta L(h)} - 1\right| + \left|\frac{W_h}{Ch^\beta L(h)}\left(\frac{L(h)}{L(M)} - 1\right)\right|.$$

The first term tends to zero uniformly in h as $M \to \infty$. For the second, we have $L(h)/L(M) \to 1$ uniformly in h as $M \to \infty$ as well, which is seen by using the representation (5). Note that these arguments hold for $\beta = 0$. Using the boundedness of x^β for $x \in (r,s]$ and $r > 0$, we obtain a limit of $-\int_r^s \dot{\Lambda}(x)x^\beta\,dx$ as $M \to \infty$. This argument works for any $\beta \in (-1,1)$.

The first interval must be treated differently – unless it is flat-top, i.e., $c > 0$, in which case it is trivially given by $W_{[cM]}$, which cancels with boundary term in the next intervals. More generally, the first interval has the form $\sum_{0 \leq |h| \leq [sM]} \Lambda(h/M)\gamma_h$, which tends to $\Lambda(0)W_\infty$ in case (ii). Otherwise in cases (i) and (iii) we have

$$\sum_{|h|=0}^{[sM]} \Lambda\left(\frac{h}{M}\right)\gamma_h = \Lambda(0)\,W_{[sM]} + 2\sum_{h=1}^{[sM]}\left[\Lambda\left(\frac{h}{M}\right) - \Lambda(0)\right]\gamma_h.$$

Note that $\beta > -1$ by assumption. The expression in brackets can be expanded in the Taylor series $\sum_{j\geq 1} \Lambda^{(j)}(0)h^j M^{-j}/j!$. Noting that $\sum_{h=1}^M h^j \gamma_h$ is divergent and asymptotic to $\beta M^j W_M/2(\beta+j)$ (proved by L'Hopital's rule and (7)) for any $j \geq 1$, we can interchange summations to obtain

$$W_{[sM]} + \beta\,W_{[sM]}\sum_{j\geq 1}\Lambda^{(j)}(0)\frac{s^j}{\beta+j} \sim W_M\left(s^\beta + \beta\int_0^s (\Lambda(x)-1)x^{\beta-1}\,dx\right)$$

after some algebra. Now piecing all intervals together, taking into account cancelations and integration by parts, we arrive at (9).

Now we turn to case (iv). Letting the partition $0 = s_0 < s_1 < \cdots s_T < s_{T+1} = 1$ denote the points of non-differentiability in Λ, we can write

$$\sum_{|h|=0}^{[sM]} \Lambda\left(\frac{h}{M}\right)\gamma_h = \sum_{j=0}^{T} \sum_{h=[s_j M]}^{[s_{j+1}M]-1} \left[\Lambda\left(\frac{h}{M}\right) - \Lambda\left(\frac{h+1}{M}\right)\right] W_h.$$

The first term is identically zero for a flat-top kernel; otherwise when $c = 0$, we can use a Taylor series expansion at h/M again. Since $\{W_h\}$ is a summable sequence, we can apply the Dominated Convergence Theorem and obtain $2r_0$ times the Taylor series at zero, as stated in the theorem. As for the other intervals, note that $\sum_{h=[s_j M]}^{[s_{j+1}M]-1} W_h = 2(r_{[s_j M]} - r_{[s_{j+1}M]})$ is tending to zero as $M \to \infty$, since $j \geq 1$. Thus these other terms decay even faster than the first term. Hence for flat-top tapers, the rate of convergence is $o(M^{-1})$. $\square$

Proof of Theorem 4. By Theorem 4.15 of Karatzas and Shreve (1991), it suffices to establish tightness and convergence of finite-dimensional distributions. Without loss of generality set $\mu = 0$. We handle tightness first, since the same argument works under P1, P2, or P3. By Problem 4.11 of Karatzas and Shreve (1991), it suffices to verify two simple conditions. First, $\mathbb{E}|V_n^{-1/2} S_{[n0]}|^2 = 0$ has bounded supremum over $n \geq 1$; second,

$$\mathbb{E}|V_n^{-1/2} S_{[tn]} - V_n^{-1/2} S_{[sn]}|^2 = V_n^{-1} \mathbb{E}|S_{[|t-s|n]}|^2 = V_{[|t-s|n]}/V_n,$$

which tends to $|t-s|^{\beta+1}$ by Proposition 1. Thus the supremum over $n \geq 1$ is bounded by $|t-s|^{\beta+1}$, and we conclude that the sequence is tight.

Now turning to the finite-dimensional distributions, first consider P1. Take any integer $m > 0$ and let $r_1 < r_2 < \cdots < r_m \in [0,1]$, and any real numbers $\alpha_1, \cdots, \alpha_m$. Also set $r_0 = 0$. Then

$$\sum_{j=1}^{m} \alpha_j V_n^{-1/2} S_{[r_j n]} = V_n^{-1/2} \sum_{i=0}^{m-1} \left(S_{[r_{i+1}n]} - S_{[r_i n]}\right) \sum_{j=i+1}^{m} \alpha_j,$$

and each summand $S_{[r_{i+1}n]} - S_{[r_i n]}$ involves a collection of random variables with disjoint index sets. We next introduce intermediary "small blocks" of random variables with roughly k terms, where $k/n + 1/k \to 0$ as $n \to \infty$:

$$S_{[r_{i+1}n]} - S_{[r_i n]} = \left(S_{[r_{i+1}n]} - S_{[r_i n]+k}\right) + \left(S_{[r_i n]+k} - S_{[r_i n]}\right)$$

for each i. The second term on the right hand side is the small block, and has variance $V_k = O(k^{\beta+1} L(k))$. Hence $V_k/V_n \to 0$ as $n \to \infty$. But the first (large) block is equal in distribution to $S_{[(r_{i+1}-r_i)n]-k}$, which is asymptotically normal with variance $V_n (r_{i+1} - r_i)^{\beta+1}$. Since each of the large blocks are distance k apart, the overall characteristic functions can be replaced by a characteristic function where the large blocks are actually independent of each other, using Lemma B.0.6 of Politis, Romano, and Wolf (1999). As a result, $\sum_{j=1}^{m} \alpha_j V_n^{-1/2} S_{[r_j n]}$ is asymptotically normal with variance $\sum_{i=0}^{m-1} (r_{i+1} - r_i)^{\beta+1} \left(\sum_{j=i+1}^m \alpha_j \right)^2$. After some algebra, it is apparent that this is equal to the variance of $\sum_{j=1}^m \alpha_j B(r_j)$, which demonstrates convergence of finite-dimensional distributions.

Now consider P2, and adapt the argument used in Theorem 5.2.3 of Taniguchi and Kakizawa (2000). The linear representation is $Y_t = \sum_j \epsilon_j \psi_{t-j}$ as in the proof of Theorem 1. We have $\sum_{j=1}^m \alpha_j S_{[r_j n]} = \sum_j \epsilon_j b_{j,n}$, with $b_{j,n} = \sum_{i=1}^m \alpha_i \sum_{t=1}^{[r_i n]} \psi_{t-j}$. Since $m < \infty$, the same types of bounds used in the proof of Theorem 5.2.3 of Taniguchi and Kakizawa (2000) still apply. Hence $\sigma_n^{-1} \sum_{j=1}^m \alpha_j S_{[r_j n]}$ is asymptotically standard normal, where the normalizing variance is

$$\begin{aligned}
\sigma_n^2 &= \sum_j b_{j,n}^2 \sigma^2 = \sum_j \left(\sum_{i=1}^m \alpha_i \sum_{t=1}^{[r_i n]} \psi_{t-j} \right)^2 \sigma^2 = \sigma^2 \sum_j \sum_{i,l=1}^m \alpha_i \alpha_l \sum_{t=1}^{[r_i n]} \sum_{s=1}^{[r_l n]} \psi_{t-j} \psi_{s-j} \\
&= \sum_{i,l=1}^m \alpha_i \alpha_l \sum_{t=1}^{[r_i n]} \sum_{s=1}^{[r_l n]} \gamma_{t-s} = \sum_{i=1}^m \alpha_i^2 \sum_{t,s=1}^{[r_i n]} \gamma_{t-s} + 2 \sum_{i<l} \alpha_i \alpha_l \sum_{t=1}^{[r_i n]} \sum_{s=1}^{[r_l n]} \gamma_{t-s} \\
&= \sum_{i=1}^m \alpha_i^2 V_{[r_i n]} + \sum_{i<l} \alpha_i \alpha_l \left(V_{[r_i n]} + V_{[r_l n]} - V_{[(r_l - r_i) n]} \right).
\end{aligned}$$

It follows that

$$\sigma_n^2 / V_n \sim \sum_{i=1}^m \alpha_i^2 r_i^{\beta+1} + \sum_{i<l} \alpha_i \alpha_l \left(r_i^{\beta+1} + r_l^{\beta+1} - (r_l - r_i)^{\beta+1} \right),$$

which is the same as the variance of $\sum_{i=1}^m \alpha_i B(r_i)$. In other words the limit of $V_n^{-1/2} \sum_{j=1}^m \alpha_j S_{[r_j n]}$ has the same distribution as $\sum_{i=1}^m \alpha_i B(r_i)$.

Finally, the P3 case is handled as follows. If LM and $\beta > 0$, we apply Theorem 5.2.2 of Taniguchi and Kakizawa (2000); note that we only obtain a FBM limit if $\tau = 1$. In the SM case a FBM limit is also obtained by the same arguments (but things are easier since all powers of the acf are summable). For the NM case, as in the proof of Theorem 1 we have (A.1), and we can substitute $[nr]$ for n in the summation. Clearly $\sum_{t=1}^{[nr]} X_t / \sqrt{V_n}$ is Gaussian, and its asymptotic variance is $D_1 r^{\beta+1}$. It is easy to see that this has the finite-dimensional distribution of FBM, which completes the proof. $\square$

Proof of Theorem 5. For the LM, SM, NM case, note that $\widehat{S}_i = S_{[rn]} - r S_n$ with $r = i/n$, and recognize the summations in the expression for $W_{\Lambda,M}$ in Proposition 4 as Riemann sums; the result follows at once from Theorem 4. For the DM case we have $\widehat{S}_i = Z_i - \frac{i}{n} Z_n - \left(1 - \frac{i}{n}\right) Z_0$, and hence without loss of generality suppose that Z_i is mean zero. Using the abbreviation $\Delta_n(i-j) =$

$2\Lambda_n(i-j) - \Lambda_n(i-j+1) - \Lambda_n(i-j-1)$,

$$nW_{\Lambda,M} = \sum_{i,j=1}^{n} \widehat{S}_i \widehat{S}_j \Delta_n(i-j) \tag{A.3}$$

$$= \sum_{i,j=1}^{n} Z_i Z_j \Delta_n(i-j) - 2Z_n \sum_{i,j=1}^{n} \frac{i}{n} Z_j \Delta_n(i-j) - 2Z_0 \sum_{i,j=1}^{n} \left(1 - \frac{i}{n}\right) Z_j \Delta_n(i-j)$$

$$+ 2Z_0 Z_n \sum_{i,j=1}^{n} \left(1 - \frac{i}{n}\right) \frac{j}{n} \Delta_n(i-j) + Z_n^2 \sum_{i,j=1}^{n} \frac{ij}{n^2} \Delta_n(i-j) + Z_0^2 \sum_{i,j=1}^{n} \left(1 - \frac{i}{n}\right)\left(1 - \frac{j}{n}\right) \Delta_n(i-j).$$

Using the Taylor series approximations for $\Delta_n(h)$ derived in the proof of Proposition 4, we can compute each of the six terms above. Each of these is of the form (up to multiplication by other random variables) $\sum_{i,j=1}^{n} X_i X_j \Delta_n(i-j)$ for some random variables X_i, which is approximately

$$- 2b^{-2} \frac{1}{n} \sum_{h=[cbn]+1}^{[bn]-1} \ddot{\Lambda}\left(\frac{h}{bn}\right) \frac{1}{n} \sum_{i=1}^{n-h} X_i X_{i+h}$$

$$- 2b^{-1} \dot{\Lambda}_+(c) \frac{1}{n} \sum_{i=1}^{n-[cbn]} X_i X_{i+[cbn]}$$

$$+ 2b^{-1} \dot{\Lambda}_-(1) \frac{1}{n} \sum_{i=1}^{n-[bn]} X_i X_{i+[bn]},$$

up to $o_P(1)$ terms. Now in the first term in (A.3), we have $X_i = Z_i$ and $\frac{1}{n}\sum_{i=1}^{n-h} Z_i Z_{i+h} \xrightarrow{P} (1 - h/n) r_h$. Now if $c = 0$ we obtain $\frac{1}{n}\sum_{i=1}^{n-[cbn]} W_i W_{i+[cbn]} \xrightarrow{P} r_0$; all other portions of the first term tend to zero in probability, whether or not $c = 0$. For the second and third terms of (A.3), we have $X_j = Z_j$ and X_i deterministic; then these terms tends to zero since Z_j is mean zero. For the fourth term of (A.3), the limit is $2Z_0 Z_*$ times

$$-b^{-2} \int_{cb<|r-s|<b} (1-r) s\, \ddot{\Lambda}\left(\frac{|r-s|}{b}\right) drds - 2b^{-1} \dot{\Lambda}_+(c) \int_0^{1-cb} (1-r)(r+cb)\, dr + 2b^{-1} \dot{\Lambda}_-(1) \int_0^{1-b} (1-r)(r+b)\, dr.$$

The fifth term of (A.3) is asymptotic to Z_*^2 times

$$-b^{-2} \int_{cb<|r-s|<b} rs\, \ddot{\Lambda}\left(\frac{|r-s|}{b}\right) drds - 2b^{-1} \dot{\Lambda}_+(c) \int_0^{1-cb} r(r+cb)\, dr + 2b^{-1} \dot{\Lambda}_-(1) \int_0^{1-b} r(r+b)\, dr,$$

while the sixth term of (A.3) tends to Z_0^2 times

$$- b^{-2} \int_{cb<|r-s|<b} (1-r)(1-s) \ddot{\Lambda}\left(\frac{|r-s|}{b}\right) drds - 2b^{-1} \dot{\Lambda}_+(c) \int_0^{1-cb} (1-r)(1-r-cb)\, dr$$

$$+ 2b^{-1} \dot{\Lambda}_-(1) \int_0^{1-b} (1-r)(1-r-b)\, dr.$$

Combining the fourth, fifth, and sixth terms together yields $Q(b)$ with $\widetilde{C}(r) = Z_0 + r(Z_* - Z_0)$, which interpolates between Z_0 and Z_*. This establishes the convergence of $nW_{\Lambda,M}$; for the joint convergence, note that we can rework the same proof for $\alpha(S_n - n\mu) + \gamma n W_{\Lambda,M}$ for any α and γ, since $S_n - n\mu = Z_n - Z_0$. Hence up to terms $o_P(1)$, $\alpha(S_n - n\mu) + \gamma n W_{\Lambda,M}$ can be expressed as a bilinear combination of Z_0 and Z_n, which converges to $\alpha(Z_* - Z_0) + \gamma P(b)$ after gathering like terms. $\square$

Proof of Proposition 5. By Theorem 5 we know that $Q_n := nW_{\Lambda,bn}/V_n \stackrel{\mathcal{L}}{\Longrightarrow} Q(b)$, and by Proposition 1 we have $V_n/n \sim n^\beta CL(n)/(\beta+1)$ in the LM/SM/NM case. Thus it follows that (17) satisfies

$$\widehat{\beta} \sim \beta + \frac{\log Q_n}{\log n} + \frac{\log\left(CL(n)/(\beta+1)\right)}{\log n}.$$

Since $Q_n = O_P(1)$ and $\log L(n)/\log n \to 0$ for slowly-varying functions L, the estimator will be consistent. In the DM case, we obtain

$$\widehat{\beta} = -1 + \frac{\log Q_n}{\log n} + \frac{\log V_n}{\log n}$$

since $Q_n = nW_{\Lambda,bn}/V_n$. Now $V_n \to 2r_0$ in the DM case, so $\widehat{\beta} \stackrel{P}{\longrightarrow} -1$. $\square$

References

[1] Andrews, D. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–858.

[2] Ango-Nze, P., Dupoiron, S., and Rios, R. (2003) Subsampling under weak dependence conditions. Technical report from CREST, $DT2003-42$.

[3] Bardet, J., Doukhan, P., Lang, G., and Ragache, N. (2008) Dependent Lindeberg Central Limit Theorem and some applications. Preprint.

[4] Beran, J. (1993) Fitting long-memory models by generalized linear regression. *Biometrika*, **80**, 817-822.

[5] Beran, J. (1994) *Statistics for Long Memory Processes*. New York: Chapman and Hall.

[6] Billingsley, P. (1995) *Probability and Measure*. John Wiley and Sons: New York.

[7] Blackman, R.B. and Tukey, J.W. (1959). *The Measurement of Power Spectra from the Point of View of Communications Engineering*, Dover, New York.

[8] Bradley, R. (2007) *Introduction to Strong Mixing Conditions*, Volume I, Kendrick Press, Heber City, Utah.

[9] Breuer, P. and Major, P. (1983) Central limit theorems for non-linear functionals of Gaussian fields. *Journal of Multivariate Analysis* **13**, 425–441.

[10] Brillinger, D.R. (1981) *Time Series: Data Analysis and Theory*, Holden-Day, New York.

[11] Brockwell, P. and Davis, R. (1991) *Time Series: Theory and Methods, 2nd Ed.* New York: Springer.

[12] Davydov, Y. (1970) The invariance principle for stationary processes. *Theory of Probability and its Applications* **14**, 487–498.

[13] Dickey, D. and Fuller, W. (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**, 427–431.

[14] Dickey, D. and Fuller, W. (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **49**, 1057–1072.

[15] Doukhan, P. and Louhichi, S. (1999) A new weak dependence and applications to moment inequalities. *Stochastic Processes and Their Applications* **84**, 313–343.

[16] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modeling Extremal Events for Insurance and Finance.* Springer-Verlag, Berlin.

[17] Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C. and Chen, B. C. (1998) New capabilities and methods of the X-12-ARIMA seasonal adjustment program. *Journal of Business and Economic Statistics* **16**, 127–177 (with discussion).

[18] Gorodetskii, V. (1977) On convergence to semi-stable Gaussian processes. *Theory Prob. Appl.* **22**, 498–508.

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

[20] Hamilton, J.D. (1994) *Time Series Analysis.* Princeton University Press, Princeton.

[21] Hannan, E.J. (1970) *Multiple Time Series*, John Wiley, New York.

[22] Giraitis, L. and Surgailis, D. (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.

[23] Giraitis, L. and Taqqu, M. (1999) Whittle estimator for finite-variance non-Gaussian time series with long memory. *The Annals of Statistics* **27**, 178–203.

[24] Hosking, J.R.M. (1981) Fractional differencing. *Biometrika*, **68**, 165-176.

[25] Hosoya, Y. (1996) Some limit theorems on stationary processes with long-range dependence. *Athens conference on applied probability and time series* **2**, 233–245. eds. P.M. Robinson and R. Rosenblatt, Springer.

[26] Hurvich, C. (2002) Multistep forecasting of long memory series using fractional exponential models. *International Journal of Forecasting* **18**, 167–179.

[27] Ibragimov, I. (1962) Some limit theorems for stationary processes. *Theory of Probability and its Applications* **7**, 349–382.

[28] Ibragimov, I. and Linnik, Y. (1971) *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhof Publishing Groningen.

[29] Jach, A., McElroy, T., and Politis, D. (2011) Subsampling inference for the mean of heavy-tailed long memory time series. Forthcoming, *Journal of Time Series Analysis.*

[30] Kiefer, N., Vogelsang, T. and Bunzel, H. (2000) Simple robust testing of regression hypotheses. *Econometrica* **68**, 695–714.

[31] Kiefer, N. and Vogelsang, T. (2002) Heteroscedastic-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica* **70**, 2093–2095.

[32] Kolmogorov, A. and Rozanov, J. (1960) On a strong mixing condition for a stationary Gaussian process. *Theory of Probability and its Applications* **5**, 204–208.

[33] Lahiri, S. (2003) *Resampling Methods for Dependent Data.* Springer-Verlag.

[34] Maravall, A. and Caporello, G. (2004) Program TSW: Revised Reference Manual. *Working Paper 2004, Research Department, Bank of Spain.* http://www.bde.es

[35] Marinucci, D. and Robinson, P. (2000) Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications* **86**, 103–120.

[36] McElroy, T. and Politis, D. (2007) Computer-intensive rate estimation, diverging statistics, and scanning. *Annals of Statistics* **35**, 1827–1848.

[37] McElroy, T. and Politis, D. (2009) Fixed-b asymptotics for the studentized mean from time series with short, long or negative memory. UC San Diego: Department of Economics, UCSD.

[38] McElroy, T. and Politis, D. (2011) Fixed-b Asymptotics for the Studentized Mean for Long and Negative Memory Time Series. Under revision at *Econometric Theory.*

[39] Palma, W. (2007) *Long-Memory Time Series*, New York: John Wiley and Sons.

[40] Parker, C., Paparoditis, E., and Politis, D. (2006) Unit root testing via the stationary bootstrap. *Journal of Econometrics* **133**, 601–638.

[41] Parzen, E. (1959) On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics* **28**, 329–348.

[42] Percival, D.B. and Walden, A.T.(1993). *Spectral analysis for physical applications. Multitaper and conventional univariate techniques.* Cambridge University Press, Cambridge.

[43] Pham, T. and Tran, L. (1985) Some mixing properties of time series models. *Stochastic Processes and Their Applications* **19**, 297–303.

[44] Phillips, P. and Perron, P. (1998) Testing for a unit root in time series regression. *Biometrika* **75**, 335–346.

[45] Politis, D. (2001) On nonparametric function estimation with infinite-order flat-top kernels, in *Probability and Statistical Models with applications*, Ch. Charalambides et al. (Eds.), Chapman and Hall/CRC: Boca Raton, pp/ 469–483.

[46] Politis, D. (2005) Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices, UCSD Dept. of Economics Discussion Paper 2005-03.

[47] Politis, D. and Romano, J. (1995) Bias-corrected nonparametric spectral estimation. *Journal of Time Series Analysis* **16**, 67–103.

[48] Politis, D., Romano, J., and Wolf, M. (1999) *Subsampling.* Springer-Verlag, New York.

[49] Priestley, M.B. (1981) *Spectral Analysis and Time Series*, Adademic Press, New York.

[50] Robinson, P. (1994) Semiparametric analysis of long memory time series. *Annals of Statistics* **22**, 515–539.

[51] Robinson, P. (2005) Robust covariance matrix estimation: HAC estimates with long memory/ antipersistence correction. *Econometric Theory* **21**, 171–180.

[52] Rosenblatt, M. (1956) A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences* **42**, 43–47.

[53] Rosenblatt, M. (1961) Independence and dependence. *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob., Vol.2*, University of California Press, 431–443.

[54] Rosenblatt, M. (1984) Stochastic processes with short-range and long-range dependence. In *Statistics: An Appraisal, Proceedings of the 50th Anniversary Conference* (David, H.A. and David, H.T., Eds.), The Iowa State University Press, Ames, IA, pp. 509520.

[55] Rosenblatt, M. (1985) *Stationary Sequences and Random Fields*, Birkhäuser, Boston.

[56] Samorodnitsky, G. and Taqqu, M. (1994) *Stable Non-Gaussian Random Processes*, New York City, New York: Chapman & Hall.

[57] Sun, Y. (2004) A convergent t-statistic in spurious regressions. *Econometric Theory* **20**, 943–962.

[58] Tanaka, K. (1990) Testing for a moving average unit root. *Econometric Theory* **6**, 433–444.

[59] Tanaka, K. (1996) *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory,* New York City, New York: John Wiley and Sons.

[60] Taniguchi, M. and Kakizawa, Y. (2000) *Asymptotic Theory of Statistical Inference for Time Series,* New York City, New York: Springer-Verlag.

[61] Taqqu, M. (1975) Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahr. Verw. Geb.* **31**, 287–302.

Differential Memory					
Sample Size	Tapers				
n= 250	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	82.3 88.9 95.5	75.9 85.5 94.5	75.3 85.3 94.7	58.5 65.3 75.4	58.9 67.6 81.7
	86.2 91.7 97.6	40.1 48.7 61.9	69.5 82.7 93.8	45.0 54.6 67.3	41.0 50.3 63.9
	100 100 100	100 100 100	89.4 92.9 95.3	100 100 100	100 100 100
$b = .5$	81.8 88.4 95.0	78.8 89.3 98.1	78.5 90.0 98.1	66.2 74.4 86.7	74.1 85.2 95.6
	88.1 93.4 97.9	87.9 95.5 99.0	81.0 92.8 97.7	62.3 74.7 90.8	84.8 97.0 99.3
	99.6 100 100	55.4 63.9 74.3	60.1 68.9 80.8	92.2 97.2 99.7	67.4 77.4 89.4
$b = 1$	82.8 88.9 94.3	83.8 91.4 97.8	75.3 85.9 94.8	72.0 82.3 92.4	77.9 86.4 95.6
	88.2 94.6 98.2	87.6 96.0 99.2	81.5 91.1 97.3	81.3 92.6 98.8	86.9 96.9 99.4
	93.9 9.80 99.8	91.7 97.3 99.4	79.2 89.2 96.6	86.9 97.1 99.6	93.5 98.9 100
n = 500	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	82.1 88.8 95.6	78.0 88.9 97.5	73.7 85.0 95.4	54.4 64.0 76.3	64.6 73.7 86.1
	86.7 92.7 98.0	79.4 90.7 98.5	79.8 89.6 97.4	50.6 58.5 71.6	56.0 68.2 86.8
	91.4 96.7 99.7	43.0 52.6 77.4	69.5 85.2 97.2	40.3 49.3 63.5	44.8 52.7 73.8
$b = .5$	80.2 87.4 94.3	78.5 88.6 98.1	78.8 89.4 98.0	67.3 75.7 87.6	74.2 84.4 95.3
	86.1 91.6 97.6	89.4 97.3 99.0	88.9 95.8 99.0	67.0 79.8 94.9	86.1 95.4 99.2
	91.1 96.9 100	96.6 98.4 99.7	91.9 96.9 99.2	66.8 79.1 97.9	91.2 97.0 99.4
$b = 1$	81.4 88.2 95.4	80.3 89.2 97.6	79.5 88.6 96.9	71.2 80.0 90.2	78.1 84.1 93.8
	87.2 93.1 98.7	88.5 95.4 98.8	81.5 89.9 97.5	84.1 93.7 99.2	82.8 93.2 99.6
	92.0 97.8 99.7	88.7 95.8 98.9	80.5 89.4 98.0	88.6 98.2 99.9	90.9 99.1 100
n = 1000	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	80.9 88.0 96.5	76.9 88.0 97.4	75.4 86.1 96.8	56.8 63.8 76.9	65.4 76.2 88.8
	86.1 92.1 97.8	82.7 92.9 98.7	76.8 90.2 97.8	51.2 58.6 71.7	62.8 74.4 92.8
	89.9 95.2 99.1	82.2 93.0 98.3	76.3 89.4 97.7	35.5 42.8 58.4	48.1 59.7 85.8
$b = .5$	81.9 87.6 95.0	75.3 86.1 97.2	77.5 87.8 98.1	67.0 73.8 83.6	70.3 79.3 93.3
	86.4 91.5 97.8	86.6 95.7 99.2	89.1 97.0 99.5	71.7 81.6 97.1	81.7 94.2 99.4
	90.0 95.5 99.8	96.3 98.2 99.6	93.9 98.1 99.7	69.3 84.4 99.3	94.0 98.5 99.5
$b = 1$	82.2 88.1 94.6	79.9 88.4 97.4	78.4 89.0 96.9	68.9 76.1 88.0	78.4 84.1 91.4
	87.1 93.1 97.9	85.8 94.4 99.6	81.2 90.9 98.4	80.1 91.8 99.4	83.2 92.3 99.3
	89.5 95.2 991.	88.7 96.9 99.0	79.3 90.8 98.3	88.7 98.8 100	91.5 99.3 100

Table 1: Empirical size for two-sided test with Type I error rate $\alpha = .10, .05, .01$ for the left, middle, and right hand entries respectively, in each cell. In each cell, the first row is for sampling fraction $a = .2$, the second row for $a = .1$, and the third row for $a = .04$. Various sample sizes and taper bandwidths b are considered. The DGP is Differential Memory.

Negative Memory: $\beta = -.8$					
Sample Size	Tapers				
n= 250	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	86.6 90.7 95.4	81.3 88.5 93.9	81.5 87.8 93.3	84.4 89.7 94.3	83.6 89.4 94.4
	88.9 93.7 97.6	76.6 83.9 90.4	81.7 89.0 95.5	83.3 89.9 95.7	82.0 87.8 93.1
	100 100 100	99.9 100 100	99.8 100 100	100 100 100	100 100 100
$b = .5$	84.5 90.6 95.0	85.0 90.2 95.8	84.5 91.1 96.5	81.7 87.6 92.2	83.4 89.0 94.5
	90.1 95.2 97.6	87.8 93.7 98.0	88.0 94.1 97.4	85.4 91.3 95.9	86.6 93.5 96.7
	99.4 100 100	93.8 96.9 99.4	89.6 94.1 97.3	99.2 99.8 100	95.8 98.8 100
$b = 1$	84.1 88.6 94.0	84.0 90.4 96.7	83.0 89.5 96.0	81.7 88.2 95.4	85.9 92.5 97.3
	90.1 94.1 97.1	88.2 94.7 97.8	88.9 94.8 98.1	89.4 94.5 98.2	88.0 94.5 98.2
	94.0 97.4 99.6	89.3 93.7 98.7	91.1 94.9 98.7	90.2 95.1 98.3	90.5 95.2 98.9
n = 500	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	84.7 90.9 96.6	82.0 87.4 92.0	80.0 85.8 91.7	80.9 86.9 93.8	78.5 84.2 91.2
	90.7 95.0 98.1	83.2 89.1 94.8	85.2 89.8 95.6	86.7 91.1 96.8	85.0 90.9 95.6
	95.7 98.4 99.6	85.2 90.6 96.5	88.4 92.2 97.7	89.7 93.5 98.2	86.2 92.3 96.8
$b = .5$	85.4 90.7 95.5	83.3 88.6 96.1	82.7 89.5 96.5	81.6 86.4 91.9	82.4 89.4 95.9
	91.4 94.8 98.0	87.4 93.3 98.5	86.9 92.9 98.0	85.5 91.2 96.9	88.2 93.0 97.6
	93.5 97.1 99.6	89.0 94.2 98.9	87.9 93.0 98.3	88.8 93.9 97.8	89.5 93.9 98.1
$b = 1$	86.8 92.4 96.6	86.8 93.0 97.4	83.7 91.1 97.4	82.0 88.5 94.4	84.1 90.1 97.0
	89.6 94.5 98.0	89.3 93.8 98.4	89.0 94.0 98.8	88.3 93.2 97.1	88.0 94.3 98.4
	90.9 96.3 99.5	89.9 93.4 98.6	89.4 95.0 98.8	90.2 95.0 99.5	89.7 94.0 98.4
n = 1000	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	86.1 90.7 96.1	84.0 89.3 95.0	80.5 87.2 93.4	82.3 89.3 94.0	81.4 86.4 92.3
	89.1 94.4 98.4	85.7 91.0 95.8	86.9 92.7 98.0	87.1 91.9 97.8	86.7 92.5 96.4
	92.4 97.4 99.2	86.9 92.8 97.8	87.9 93.3 98.5	86.3 91.2 97.6	87.6 93.2 98.2
$b = .5$	84.4 89.5 95.6	82.1 89.9 96.9	83.2 88.9 97.5	83.6 89.0 94.0	83.5 89.7 95.6
	87.9 93.6 97.9	88.7 94.0 98.4	89.2 94.2 98.2	86.7 91.5 97.4	87.6 93.1 98.5
	92.0 96.1 99.6	88.5 92.7 98.5	90.6 95.6 98.5	90.5 94.0 98.6	89.4 94.1 98.8
$b = 1$	85.3 90.5 95.1	85.2 93.0 98.5	84.6 90.7 97.9	85.0 92.0 96.9	84.9 91.6 97.3
	91.4 95.5 98.5	90.2 94.9 99.3	88.2 93.9 98.9	86.9 92.6 97.3	91.1 95.7 98.9
	92.5 97.3 99.7	88.8 94.3 98.8	91.1 95.4 99.3	88.0 94.8 99.1	90.8 95.6 99.4

Table 2: Empirical size for two-sided test with Type I error rate $\alpha = .10, .05, .01$ for the left, middle, and right hand entries respectively, in each cell. In each cell, the first row is for sampling fraction $a = .2$, the second row for $a = .1$, and the third row for $a = .04$. Various sample sizes and taper bandwidths b are considered. The DGP is Negative Memory with $\beta = -.8$.

Negative Memory: $\beta = -.6$					
Sample Size	Tapers				
n= 250	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	84.0 88.2 94.1	81.9 86.3 91.7	78.1 84.7 90.5	82.0 87.6 92.4	82.8 88.0 92.6
	89.9 94.8 97.6	81.2 87.0 92.9	84.2 90.3 95.4	85.3 92.3 96.5	84.5 90.5 94.5
	99.8 100 100	99.9 100 100	99.6 100 100	100 100 100	99.9 99.9 100
$b = .5$	85.3 89.8 94.0	84.8 89.0 94.9	81.8 89.6 96.5	82.4 87.6 91.8	82.2 88.0 94.3
	88.3 92.9 96.6	87.6 92.4 96.7	87.4 92.3 96.4	85.4 91.2 95.6	85.8 91.6 96.5
	98.4 99.1 99.5	94.0 97.0 98.8	90.9 93.4 97.4	97.4 99.2 100	95.7 98.2 99.5
$b = 1$	85.9 90.8 93.2	84.5 90.0 96.0	83.7 90.7 95.8	84.1 88.4 94.4	85.7 90.4 96.6
	89.0 93.3 96.6	87.5 93.3 97.3	89.1 93.5 97.6	86.3 92.0 96.1	88.7 94.4 99.0
	92.0 96.5 99.2	91.6 95.7 98.4	89.7 94.6 98.8	89.6 94.3 97.8	90.5 94.6 98.7
n = 500	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	84.7 88.8 94.4	80.3 86.7 92.8	82.8 88.2 92.6	84.7 89.8 95.3	84.2 88.8 95.3
	89.5 94.7 98.5	85.7 92.2 95.9	85.2 90.4 95.4	87.1 92.4 96.9	86.7 92.2 96.5
	92.9 96.0 98.8	86.2 91.8 97.3	88.9 94.0 98.6	89.9 95.6 98.6	88.1 92.5 97.4
$b = .5$	82.8 87.7 92.5	81.9 89.5 96.0	82.3 88.8 96.2	79.7 84.8 91.1	81.0 88.0 94.4
	87.6 91.9 96.7	86.5 91.9 97.3	86.6 92.6 98.3	87.5 92.7 96.8	87.8 92.4 97.9
	91.8 96.2 98.9	89.5 94.4 98.4	88.8 93.5 98.1	88.7 93.8 97.2	89.6 94.5 98.9
$b = 1$	86.9 91.0 95.2	83.6 90.7 97.3	85.3 91.2 97.7	79.4 87.0 94.0	85.2 90.7 97.1
	89.0 93.5 97.2	89.1 94.4 98.5	88.4 93.1 98.6	87.1 91.9 97.1	86.5 92.6 98.3
	92.6 96.5 99.2	90.0 94.9 98.8	87.3 93.4 98.4	89.3 94.0 98.6	89.7 95.2 99.0
n = 1000	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	83.6 89.1 93.7	82.1 86.5 91.5	83.3 88.1 92.8	83.7 89.7 94.4	81.6 87.6 93.7
	88.3 93.2 97.4	85.5 91.1 96.5	86.6 92.7 97.1	87.2 93.0 96.8	84.0 90.2 95.3
	93.2 96.4 99.1	90.5 95.3 98.7	88.3 92.6 97.8	90.3 94.0 97.9	89.0 93.2 98.2
$b = .5$	82.9 89.4 94.5	83.3 89.5 96.7	83.7 89.9 97.3	78.8 84.8 92.4	82.8 88.0 94.5
	90.3 93.9 97.9	85.8 92.5 97.9	88.0 93.4 98.0	86.5 91.8 96.8	87.0 92.3 97.5
	90.6 95.1 99.1	89.7 93.4 98.7	88.3 93.6 98.8	89.0 93.9 98.3	86.9 93.0 98.5
$b = 1$	86.0 91.0 96.0	85.5 91.0 96.0	86.0 92.4 97.4	84.5 89.8 96.0	84.9 91.3 97.1
	88.1 93.4 97.2	87.4 93.3 98.0	87.2 91.8 98.4	85.2 91.7 97.1	87.5 94.1 98.7
	91.8 95.9 99.1	88.2 94.3 98.7	88.7 93.3 99.2	88.3 93.5 98.5	89.5 95.5 99.5

Table 3: Empirical size for two-sided test with Type I error rate $\alpha = .10, .05, .01$ for the left, middle, and right hand entries respectively, in each cell. In each cell, the first row is for sampling fraction $a = .2$, the second row for $a = .1$, and the third row for $a = .04$. Various sample sizes and taper bandwidths b are considered. The DGP is Negative Memory with $\beta = -.6$.

Negative Memory: $\beta = -.4$					
Sample Size	Tapers				
n= 250	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	85.2 89.0 93.2	82.2 86.5 92.0	82.2 86.6 90.7	83.7 88.7 92.6	79.7 84.9 90.8
	87.5 92.0 96.0	84.3 89.1 93.8	83.9 88.5 93.6	86.6 90.8 94.4	83.0 88.7 93.3
	99.5 99.9 100	99.3 99.9 100	98.8 99.6 99.9	99.6 100 100	99.5 100 100
$b = .5$	79.2 83.2 88.9	80.8 86.8 93.4	82.2 87.8 95.5	81.3 85.8 90.2	80.7 86.0 92.8
	84.9 90.5 95.2	84.9 91.2 96.5	85.0 92.6 96.8	85.1 90.1 94.5	87.2 92.9 97.2
	96.9 98.7 99.7	92.5 96.0 98.6	90.3 94.1 97.7	96.7 98.6 99.6	94.1 96.9 99.3
$b = 1$	80.7 87.5 91.8	85.2 91.0 97.3	83.8 89.6 95.8	83.2 88.1 92.8	86.4 91.6 97.0
	89.0 94.0 97.4	90.1 95.1 98.3	87.6 94.6 98.2	86.8 92.2 96.6	86.0 92.0 97.0
	90.1 95.1 98.3	90.5 94.8 98.1	90.4 94.5 98.5	90.3 94.9 98.0	89.9 94.4 98.3
n = 500	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	81.2 86.8 92.4	80.0 85.3 90.5	82.0 87.1 92.1	83.5 88.0 92.8	82.3 88.1 93.0
	87.0 90.4 95.8	85.0 89.4 94.7	86.6 91.0 95.0	85.9 91.2 95.3	87.2 91.5 95.2
	90.9 95.8 99.0	86.6 91.4 96.2	89.7 94.2 98.6	89.0 93.9 97.6	88.6 93.2 97.6
$b = .5$	82.4 87.3 92.2	81.6 88.9 95.7	81.1 88.9 95.6	81.5 85.8 89.6	81.7 87.7 94.3
	87.2 92.1 96.2	87.7 93.3 97.8	86.9 92.6 98.2	85.9 91.1 96.4	85.8 92.0 97.1
	90.8 96.0 98.5	87.3 94.2 98.1	88.9 94.4 98.2	88.6 93.5 97.6	88.7 94.1 98.3
$b = 1$	84.1 87.1 91.5	84.0 90.5 96.6	83.2 89.8 96.9	82.3 87.7 93.4	82.9 90.1 96.8
	86.2 91.2 96.4	88.3 94.0 98.1	86.1 92.1 97.8	87.5 92.5 96.6	88.1 93.5 98.4
	91.2 95.1 98.5	88.4 93.6 99.0	88.9 94.7 98.9	89.1 94.7 99.3	88.9 94.1 98.7
n = 1000	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	84.9 88.7 92.7	82.9 87.8 92.1	82.0 87.2 91.1	83.3 88.0 92.9	83.3 89.1 92.7
	88.5 92.3 95.7	84.2 90.2 95.2	87.2 92.2 96.2	85.9 90.3 95.7	86.4 90.8 95.4
	88.4 93.4 97.6	89.5 93.8 97.6	89.3 93.9 97.5	87.7 93.0 98.1	88.3 93.0 96.7
$b = .5$	82.6 87.1 92.6	83.0 87.8 94.9	83.6 88.2 94.5	80.9 85.7 90.5	81.0 86.8 93.1
	88.7 92.8 96.2	84.6 90.9 98.4	86.8 93.1 98.2	85.5 92.2 95.9	86.5 91.2 97.1
	90.0 94.0 97.9	89.4 95.0 99.2	86.8 93.8 99.3	89.3 95.0 98.4	89.2 94.3 98.0
$b = 1$	82.6 88.6 93.1	84.3 90.6 97.3	86.4 92.2 98.4	81.0 86.8 93.2	86.1 91.3 97.1
	86.6 91.6 96.3	87.1 93.6 98.4	86.7 93.5 98.9	86.5 91.9 96.2	86.5 92.8 98.1
	88.0 94.1 98.3	88.1 93.7 98.6	87.6 94.4 99.4	88.9 93.8 98.4	90.4 94.7 99.1

Table 4: Empirical size for two-sided test with Type I error rate $\alpha = .10, .05, .01$ for the left, middle, and right hand entries respectively, in each cell. In each cell, the first row is for sampling fraction $a = .2$, the second row for $a = .1$, and the third row for $a = .04$. Various sample sizes and taper bandwidths b are considered. The DGP is Negative Memory with $\beta = -.4$.

Negative Memory: $\beta = -.2$					
Sample Size	Tapers				
n= 250	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	79.0 83.8 88.6	79.3 84.1 88.1	80.1 83.7 88.1	79.8 85.4 90.9	78.6 84.2 89.7
	84.9 89.7 93.7	83.2 87.7 92.2	83.8 88.3 92.8	84.0 89.6 93.8	82.8 88.9 92.9
	97.6 99.3 99.8	97.1 99.0 99.9	96.4 98.4 99.5	97.8 99.1 99.6	96.6 99.2 99.8
$b = .5$	80.1 85.3 90.3	80.4 86.5 92.6	80.7 87.2 94.7	80.6 84.6 89.0	79.6 84.7 91.1
	85.5 92.3 94.8	84.6 90.2 96.7	84.7 91.3 96.1	86.2 90.9 95.2	84.8 90.0 95.2
	94.2 97.1 99.7	91.3 94.4 98.5	90.3 93.9 97.8	93.3 96.5 99.3	91.6 94.7 98.4
$b = 1$	80.2 86.0 91.0	82.7 88.6 95.4	81.3 87.8 94.8	81.0 85.8 91.2	83.5 90.1 96.1
	85.6 90.4 94.4	88.0 93.6 97.5	86.6 93.5 97.3	87.3 91.6 95.7	87.2 93.8 98.0
	87.2 93.8 97.7	89.2 93.9 98.7	87.7 93.3 98.2	87.6 92.5 97.3	88.6 94.8 98.4
n = 500	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	80.4 85.8 90.0	80.9 85.4 90.1	79.8 84.6 89.0	80.8 85.1 90.9	80.5 84.3 90.3
	87.1 91.8 95.3	86.8 90.0 93.8	85.3 90.7 95.4	85.6 89.3 94.3	85.5 89.3 93.1
	90.7 95.4 98.5	86.5 92.0 96.8	89.8 94.9 98.5	86.4 91.6 97.1	90.3 93.9 97.6
$b = .5$	79.1 84.6 88.9	81.3 87.6 94.3	82.0 89.7 95.2	81.3 85.8 91.4	81.0 86.8 92.6
	88.2 91.6 95.7	86.3 91.7 96.6	87.2 92.9 98.1	84.3 89.0 94.3	84.4 89.5 95.2
	89.9 94.7 98.2	90.3 94.3 98.0	86.7 93.1 98.6	87.6 92.5 96.8	88.2 93.9 98.3
$b = 1$	81.4 86.4 91.2	82.6 88.8 96.6	84.8 90.6 96.7	80.9 86.6 92.2	83.3 89.7 95.7
	84.5 89.9 95.1	87.8 92.6 97.7	86.7 92.2 98.0	87.9 93.0 96.5	88.0 93.0 98.5
	89.8 94.8 98.5	89.8 94.7 98.9	87.4 93.1 98.7	88.6 93.9 98.2	90.5 94.9 99.4
n = 1000	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	81.2 86.2 90.9	78.9 83.8 88.7	79.9 85.7 90.4	80.1 85.5 90.1	81.4 86.2 91.0
	86.9 89.5 94.6	85.6 90.2 95.1	86.4 90.3 94.8	84.6 89.8 94.3	83.8 89.9 95.5
	89.4 94.3 97.9	88.6 93.3 97.2	89.1 94.3 97.7	90.6 94.6 97.8	88.4 92.5 97.3
$b = .5$	80.7 85.1 89.7	82.7 88.2 94.8	81.5 87.0 95.7	79.2 84.1 90.3	80.6 87.5 93.5
	85.4 90.5 96.0	86.7 92.1 97.9	84.4 91.3 98.3	84.7 90.8 95.3	85.7 89.7 97.0
	87.0 91.9 97.9	86.6 92.3 97.6	87.9 94.1 98.9	88.0 92.8 97.7	87.7 93.1 98.6
$b = 1$	82.9 87.2 90.7	85.3 91.4 97.0	83.3 89.1 96.9	83.0 88.7 94.1	83.1 89.3 96.1
	85.8 91.2 95.6	86.5 92.2 98.2	87.5 92.8 98.3	86.7 92.4 97.1	86.5 91.6 97.5
	89.7 94.4 98.2	88.4 93.4 98.5	87.8 93.2 98.6	87.2 93.0 97.5	87.9 93.3 98.4

Table 5: Empirical size for two-sided test with Type I error rate $\alpha = .10, .05, .01$ for the left, middle, and right hand entries respectively, in each cell. In each cell, the first row is for sampling fraction $a = .2$, the second row for $a = .1$, and the third row for $a = .04$. Various sample sizes and taper bandwidths b are considered. The DGP is Negative Memory with $\beta = -.2$.

Short Memory					
Sample Size	Tapers				
n= 250	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	77.8 82.0 86.2	77.3 82.1 85.7	77.8 82.8 87.6	78.6 83.7 88.4	77.0 82.6 85.7
	84.1 89.9 93.1	80.7 86.3 90.6	81.8 88.4 93.0	85.7 89.9 93.3	85.8 90.0 93.4
	91.6 95.8 98.0	91.5 95.5 98.6	92.2 95.9 98.7	91.4 96.4 98.9	92.6 95.7 98.8
$b = .5$	79.4 84.7 88.0	79.3 85.4 91.2	81.5 86.4 93.8	80.0 84.1 87.5	77.3 83.2 89.6
	84.3 89.9 93.7	84.6 89.4 94.9	85.2 92.4 96.9	85.0 90.2 93.6	83.3 89.3 94.9
	89.6 94.7 98.3	89.7 94.1 97.8	89.0 92.8 97.2	91.2 95.3 98.4	90.1 93.6 98.2
$b = 1$	79.3 83.4 87.9	81.3 88.6 94.9	82.7 89.9 95.8	79.5 84.6 88.9	83.0 88.4 94.2
	81.4 87.9 92.4	87.1 92.9 97.2	85.5 91.9 96.9	84.0 90.7 94.7	84.2 91.8 97.8
	87.4 92.0 97.2	88.9 93.3 97.7	88.7 92.9 98.1	89.5 94.2 98.2	89.0 93.9 97.8
n = 500	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	77.7 83.9 87.5	77.4 81.3 85.0	81.7 86.7 90.6	78.7 82.5 87.0	77.6 82.8 86.7
	84.6 88.8 92.1	84.5 89.4 94.4	85.4 90.3 94.1	84.8 88.9 93.3	82.9 88.1 92.6
	88.2 92.9 97.2	87.7 92.4 96.7	87.4 93.1 97.7	86.9 92.1 97.0	88.3 93.6 97.3
$b = .5$	80.6 84.9 89.1	81.1 85.2 91.3	81.7 88.7 94.8	75.7 80.4 86.3	78.7 83.9 90.7
	85.1 89.4 93.9	86.5 91.6 97.7	85.6 91.9 97.2	84.0 88.9 93.4	83.7 89.7 96.5
	86.9 91.6 96.8	88.2 93.8 98.7	87.0 92.8 98.4	87.0 92.3 97.2	90.0 93.4 98.3
$b = 1$	78.5 84.6 88.8	83.2 89.1 95.3	83.6 89.5 95.5	81.2 85.6 90.1	82.8 88.4 94.5
	85.3 91.1 95.6	85.3 90.6 97.0	88.3 93.2 97.7	86.7 90.6 95.6	87.2 93.3 97.6
	86.7 92.5 97.3	87.2 93.1 98.6	87.8 94.4 98.4	86.5 92.5 98.0	87.4 93.9 99.0
n = 1000	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	77.5 82.3 86.3	77.9 82.2 86.9	80.1 84.2 87.8	78.9 82.9 87.8	78.7 82.4 87.1
	84.4 89.1 93.6	84.2 89.5 92.9	84.9 89.3 93.7	84.8 89.0 93.2	84.6 89.1 93.0
	86.4 93.1 97.2	87.8 92.5 97.7	88.6 93.9 97.5	88.5 93.1 96.8	87.2 91.9 97.5
$b = .5$	76.6 81.6 86.2	79.6 85.3 92.5	80.5 87.0 94.0	77.5 83.6 88.0	80.0 84.7 91.1
	83.4 88.1 93.9	84.6 90.7 97.5	85.8 91.9 97.8	83.1 88.8 93.6	86.1 90.1 95.9
	88.0 93.0 96.7	89.2 93.4 97.7	88.2 92.4 98.4	88.6 93.7 97.7	86.7 91.3 98.0
$b = 1$	77.2 82.3 88.2	82.9 89.4 96.2	80.8 87.7 95.9	76.0 82.2 87.7	81.2 87.3 94.0
	85.3 90.2 93.9	88.4 93.2 99.1	85.1 91.8 97.7	88.3 92.6 96.3	85.2 91.9 97.6
	87.9 92.4 96.9	87.0 93.3 98.1	89.6 93.7 99.0	88.5 92.6 98.2	88.7 93.6 98.6

Table 6: Empirical size for two-sided test with Type I error rate $\alpha = .10, .05, .01$ for the left, middle, and right hand entries respectively, in each cell. In each cell, the first row is for sampling fraction $a = .2$, the second row for $a = .1$, and the third row for $a = .04$. Various sample sizes and taper bandwidths b are considered. The DGP is Short Memory.

Long Memory: $\beta = .2$					
Sample Size	Tapers				
n= 250	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	78.2 82.1 84.7	81.3 85.6 88.5	75.0 78.4 82.4	80.0 82.6 85.9	80.2 83.7 86.6
	89.5 92.4 94.3	90.1 93.1 94.6	87.7 91.8 94.1	91.0 94.2 95.6	89.7 92.5 94.4
	80.8 87.6 94.7	81.4 88.6 95.6	84.0 90.3 95.5	75.7 84.5 93.6	84.3 90.0 95.6
$b = .5$	79.2 82.6 85.3	78.6 83.4 88.7	81.1 85.7 91.7	77.9 81.3 84.4	78.3 82.8 87.2
	85.6 90.9 94.1	83.5 89.2 93.3	84.5 90.2 95.5	86.2 90.6 95.0	83.3 89.3 94.3
	93.8 96.2 98.3	94.4 97.1 99.2	95.2 97.2 99.1	92.7 96.1 98.5	95.8 97.6 99.1
$b = 1$	77.5 81.8 86.3	81.1 87.4 92.7	81.4 87.4 94.8	76.8 81.6 86.8	80.3 86.5 91.4
	87.4 92.8 95.6	85.0 91.6 97.0	88.8 94.1 98.1	83.7 90.2 94.4	84.7 90.7 96.3
	91.1 94.8 98.3	91.8 95.2 98.8	89.7 94.0 97.8	91.8 96.0 99.0	90.1 94.7 98.7
n = 500	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	78.2 82.9 85.8	77.5 81.9 85.1	75.8 79.2 83.3	78.2 81.7 85.4	77.1 82.4 85.8
	87.7 91.6 94.0	85.1 88.8 92.6	87.3 91.4 94.6	89.7 91.5 94.9	87.6 90.9 93.8
	94.3 97.3 98.9	93.4 96.0 98.4	94.4 96.9 98.7	95.1 97.5 98.8	95.4 97.5 98.9
$b = .5$	77.6 81.1 85.2	77.1 81.8 87.7	77.9 84.6 92.4	77.8 82.6 86.1	79.8 84.9 89.3
	85.5 90.7 94.6	84.3 89.6 95.7	84.0 90.3 95.9	85.4 90.0 93.9	83.8 89.3 94.7
	92.3 95.8 98.3	91.5 94.1 98.3	89.2 94.8 99.4	89.9 94.1 98.1	89.0 93.6 98.3
$b = 1$	79.3 84.1 87.4	79.3 86.7 93.1	81.5 87.9 95.1	77.5 83.0 88.4	79.6 85.9 92.9
	88.4 92.6 96.0	86.7 92.5 97.5	88.3 92.3 97.9	84.9 88.9 93.5	86.7 90.7 96.5
	91.5 94.9 97.7	89.3 94.5 98.8	92.3 95.4 98.8	89.4 94.7 98.6	89.3 94.7 97.6
n = 1000	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	79.3 82.5 86.1	78.1 82.1 84.8	78.3 82.6 84.7 ‘	78.5 82.4 86.5	78.0 82.1 84.6
	85.8 90.4 93.8	85.2 89.8 93.8	86.6 90.4 94.9	88.4 91.0 93.6	85.4 89.9 93.1
	93.1 95.8 98.1	93.1 96.1 98.3	92.6 95.6 98.0	93.8 97.0 99.0	92.6 96.1 98.8
$b = .5$	76.1 81.2 85.0	78.2 83.6 89.0	77.2 83.8 92.7	76.2 79.7 83.1	77.7 82.3 87.1
	84.5 89.8 93.2	83.7 90.0 94.7	85.5 91.9 96.7	84.0 89.1 92.9	84.8 89.7 95.1
	91.3 95.3 98.8	87.3 92.5 98.2	90.2 95.1 98.4	90.6 94.2 98.2	90.5 94.6 98.0
$b = 1$	77.9 83.0 87.8	82.8 88.0 94.5	81.1 89.1 96.1	78.7 83.1 87.9	80.4 86.0 92.8
	84.4 90.1 94.7	86.8 93.0 97.5	88.1 92.8 98.5	86.1 91.3 94.9	84.0 90.7 96.8
	91.7 94.7 98.3	90.2 94.2 98.6	91.4 95.7 99.5	88.9 93.8 97.6	90.7 94.2 98.4

Table 7: Empirical size for two-sided test with Type I error rate $\alpha = .10, .05, .01$ for the left, middle, and right hand entries respectively, in each cell. In each cell, the first row is for sampling fraction $a = .2$, the second row for $a = .1$, and the third row for $a = .04$. Various sample sizes and taper bandwidths b are considered. The DGP is Long Memory with $\beta = .2$.

Long Memory: $\beta = .4$					
Sample Size	Tapers				
n= 250	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	76.7 80.6 83.4	74.3 78.0 80.3	76.1 79.2 82.1	75.9 78.9 81.7	76.0 79.5 83.0
	87.4 90.7 94.1	87.5 92.0 94.4	85.7 89.6 92.0	87.0 90.9 93.2	86.8 90.1 93.4
	73.4 80.8 88.1	76.1 83.9 91.6	81.6 87.1 93.5	69.8 77.5 86.8	76.0 83.7 91.7
$b = .5$	72.6 76.8 81.7	76.7 80.8 85.9	79.2 84.9 90.0	74.6 79.1 81.8	74.4 78.4 84.4
	83.6 88.2 92.3	81.9 88.0 93.6	83.3 89.6 95.1	84.8 90.0 92.6	82.8 88.9 94.0
	90.0 95.0 98.0	93.3 96.5 98.4	94.4 96.7 98.7	90.3 94.5 97.4	91.0 95.0 97.3
$b = 1$	73.3 79.2 83.8	76.7 83.6 91.7	78.3 85.8 93.0	76.5 81.3 86.3	76.4 82.8 89.2
	83.2 88.5 92.1	84.6 90.4 96.2	84.2 91.6 97.6	82.7 88.5 93.1	84.0 89.6 94.8
	91.2 95.7 98.5	89.0 94.4 98.0	91.1 95.4 98.6	90.7 95.5 99.2	89.5 93.8 97.8
n = 500	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	75.8 80.7 83.7	74.7 77.8 80.5	74.3 78.6 81.9	78.2 81.5 83.8	76.9 80.0 83.4
	86.8 90.3 93.4	86.0 89.4 92.1	85.4 89.3 92.6	86.9 89.8 94.2	84.9 88.5 91.8
	92.2 95.7 98.0	93.4 96.3 98.7	93.6 96.6 98.7	94.1 96.4 98.8	95.0 97.7 98.9
$b = .5$	75.0 79.3 83.0	75.2 80.7 85.9	76.6 81.8 89.7	73.3 78.3 81.7	76.1 79.6 85.2
	83.6 87.9 91.9	83.2 87.8 94.1	83.4 89.1 95.7	81.8 86.9 91.7	85.1 91.0 94.7
	89.2 93.9 98.3	90.3 93.5 98.3	89.4 93.1 97.5	89.2 94.1 97.6	87.5 93.5 97.3
$b = 1$	73.9 78.8 83.8	79.5 85.2 93.6	79.0 85.6 94.7	75.6 80.0 85.8	79.1 84.0 90.5
	83.1 88.2 93.3	83.8 90.4 96.4	84.4 90.7 96.4	82.8 88.2 92.6	82.5 88.5 94.6
	90.7 94.3 98.1	87.8 93.0 98.1	90.8 94.4 98.7	90.2 94.3 98.7	89.3 94.5 98.4
n = 1000	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	75.5 78.7 82.0	74.6 78.1 80.8	71.5 76.1 80.6	74.9 78.2 82.5	74.6 78.6 82.4
	87.6 90.2 93.0	85.1 89.1 92.0	84.2 87.8 91.9	85.1 89.0 91.7	84.9 88.8 91.8
	91.8 95.7 98.2	91.1 94.0 97.7	89.3 93.9 97.6	92.9 96.5 99.0	90.7 94.0 97.0
$b = .5$	73.7 77.8 81.3	75.5 80.2 85.4	75.3 82.3 89.9	74.9 80.2 84.3	73.3 79.5 85.9
	83.5 89.4 93.6	82.4 88.8 93.8	83.8 89.8 96.0	83.4 87.8 92.0	84.1 89.8 94.7
	90.4 94.7 97.8	88.5 94.2 97.9	88.3 92.7 97.4	86.7 91.4 96.1	88.4 93.0 97.4
$b = 1$	76.0 80.9 85.0	79.0 85.0 93.1	80.4 86.8 94.3	75.1 82.0 87.3	77.9 83.9 91.0
	81.8 86.7 91.1	85.0 91.2 97.7	84.6 91.4 97.0	82.2 87.0 92.6	87.3 92.2 96.5
	89.3 93.6 97.9	88.2 93.1 98.5	88.8 94.4 98.5	90.1 94.4 98.2	88.8 93.9 98.3

Table 8: Empirical size for two-sided test with Type I error rate $\alpha = .10, .05, .01$ for the left, middle, and right hand entries respectively, in each cell. In each cell, the first row is for sampling fraction $a = .2$, the second row for $a = .1$, and the third row for $a = .04$. Various sample sizes and taper bandwidths b are considered. The DGP is Long Memory with $\beta = .4$.

Long Memory: $\beta = .6$					
Sample Size	Tapers				
n= 250	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	70.9 74.7 77.4	71.6 76.6 79.3	70.5 74.6 77.5	71.3 74.6 78.1	70.9 74.9 76.7
	82.3 86.2 89.1	84.0 88.5 91.7	81.5 86.8 90.5	81.8 85.8 88.9	84.9 88.5 91.7
	61.2 70.5 81.4	65.3 73.2 83.4	73.3 80.3 87.9	59.0 67.6 80.2	63.7 72.0 81.0
$b = .5$	71.6 76.6 80.5	70.8 78.0 84.1	72.0 78.3 84.9	67.7 72.6 75.1	71.3 76.3 82.5
	79.4 84.8 89.9	79.5 84.7 90.9	79.9 86.9 93.8	80.4 85.9 89.3	80.2 86.8 92.2
	86.7 90.4 95.1	89.5 93.4 96.6	88.9 93.5 97.0	86.9 91.0 95.1	87.1 90.9 95.3
$b = 1$	71.2 75.9 80.2	73.9 80.8 90.8	74.8 82.4 90.9	69.7 75.5 80.8	75.9 81.4 88.3
	82.0 87.7 92.0	81.3 88.9 94.0	82.8 90.0 95.3	80.6 85.8 91.8	83.2 89.3 95.0
	86.8 91.1 96.4	89.6 94.0 98.1	88.2 93.8 97.7	88.3 93.0 97.9	87.9 92.4 97.4
n = 500	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	70.6 74.3 77.7	67.3 71.7 74.9	67.8 71.5 74.6	69.4 73.5 76.8	70.4 74.8 78.0
	79.4 83.6 88.7	83.1 86.8 90.2	81.5 85.9 89.1	81.3 84.8 88.1	79.1 84.6 88.2
	88.3 92.5 97.0	88.5 92.6 97.6	89.8 94.0 97.5	91.1 94.3 97.1	89.0 93.2 97.0
$b = .5$	70.2 74.2 78.2	73.8 79.5 84.6	73.0 80.2 87.8	68.9 73.2 76.7	72.4 76.6 81.6
	81.6 87.2 90.9	77.2 84.6 92.6	79.2 86.5 93.9	79.3 84.2 88.8	81.7 88.2 92.7
	85.9 92.2 96.9	86.2 91.3 97.3	84.8 91.3 97.6	87.2 91.8 97.4	85.7 90.3 96.3
$b = 1$	71.5 77.1 81.6	75.6 82.4 92.9	77.1 84.2 93.0	68.7 75.7 81.9	74.2 81.0 88.2
	77.6 83.5 88.9	84.0 89.4 96.8	82.3 88.9 96.5	82.7 86.9 92.8	80.0 86.0 93.9
	87.2 91.5 96.5	87.7 92.8 98.2	88.2 94.6 98.5	87.8 93.0 97.6	89.2 94.4 99.0
n = 1000	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	68.6 73.0 76.4	69.8 73.8 76.2	67.8 71.8 75.2	67.0 70.6 74.1	68.1 71.5 74.8
	79.4 83.9 88.2	79.7 84.5 88.7	77.7 82.5 87.6	81.2 85.1 88.9	81.1 86.9 90.5
	87.4 91.2 95.8	88.0 92.2 96.3	86.6 93.0 97.3	88.2 92.6 96.0	88.3 93.3 96.3
$b = .5$	71.9 77.1 80.8	71.3 77.3 83.0	70.8 76.9 85.4	68.5 74.2 77.6	72.2 77.6 83.5
	79.5 84.7 90.0	80.9 86.5 93.2	80.4 87.7 94.7	77.2 82.8 87.8	79.3 85.7 92.8
	85.1 90.9 96.0	86.8 91.5 97.7	87.2 94.0 98.2	84.5 90.3 95.7	86.7 91.5 97.4
$b = 1$	69.5 74.9 79.6	73.8 81.3 91.4	74.4 82.2 92.3	72.4 78.8 83.2	72.8 80.3 87.9
	80.3 85.6 93.1	83.5 89.9 96.5	83.1 90.9 97.5	79.2 84.8 92.6	81.4 89.0 96.1
	86.4 92.5 97.7	86.7 92.7 97.8	86.4 92.8 98.4	85.1 91.6 96.9	86.3 91.5 97.2

Table 9: Empirical size for two-sided test with Type I error rate $\alpha = .10, .05, .01$ for the left, middle, and right hand entries respectively, in each cell. In each cell, the first row is for sampling fraction $a = .2$, the second row for $a = .1$, and the third row for $a = .04$. Various sample sizes and taper bandwidths b are considered. The DGP is Long Memory with $\beta = .6$.

Long Memory: $\beta = .8$					
Sample Size	Tapers				
n= 250	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	59.8 63.0 65.8	59.2 62.4 64.7	60.6 64.2 68.3	58.9 63.3 65.9	60.4 63.9 66.5
	72.8 78.4 81.6	75.8 80.5 83.9	76.0 81.2 84.9	73.3 78.6 83.4	73.9 78.9 82.4
	49.3 58.7 68.3	50.0 59.6 69.9	56.0 64.1 73.7	43.6 50.9 61.5	49.8 56.9 68.4
$b = .5$	61.1 66.8 70.9	61.8 68.3 76.2	65.0 71.6 81.2	58.8 63.8 67.5	62.6 67.0 72.7
	73.9 81.2 86.5	70.6 78.6 86.8	76.4 84.0 92.4	72.0 78.7 84.5	74.2 81.2 86.5
	74.8 81.9 89.0	77.6 84.5 91.6	82.1 87.3 92.8	72.4 78.9 87.4	79.4 85.2 91.8
$b = 1$	62.3 67.2 72.6	68.1 76.8 85.0	66.8 75.9 85.9	64.2 70.5 76.6	65.4 72.0 83.0
	74.9 80.6 86.4	75.9 84.4 92.8	79.0 87.0 94.4	75.5 82.3 89.3	76.1 84.9 92.0
	82.1 88.6 93.9	85.0 91.1 97.7	85.4 91.9 97.3	80.8 87.3 94.2	83.7 89.9 96.5
n = 500	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	59.0 62.9 66.0	59.6 63.7 66.8	59.5 63.9 66.0	55.6 59.0 62.4	57.2 61.2 63.6
	69.0 74.1 79.1	71.1 75.9 80.9	69.4 74.7 79.4	72.3 76.7 81.8	71.8 75.7 81.8
	79.3 84.9 92.3	81.0 85.9 91.7	81.2 86.2 92.2	81.4 87.0 92.1	80.9 86.2 92.8
$b = .5$	59.5 64.2 68.8	64.4 70.8 78.1	63.0 70.9 81.1	60.1 64.1 68.6	62.0 68.0 74.7
	69.2 75.5 83.7	74.1 81.9 89.9	72.2 80.3 91.3	72.9 79.8 85.9	73.5 80.9 88.7
	82.8 88.3 94.5	78.4 86.3 95.3	82.5 89.1 96.6	81.7 87.5 95.1	81.9 87.7 96.0
$b = 1$	63.3 69.9 75.0	68.6 77.6 87.3	63.2 73.5 86.1	62.8 70.0 75.4	66.7 74.7 84.0
	75.1 79.7 88.0	78.3 85.0 94.0	77.6 84.3 94.7	73.6 80.2 86.8	76.3 85.2 93.8
	79.9 87.1 94.3	84.0 90.4 97.6	83.9 90.2 96.8	80.3 88.3 95.5	82.0 89.2 96.3
n = 1000	Bartlett	Trapezoid (.25)	Trapezoid (.5)	Parzen	Daniell
$b = .1$	56.6 60.3 63.7	55.6 60.3 63.2	54.8 60.4 64.6	55.4 58.4 61.2	57.7 61.1 64.9
	68.5 74.6 80.4	70.4 75.5 79.7	67.6 74.1 80.5	65.5 71.9 77.7	68.1 73.3 78.6
	79.3 85.7 92.1	80.2 86.4 91.0	78.8 83.1 89.9	80.6 85.3 91.8	80.2 86.5 92.4
$b = .5$	63.1 67.4 70.7	63.1 69.2 77.3	67.1 74.8 83.9	61.1 65.2 68.8	62.4 68.8 76.9
	69.5 76.3 83.6	75.9 82.1 89.5	75.5 83.5 91.9	71.4 78.4 83.8	69.7 78.5 86.2
	83.8 88.7 95.1	79.2 86.9 95.8	80.8 87.7 95.4	81.5 87.6 93.4	77.3 85.8 94.2
$b = 1$	61.3 66.7 71.6	69.0 76.7 87.3	66.4 75.5 87.3	61.0 67.7 74.5	67.9 74.7 83.2
	74.0 80.0 86.2	78.1 86.1 94.8	74.4 81.9 93.1	74.4 81.9 89.1	76.1 82.6 91.8
	79.8 86.4 93.8	80.4 87.4 96.3	82.8 89.8 98.5	78.6 86.3 94.4	83.5 89.9 96.4

Table 10: Empirical size for two-sided test with Type I error rate $\alpha = .10, .05, .01$ for the left, middle, and right hand entries respectively, in each cell. In each cell, the first row is for sampling fraction $a = .2$, the second row for $a = .1$, and the third row for $a = .04$. Various sample sizes and taper bandwidths b are considered. The DGP is Long Memory with $\beta = .8$.

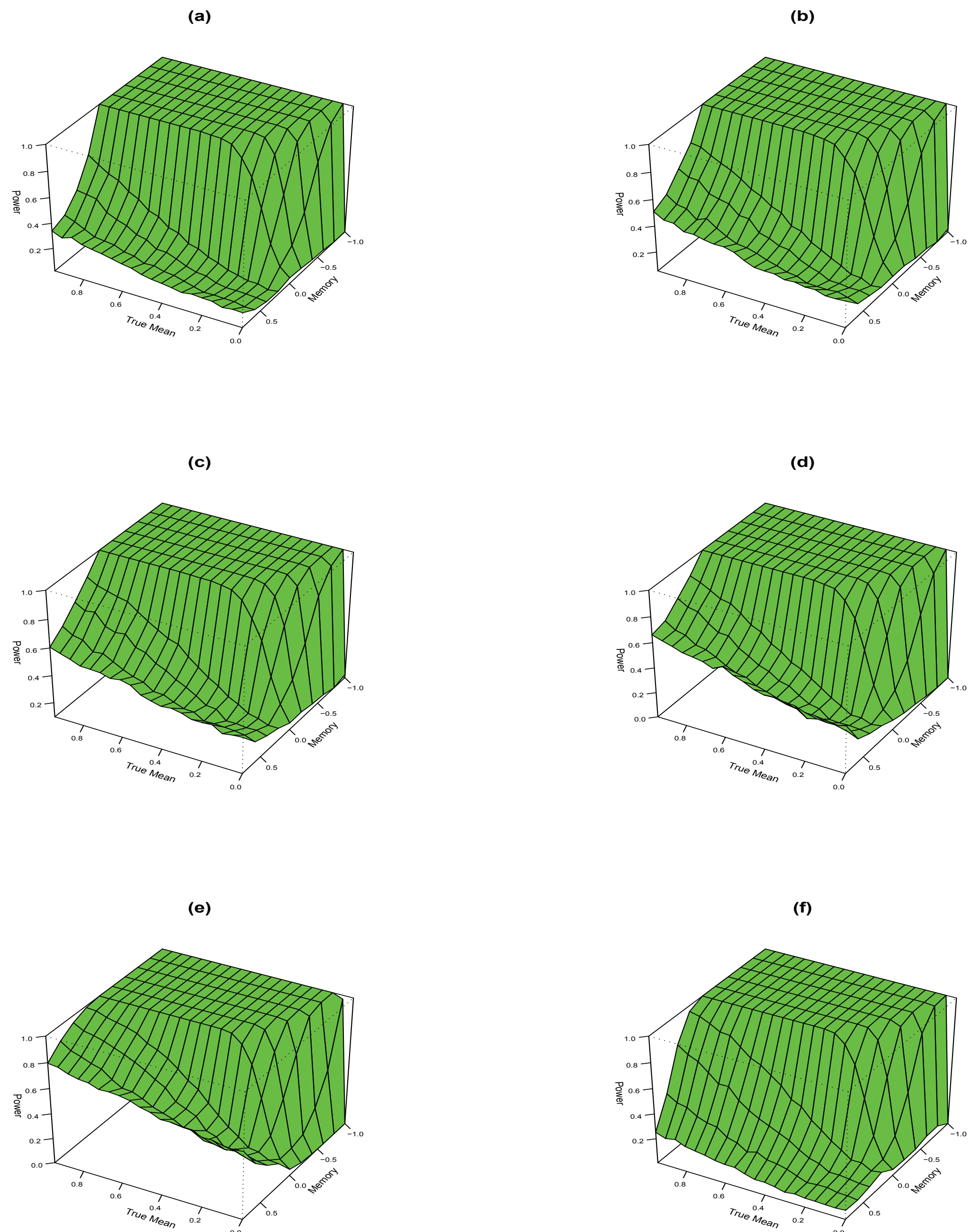

Figure 1: Power surfaces by Memory parameter $\beta \in \{-1, -.8, -.6, -.4, -.2, 0, .2, .4, .6., .8\}$ and true mean $\mu \in [0, 1)$. These power surfaces are for the Bartlett taper with bandwidth fraction $b = .2$. Panel (a) corresponds to subsampling with sampling fraction $a = .04$; panel (b) corresponds to subsampling with sampling fraction $a = .12$; panel (c) corresponds to subsampling with sampling fraction $a = .2$. Panel (d) corresponds to using the plug-in estimate of β; panel (e) corresponds to using the plug-in method with $\beta = 0$; panel (f) corresponds to using the true (unknown) value of β.

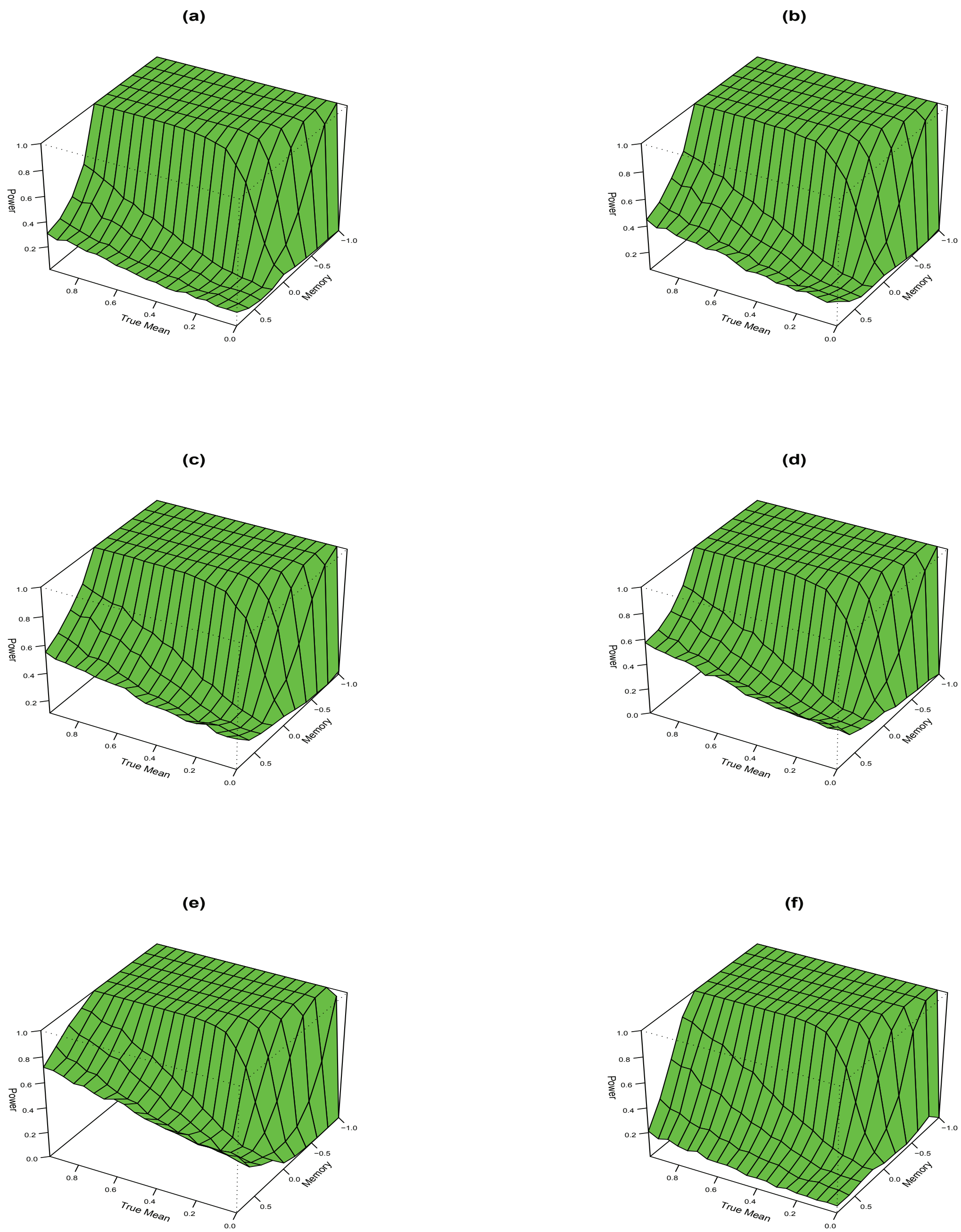

Figure 2: Power surfaces by Memory parameter $\beta \in \{-1, -.8, -.6, -.4, -.2, 0, .2, .4, .6., .8\}$ and true mean $\mu \in [0, 1)$. These power surfaces are for the Bartlett taper with bandwidth fraction $b = .5$. Panel (a) corresponds to subsampling with sampling fraction $a = .04$; panel (b) corresponds to subsampling with sampling fraction $a = .12$; panel (c) corresponds to subsampling with sampling fraction $a = .2$. Panel (d) corresponds to using the plug-in estimate of β; panel (e) corresponds to using the plug-in method with $\beta = 0$; panel (f) corresponds to using the true (unknown) value of β.

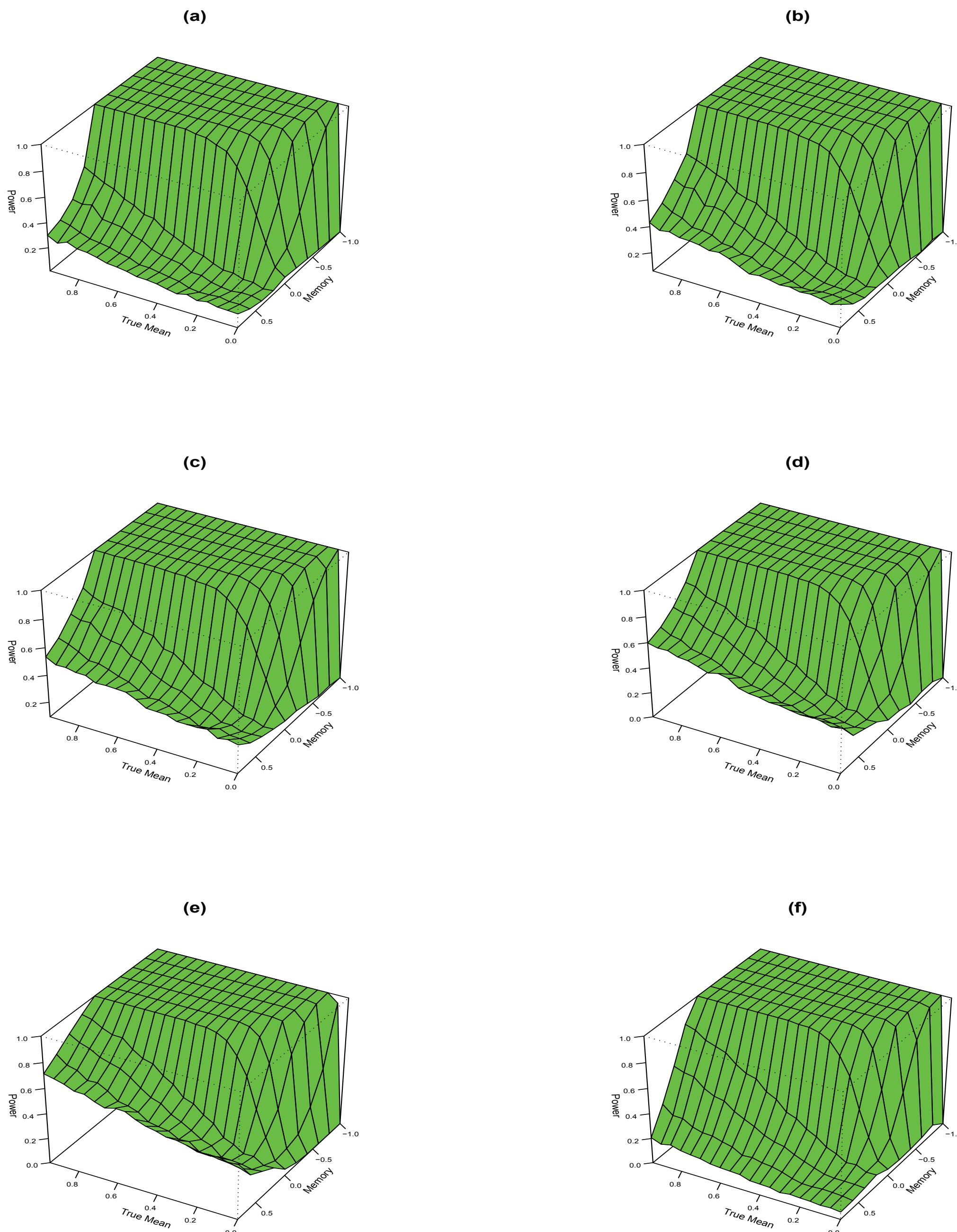

Figure 3: Power surfaces by Memory parameter $\beta \in \{-1, -.8, -.6, -.4, -.2, 0, .2, .4, .6., .8\}$ and true mean $\mu \in [0, 1)$. These power surfaces are for the Bartlett taper with bandwidth fraction $b = 1$. Panel (a) corresponds to subsampling with sampling fraction $a = .04$; panel (b) corresponds to subsampling with sampling fraction $a = .12$; panel (c) corresponds to subsampling with sampling fraction $a = .2$. Panel (d) corresponds to using the plug-in estimate of β; panel (e) corresponds to using the plug-in method with $\beta = 0$; panel (f) corresponds to using the true (unknown) value of β.

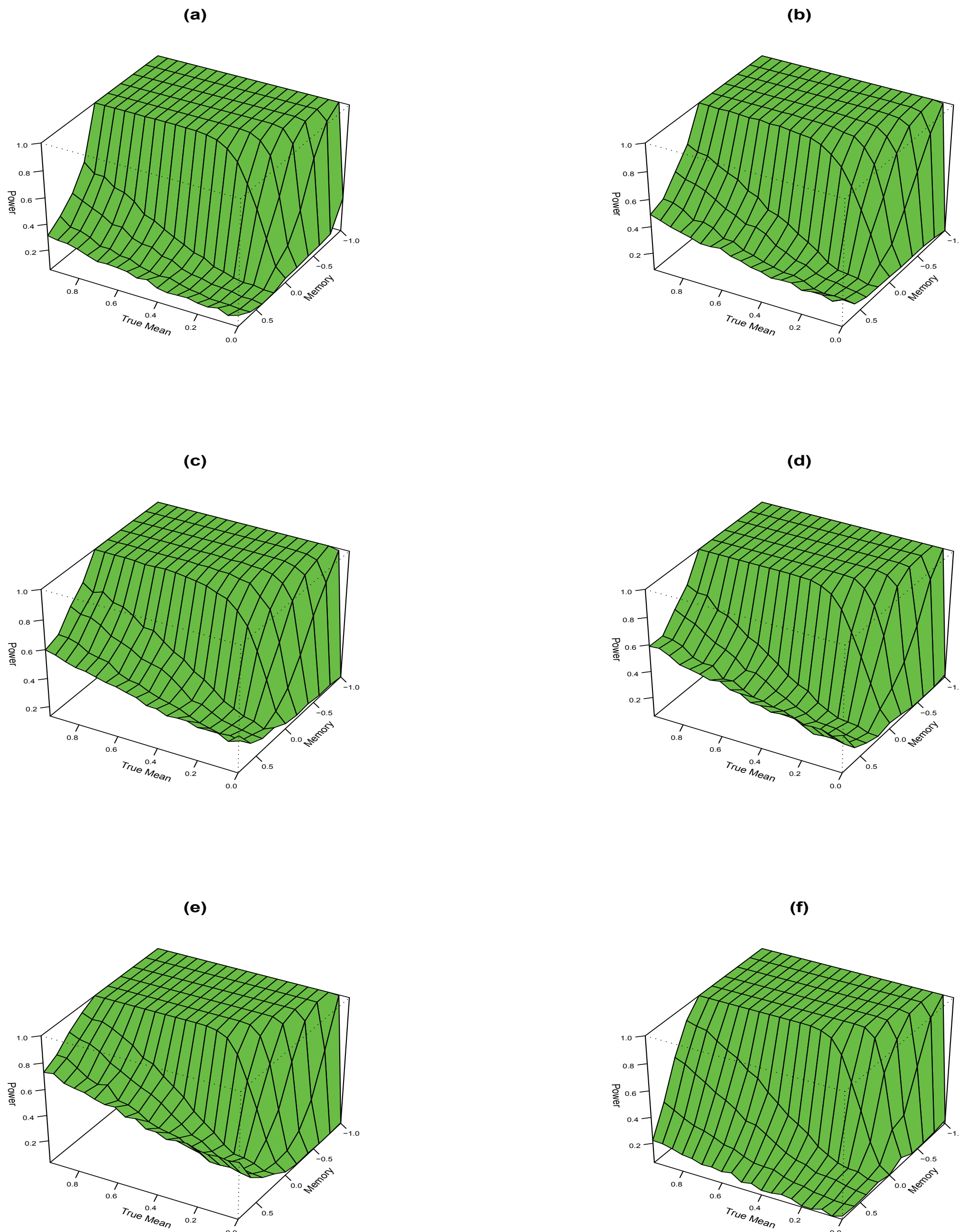

Figure 4: Power surfaces by Memory parameter $\beta \in \{-1, -.8, -.6, -.4, -.2, 0, .2, .4, .6., .8\}$ and true mean $\mu \in [0, 1)$. These power surfaces are for the Trapezoidal (.25) taper with bandwidth fraction $b = .2$. Panel (a) corresponds to subsampling with sampling fraction $a = .04$; panel (b) corresponds to subsampling with sampling fraction $a = .12$; panel (c) corresponds to subsampling with sampling fraction $a = .2$. Panel (d) corresponds to using the plug-in estimate of β; panel (e) corresponds to using the plug-in method with $\beta = 0$; panel (f) corresponds to using the true (unknown) value of β.

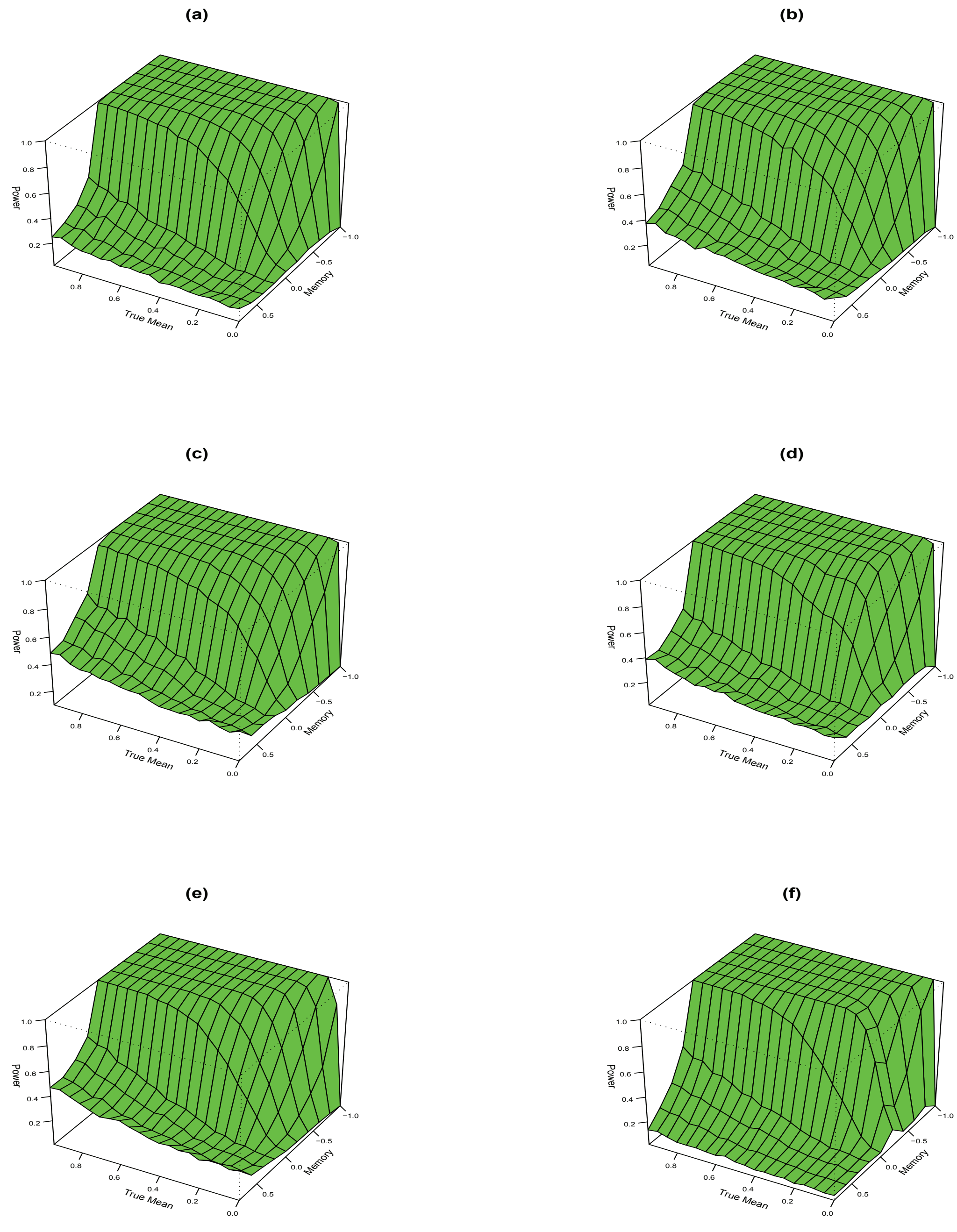

Figure 5: Power surfaces by Memory parameter $\beta \in \{-1, -.8, -.6, -.4, -.2, 0, .2, .4, .6., .8\}$ and true mean $\mu \in [0, 1)$. These power surfaces are for the Trapezoidal (.25) taper with bandwidth fraction $b = .2$. Panel (a) corresponds to subsampling with sampling fraction $a = .04$; panel (b) corresponds to subsampling with sampling fraction $a = .12$; panel (c) corresponds to subsampling with sampling fraction $a = .2$. Panel (d) corresponds to using the plug-in estimate of β; panel (e) corresponds to using the plug-in method with $\beta = 0$; panel (f) corresponds to using the true (unknown) value of β.

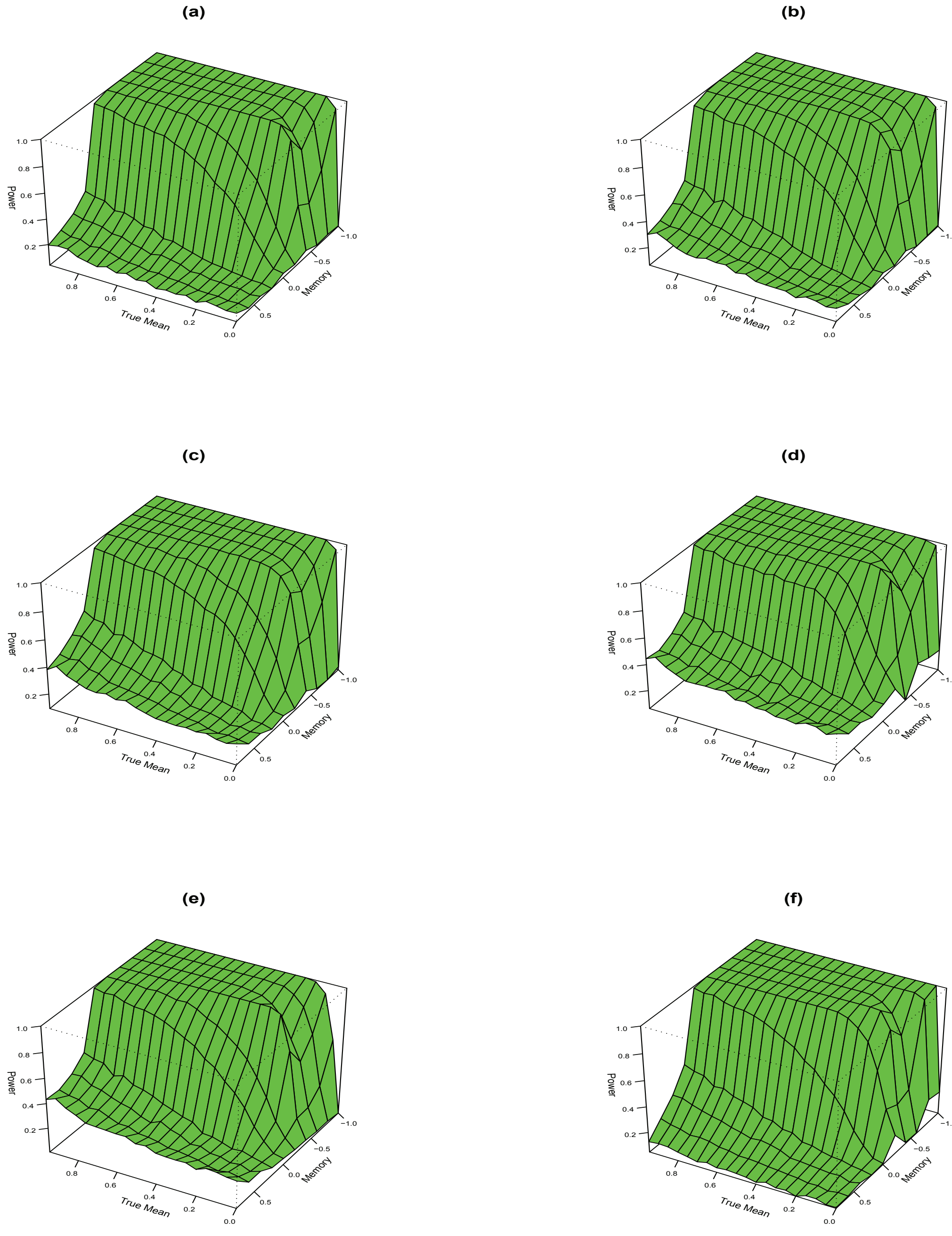

Figure 6: Power surfaces by Memory parameter $\beta \in \{-1, -.8, -.6, -.4, -.2, 0, .2, .4, .6., .8\}$ and true mean $\mu \in [0, 1)$. These power surfaces are for the Trapezoidal (.25) taper with bandwidth fraction $b = 1$. Panel (a) corresponds to subsampling with sampling fraction $a = .04$; panel (b) corresponds to subsampling with sampling fraction $a = .12$; panel (c) corresponds to subsampling with sampling fraction $a = .2$. Panel (d) corresponds to using the plug-in estimate of β; panel (e) corresponds to using the plug-in method with $\beta = 0$; panel (f) corresponds to using the true (unknown) value of β.

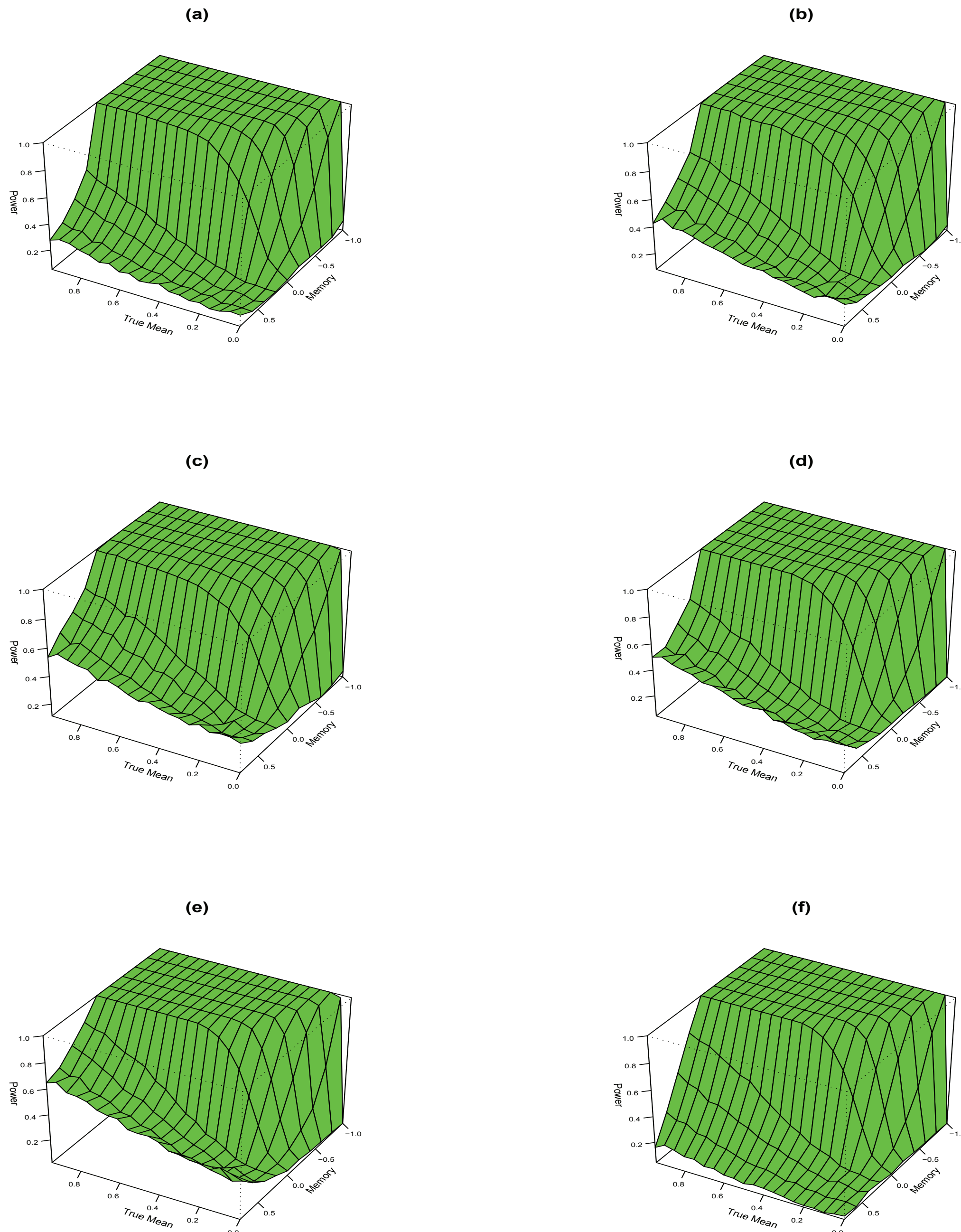

Figure 7: Power surfaces by Memory parameter $\beta \in \{-1, -.8, -.6, -.4, -.2, 0, .2, .4, .6., .8\}$ and true mean $\mu \in [0, 1)$. These power surfaces are for the Trapezoidal (.5) taper with bandwidth fraction $b = .2$. Panel (a) corresponds to subsampling with sampling fraction $a = .04$; panel (b) corresponds to subsampling with sampling fraction $a = .12$; panel (c) corresponds to subsampling with sampling fraction $a = .2$. Panel (d) corresponds to using the plug-in estimate of β; panel (e) corresponds to using the plug-in method with $\beta = 0$; panel (f) corresponds to using the true (unknown) value of β.

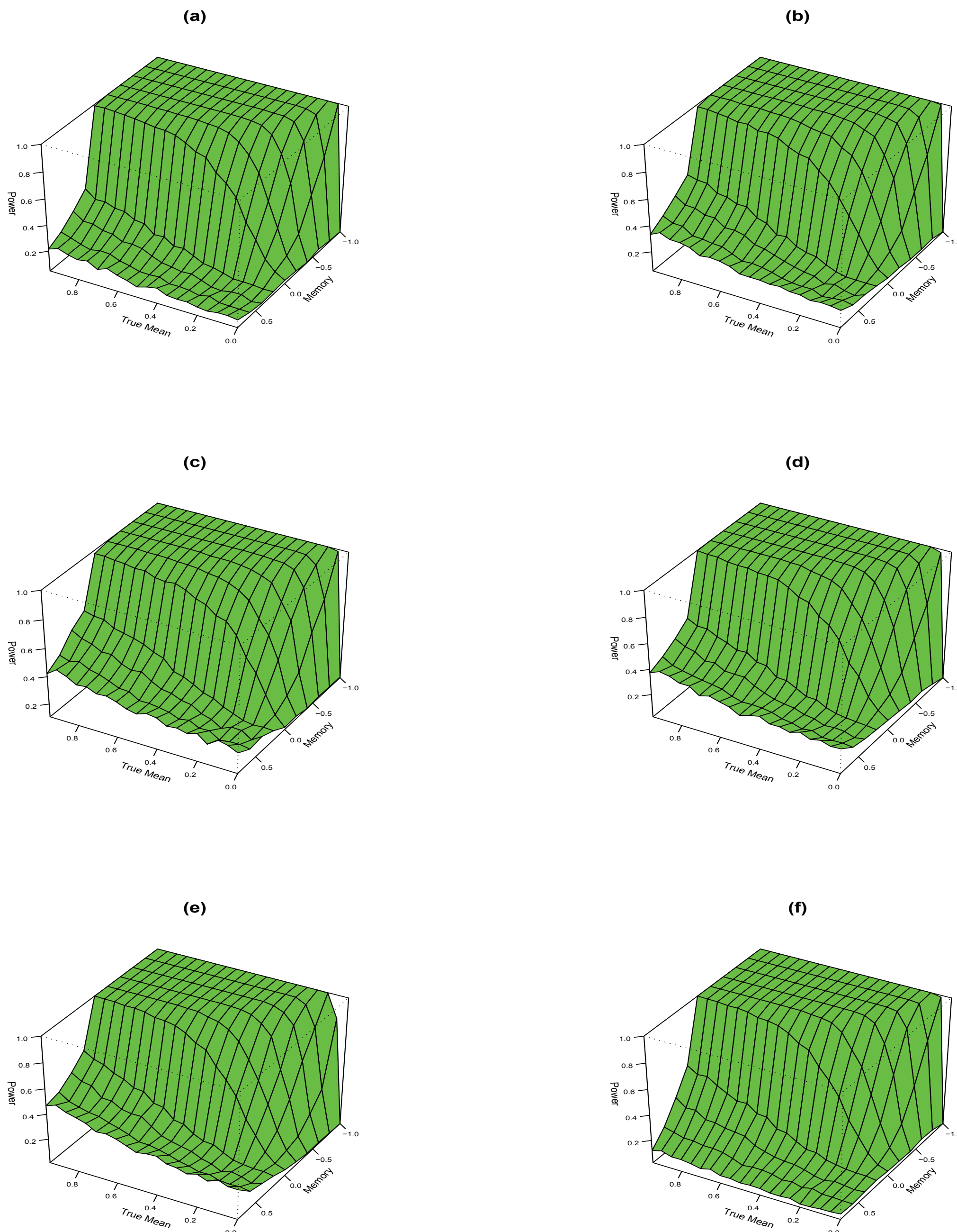

Figure 8: Power surfaces by Memory parameter $\beta \in \{-1, -.8, -.6, -.4, -.2, 0, .2, .4, .6., .8\}$ and true mean $\mu \in [0, 1)$. These power surfaces are for the Trapezoidal (.5) taper with bandwidth fraction $b = .5$. Panel (a) corresponds to subsampling with sampling fraction $a = .04$; panel (b) corresponds to subsampling with sampling fraction $a = .12$; panel (c) corresponds to subsampling with sampling fraction $a = .2$. Panel (d) corresponds to using the plug-in estimate of β; panel (e) corresponds to using the plug-in method with $\beta = 0$; panel (f) corresponds to using the true (unknown) value of β.

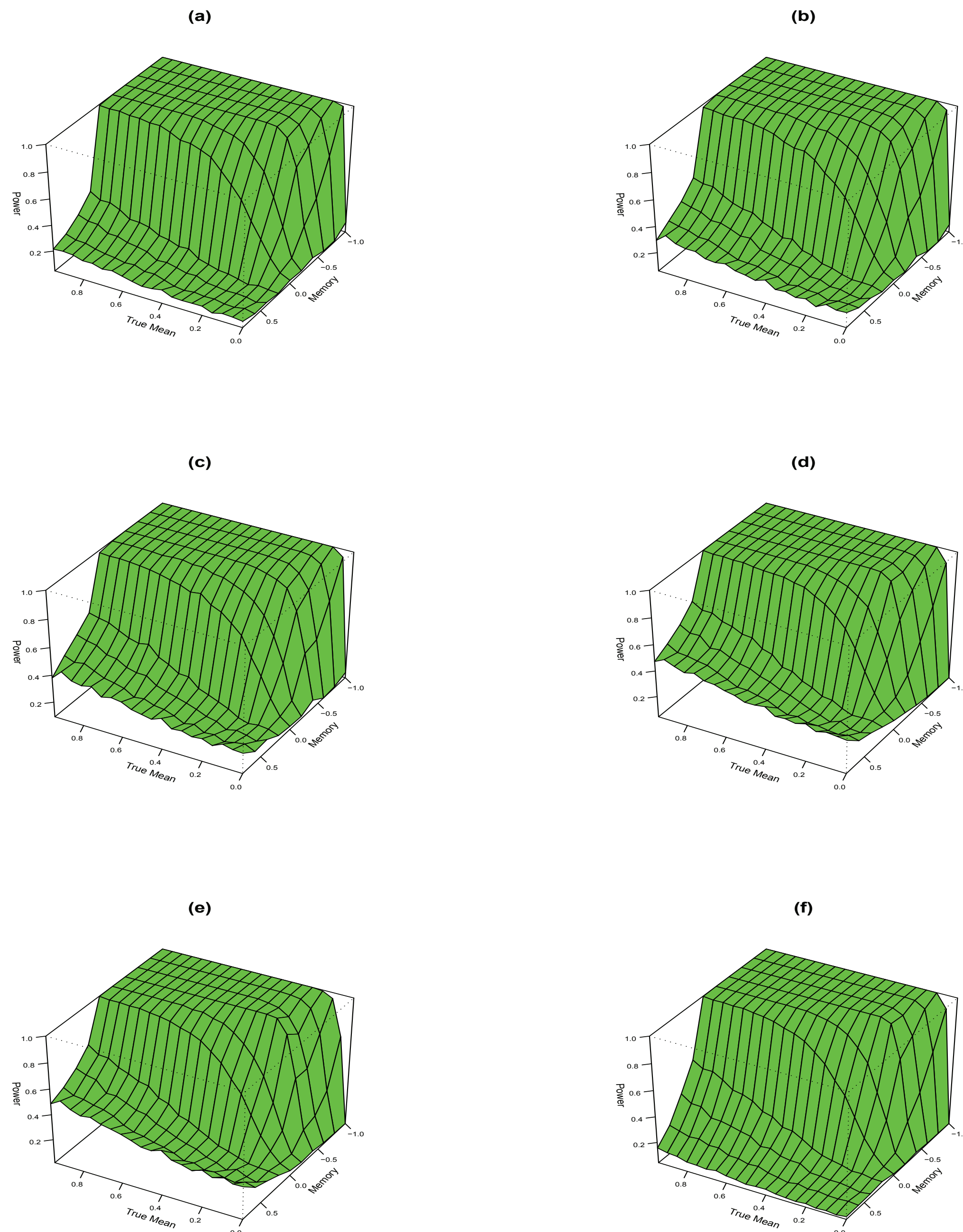

Figure 9: Power surfaces by Memory parameter $\beta \in \{-1, -.8, -.6, -.4, -.2, 0, .2, .4, .6., .8\}$ and true mean $\mu \in [0, 1)$. These power surfaces are for the Trapezoidal (.5) taper with bandwidth fraction $b = .5$. Panel (a) corresponds to subsampling with sampling fraction $a = .04$; panel (b) corresponds to subsampling with sampling fraction $a = .12$; panel (c) corresponds to subsampling with sampling fraction $a = .2$. Panel (d) corresponds to using the plug-in estimate of β; panel (e) corresponds to using the plug-in method with $\beta = 0$; panel (f) corresponds to using the true (unknown) value of β.

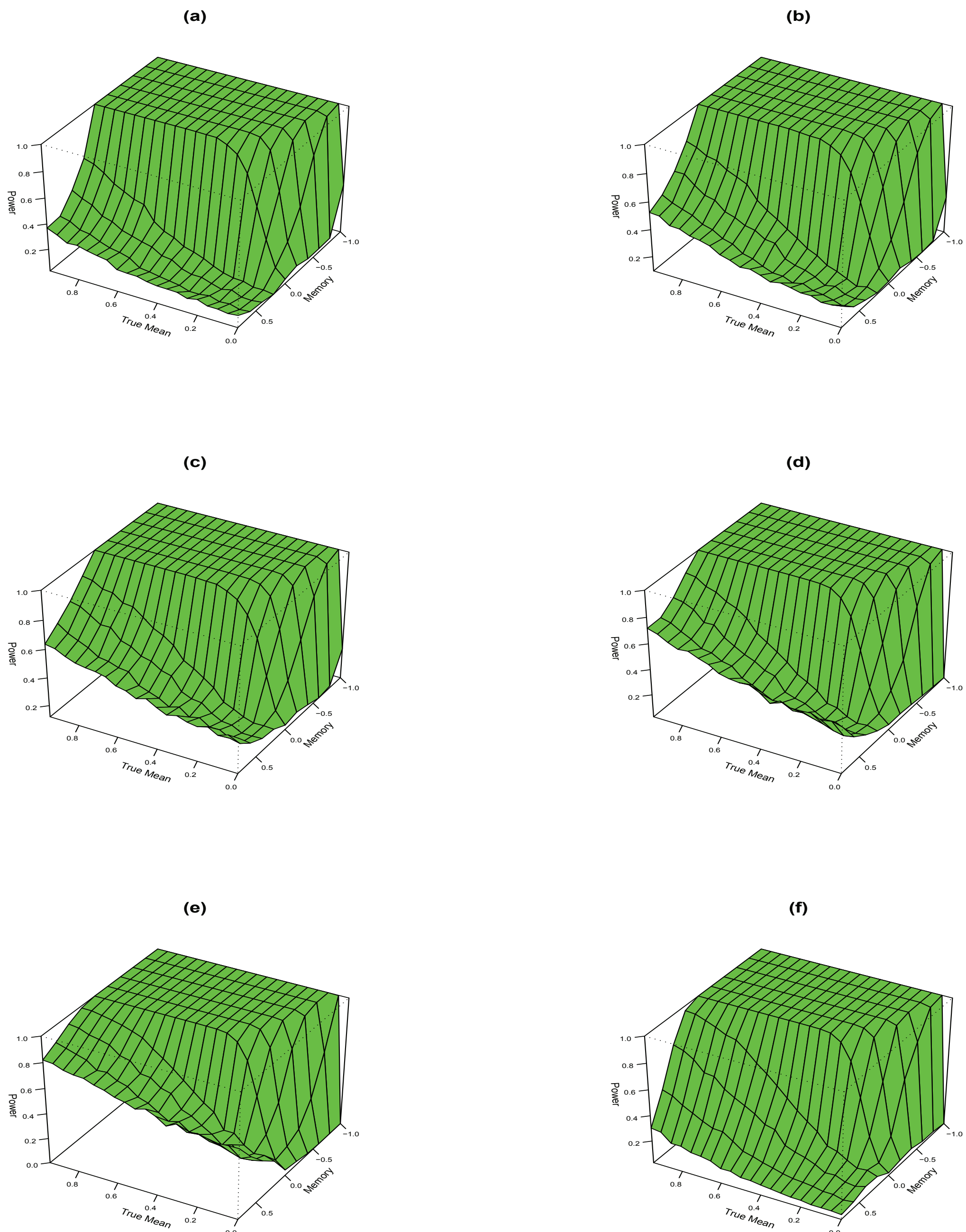

Figure 10: Power surfaces by Memory parameter $\beta \in \{-1, -.8, -.6, -.4, -.2, 0, .2, .4, .6., .8\}$ and true mean $\mu \in [0, 1)$. These power surfaces are for the Parzen taper with bandwidth fraction $b = .2$. Panel (a) corresponds to subsampling with sampling fraction $a = .04$; panel (b) corresponds to subsampling with sampling fraction $a = .12$; panel (c) corresponds to subsampling with sampling fraction $a = .2$. Panel (d) corresponds to using the plug-in estimate of β; panel (e) corresponds to using the plug-in method with $\beta = 0$; panel (f) corresponds to using the true (unknown) value of β.

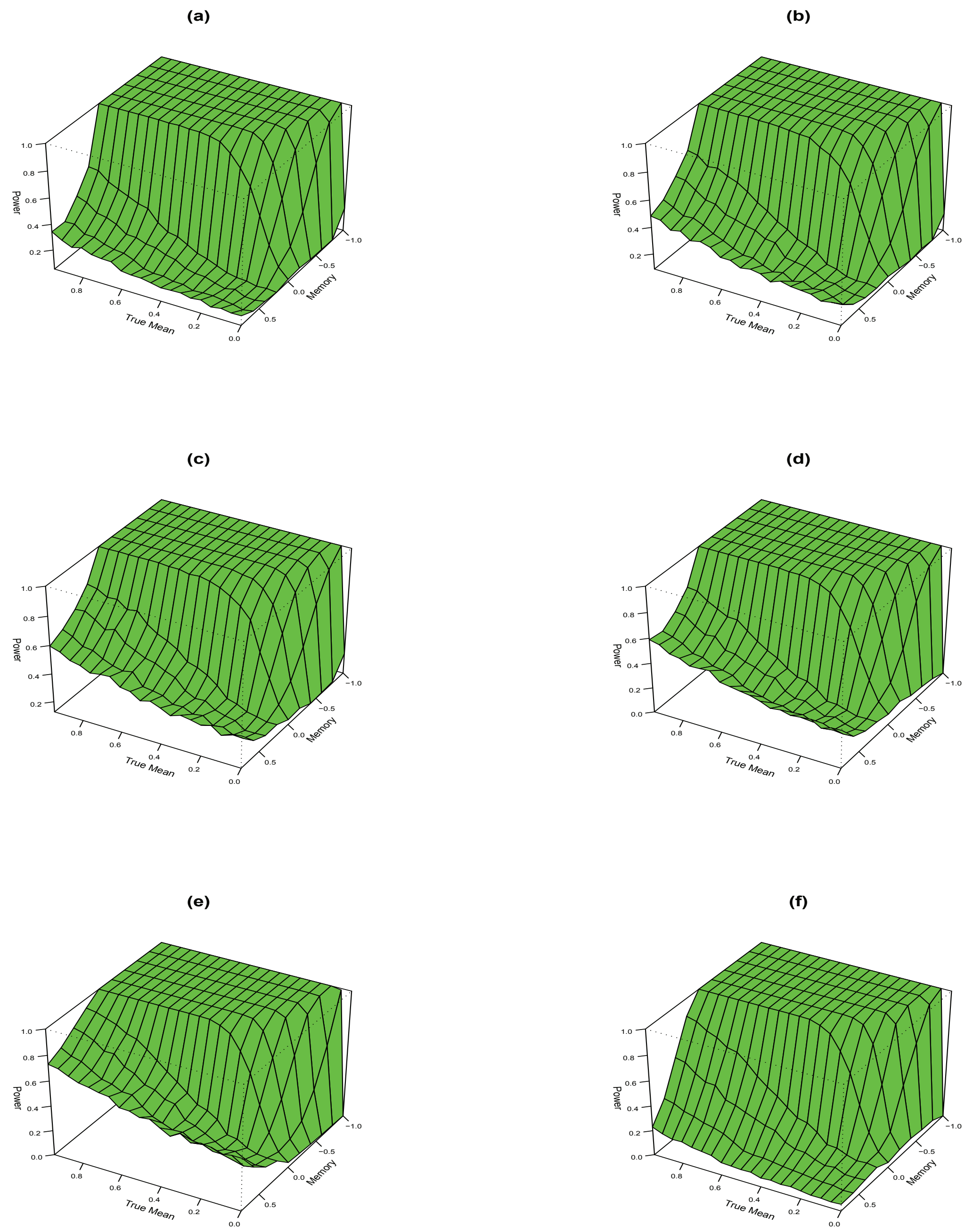

Figure 11: Power surfaces by Memory parameter $\beta \in \{-1, -.8, -.6, -.4, -.2, 0, .2, .4, .6., .8\}$ and true mean $\mu \in [0, 1)$. These power surfaces are for the Parzen taper with bandwidth fraction $b = .5$. Panel (a) corresponds to subsampling with sampling fraction $a = .04$; panel (b) corresponds to subsampling with sampling fraction $a = .12$; panel (c) corresponds to subsampling with sampling fraction $a = .2$. Panel (d) corresponds to using the plug-in estimate of β; panel (e) corresponds to using the plug-in method with $\beta = 0$; panel (f) corresponds to using the true (unknown) value of β.

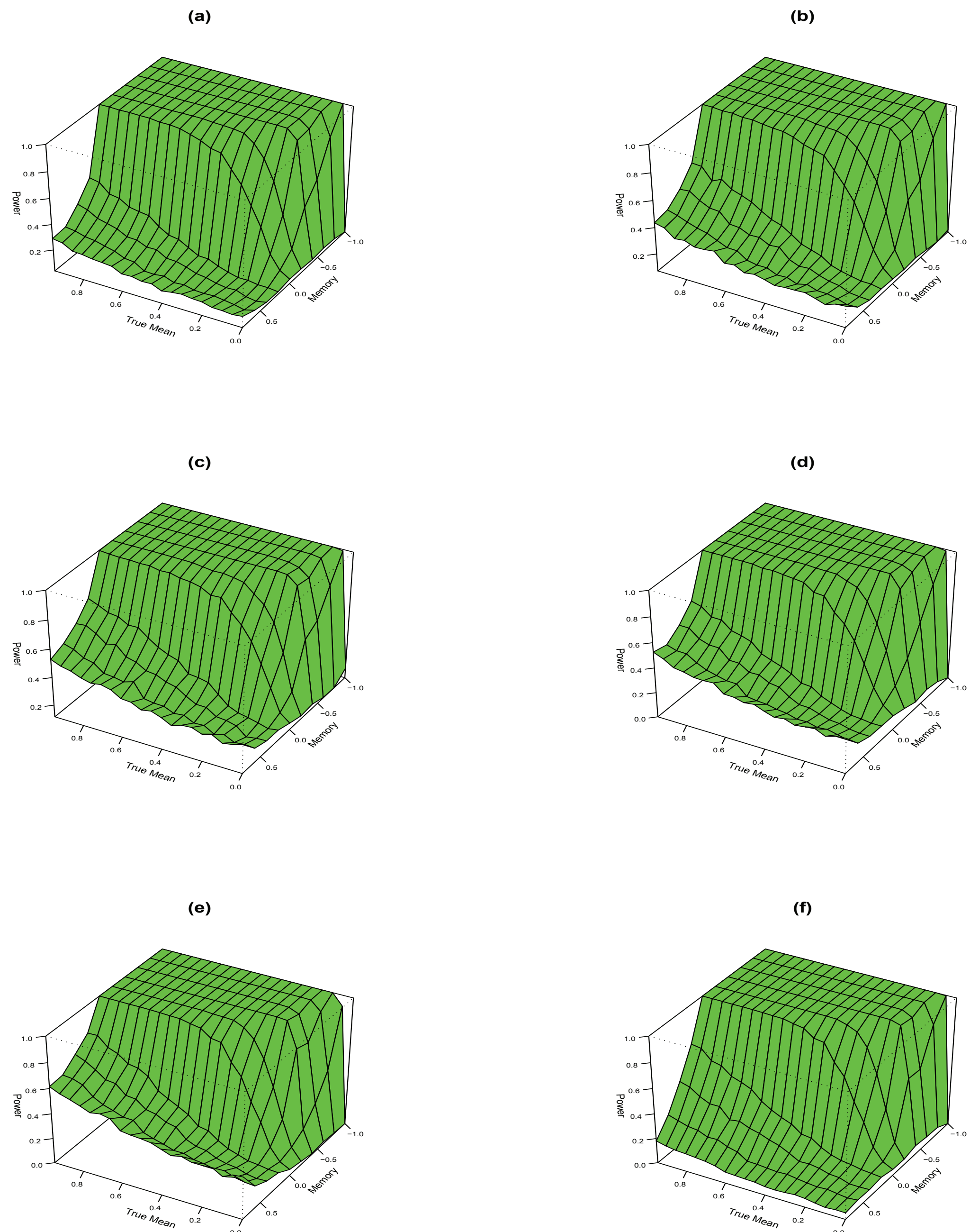

Figure 12: Power surfaces by Memory parameter $\beta \in \{-1, -.8, -.6, -.4, -.2, 0, .2, .4, .6., .8\}$ and true mean $\mu \in [0, 1)$. These power surfaces are for the Parzen taper with bandwidth fraction $b = 1$. Panel (a) corresponds to subsampling with sampling fraction $a = .04$; panel (b) corresponds to subsampling with sampling fraction $a = .12$; panel (c) corresponds to subsampling with sampling fraction $a = .2$. Panel (d) corresponds to using the plug-in estimate of β; panel (e) corresponds to using the plug-in method with $\beta = 0$; panel (f) corresponds to using the true (unknown) value of β.

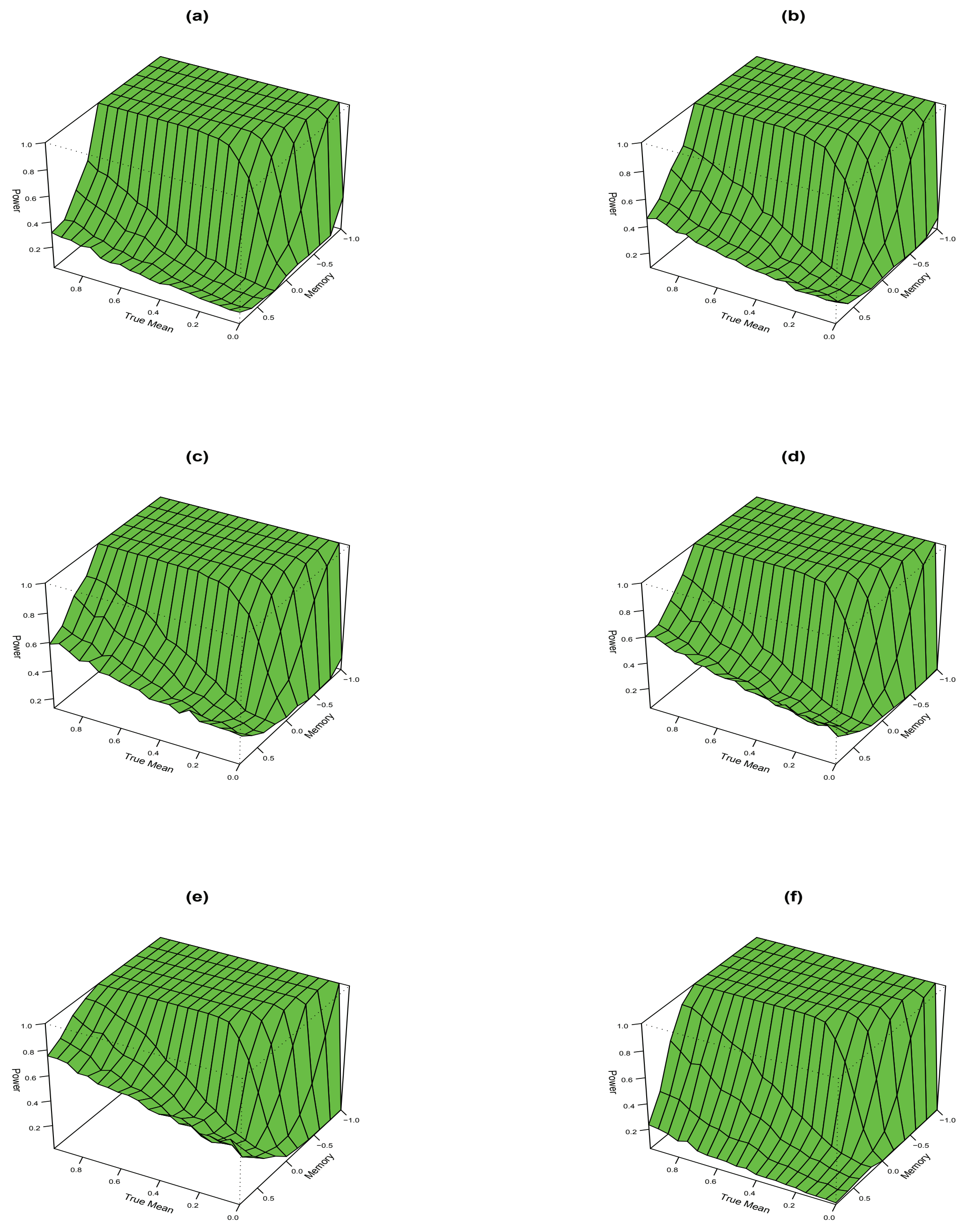

Figure 13: Power surfaces by Memory parameter $\beta \in \{-1, -.8, -.6, -.4, -.2, 0, .2, .4, .6., .8\}$ and true mean $\mu \in [0, 1)$. These power surfaces are for the Daniell taper with bandwidth fraction $b = .2$. Panel (a) corresponds to subsampling with sampling fraction $a = .04$; panel (b) corresponds to subsampling with sampling fraction $a = .12$; panel (c) corresponds to subsampling with sampling fraction $a = .2$. Panel (d) corresponds to using the plug-in estimate of β; panel (e) corresponds to using the plug-in method with $\beta = 0$; panel (f) corresponds to using the true (unknown) value of β.

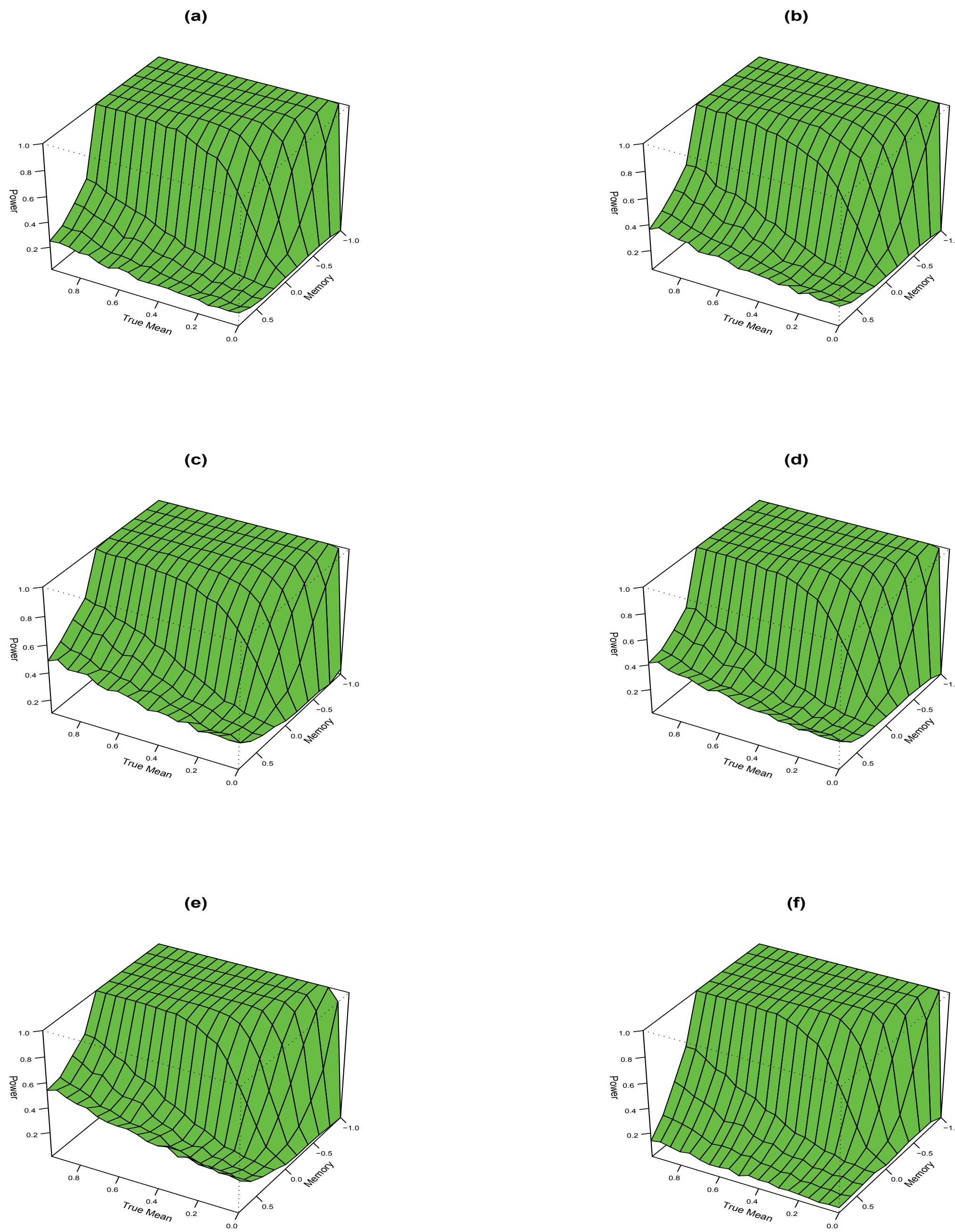

Figure 14: Power surfaces by Memory parameter $\beta \in \{-1, -.8, -.6, -.4, -.2, 0, .2, .4, .6., .8\}$ and true mean $\mu \in [0, 1)$. These power surfaces are for the Daniell taper with bandwidth fraction $b = .5$. Panel (a) corresponds to subsampling with sampling fraction $a = .04$; panel (b) corresponds to subsampling with sampling fraction $a = .12$; panel (c) corresponds to subsampling with sampling fraction $a = .2$. Panel (d) corresponds to using the plug-in estimate of β; panel (e) corresponds to using the plug-in method with $\beta = 0$; panel (f) corresponds to using the true (unknown) value of β.

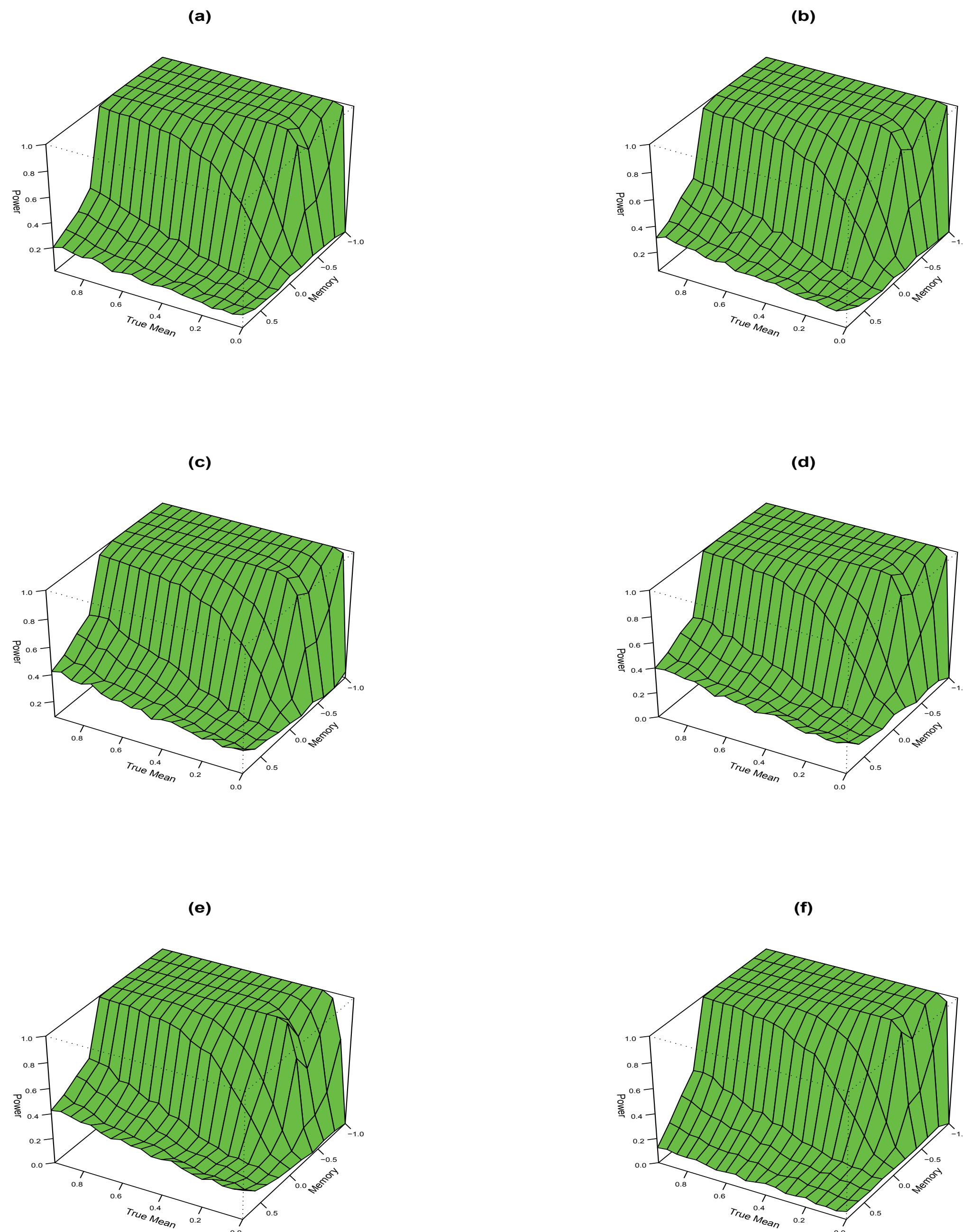

Figure 15: Power surfaces by Memory parameter $\beta \in \{-1, -.8, -.6, -.4, -.2, 0, .2, .4, .6., .8\}$ and true mean $\mu \in [0, 1)$. These power surfaces are for the Daniell taper with bandwidth fraction $b = 1$. Panel (a) corresponds to subsampling with sampling fraction $a = .04$; panel (b) corresponds to subsampling with sampling fraction $a = .12$; panel (c) corresponds to subsampling with sampling fraction $a = .2$. Panel (d) corresponds to using the plug-in estimate of β; panel (e) corresponds to using the plug-in method with $\beta = 0$; panel (f) corresponds to using the true (unknown) value of β.