Memory Parameter Estimation in the Presence of Level Shifts and Deterministic Trends*

Adam McCloskey† **Pierre Perron**‡

Boston University **Boston University**

January, 2010

Abstract

We propose estimators of the memory parameter of a time series that are robust to a wide variety of random level shift processes, deterministic level shifts and deterministic time trends. The estimators are simple trimmed versions of the popular log-periodogram regression estimator that employ certain sample size-dependent, and in some cases, data-dependent trimmings which discard low-frequency components. Regardless of whether the underlying long/short-memory process is contaminated by level shifts or deterministic trends, our estimators are shown to be consistent and asymptotically normal with the same limiting variance as the standard log-periodogram estimator. An extensive simulation study shows that our estimators perform their intended purpose quite well, substantially decreasing both finite sample bias and root mean-squared error in the presence of these contaminating components. Furthermore, we assess the tradeoffs involved with their use when such components are not present but the underlying process exhibits strong short-memory dynamics or is contaminated by noise. To balance the potential finite sample biases involved in estimating the memory parameter, we recommend a particular version of our estimators that performs well in a wide variety of circumstances. Finally, we apply our estimators to stock market volatility and hydrological data to find that many of the time series typically thought to be long-memory processes actually appear to be short-memory processes contaminated by level shifts or deterministic trends.

JEL Classification Numbers: C22, C13, C14

Keywords: long-memory processes, semiparametric estimators, level shifts, structural change, deterministic trends

*The authors are grateful to Shinsuke Ikeda for kindly sharing S&P 500 futures realized volatility data.

†Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (mcclosk@bu.edu, http://people.bu.edu/mcclosk/).

‡Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (perron@bu.edu, http://people.bu.edu/perron/).

1 Introduction

There has long been interest in time series that are stationary yet exhibit persistence beyond that of $I(0)$ variates, the so-called "long-memory" processes. They have been studied as far back as the 1950's with the seminal contribution of Hurst (1951) in the context of hydrology. Long-memory processes are usually characterized in the time domain by an autocorrelation function that is not absolutely summable and decays hyperbolically at long lags. In the frequency domain, it is typically characterized by a spectral density function that is proportional to λ^{-2d} as λ approaches zero, where d is known as the "memory parameter" of the process. Such a process is stationary when $d \in (-1/2, 1/2)$, a specification nesting the short-memory processes ($d = 0$). Independently, Granger and Joyeux (1980) and Hosking (1981) introduced the fractionally integrated ARFIMA(p, d, q) process, a long-memory generalization of the $I(0)$ ARMA(p, q) process. Parametric estimates of d, requiring the specification of the entire spectral density function, have been proposed by Fox and Taqqu (1986) and Dahlhaus (1989), among others. Semiparametric estimates of the memory parameter have grown popular as they do not require specification of the "short-memory" parameters of a process. The most widely used examples of semiparametric estimators are the log-periodogram (LP) estimator of Geweke and Porter-Hudak (1983) and the local Whittle estimator proposed by Künsch (1987).

The fact that the presence of level shifts or deterministic time trends affects the persistence properties of time series has long been recognized by statisticians and econometricians. Perron (1989) showed that the presence of one-time shifts in a time trend will often induce spurious non-rejection of the unit root hypothesis for detrended data. Bhattacharya et al. (1983) demonstrated similar findings with regard to deterministic trends. More recently, researchers have shown that short-memory time series contaminated by level shifts or certain deterministic trends display many of the same properties of long-memory time series, inducing "spurious long-memory" effects. For example, such a process will exhibit hyperbolically decaying autocorrelations as well as a pole at the null frequency of its spectral density function. Among others, Diebold and Inoue (2001), Granger and Hyung (2004), Mikosch and Stărică (2004) and Perron and Qu (2010) provide theoretical reasons for and simulation evidence of this phenomenon. A short-memory time series contaminated by level shifts or a deterministic trend will thus frequently cause spurious rejection of a short-memory null hypothesis and bias memory parameter estimates upward. Overestimation of the memory parameter has important practical implications in economics and finance. For example, Taylor (2000) has shown that the assumed memory parameter significantly affects implied volatilities while Ohanissian et al. (2004) have shown its important impact on the pricing of call options.

In recent years, a handful of papers aimed at distinguishing true from spurious long-memory has emerged. Tests in both the time and frequency domains have been proposed by Dolado et al. (2005), Shimotsu (2006), Ohanissian et al. (2008), Perron and Qu (2010) and Qu (2008) with varying degrees of success. Many of these authors have argued that the long-memory properties of many economic time series are indeed spurious. However, scant attention has been paid to *estimation* of the memory parameter in the presence of contaminating elements. This may be partly due to the existing focus on two specific alternative processes: short-memory contaminated by level shifts or deterministic trends vs. pure long-memory. Though Granger and Hyung (2004) suggest a forecasting procedure, authors have not adequately explored the implications of *long-memory* processes contaminated by these elements, especially level shifts.[1] As we will show later, level shifts and deterministic trends induce an upward bias in memory parameter estimates whether the contaminated process exhibits short *or* long-memory. Tests focusing on the specific alternative of contaminated short-memory may reject if the underlying process is contaminated long-memory, another reason estimation is important. In this paper, we propose very simple estimators of the memory parameter that are robust to the presence of the contaminating elements that cause spurious long-memory. They are simply "trimmed" versions of the popular LP estimator.

Smith (2005) has also attempted to address the issue of memory parameter estimation in the presence of level shifts. He derived a bias correction for the LP estimator from a stationary, mean-reverting mean-plus-noise model. This bias correction is based upon the assumption of a short-memory process contaminated by a stationary random level shift (RLS) process, meaning that it is not necessarily valid when the process is contaminated long-memory. The standard LP estimator is already consistent under the data generating processes (DGPs) he considers (Hurvich et al., 1998) although its finite sample performance is often inferior to that of his estimator. In contrast, we provide estimators that are not only consistent under all of the DGPs Smith (2005) considers but also under many others, including those contaminated by non-stationary RLS processes and smooth or monotonic trends. In terms of RLS processes, those we consider are arguably more practically relevant than those considered by Smith (2005) as they imply periodograms that diverge in expectation

[1] Some authors have addressed memory parameter estimation in the presence of certain types of deterministic trends. See Robinson (1997) and Hurvich et al. (2005) for examples.

at the zero frequency rather than flattening out and converging to some constant. These non-stationary RLS processes have received considerable attention in the econometrics literature (e.g., see Chen and Tiao, 1990; Diebold and Inoue, 2001; Granger and Hyung, 2004 and Perron and Qu, 2010).

It turns out that our estimators have many desirable asymptotic and finite sample properties that we explore later. The estimators employ the usual bandwidth parameter to determine the highest periodogram frequency used in estimation as well as a trimming parameter to determine the lowest periodogram frequency used. They are consistent and asymptotically normal under mild conditions on the contaminating processes, the spectral density function of the contaminated process and the user-chosen trimming and bandwidth parameters. The limiting variance of our trimmed estimators is the same as that for the standard LP estimator, implying no asymptotic efficiency loss. In finite samples, the estimators significantly reduce the upward bias caused by level shift or deterministic components, often nearly eliminating it entirely. In addition, our estimators dominate the standard LP estimator in terms of mean-squared error for a wide variety of DGPs when contaminating elements are present. The finite sample properties of the estimators depend crucially upon their trimming and bandwidth parameters much like the standard LP estimator depends upon its bandwidth parameter. Through Monte Carlo simulation, we explore the finite sample properties of the estimators for different combinations of these parameters, providing practical suggestions for choosing them. Moreover, we provide the asymptotic means squared error (MSE) minimizing choice of the bandwidth parameter by extending the results of Hurvich et al. (1998) (HDB, henceforth) to the DGPs in question.

After the work of this paper was completed, we became aware of related research by Iacone (2010), who also uses trimming in the context of the local Whittle estimator. The class of processes he considers are, however, much more restrictive and less practically relevant. As stated in Remark 1, we can use our theoretical results to show the consistency and asymptotic normality of his trimmed local Whittle estimator for a much larger class of processes. Furthermore, since for some reason his simulations use a trimmed LP estimator (rather than local Whittle), our results also provide the required theoretical grounding to substantiate his Monte Carlo experiments.

The remainder of this paper is composed as follows. Section 2 introduces our estimators. Section 3 describes the DGPs we consider, imposing specific assumptions, and then details a crucial asymptotic property of the periodogram of these processes. Section 4 explores the asymptotic properties of our trimmed LP estimators. Section 5 provides an extensive

Monte Carlo study of the finite sample properties of our estimators, in comparison with the standard LP estimator, under a wide variety of DGPs. In this section, we advocate the use of a particular version of our estimators that seems to best balance considerations arising from potential finite sample biases when the empiricist has little prior information on the DGP at hand. We also show through simulation evidence that the assumption of Gaussianity on the contaminated process imposed in Section 3 can likely be relaxed. Section 6 focuses on empirical application of our estimators to stock market volatility data and the classic Nile River level series. We find that many of these series which have been previously typified as long-memory processes appear actually to be contaminated short-memory processes as our robust estimators of their memory parameters are very near zero. Section 7 is composed of concluding remarks while the proofs of our theoretical results are given in a mathematical appendix.

2 A Robust Memory Parameter Estimator

Log periodogram estimators appear to be the most popular memory parameter estimators among empiricists due to their simplicity, intuitiveness and ease of use. As mentioned, our estimators are "trimmed" versions of the standard LP estimator that employ specific sample size-dependent, and in some cases, data-dependent trimmings that discard low frequency components. The LP estimator is based upon the spectral characterization of a long-memory process which implies the following relation:

$$\log f(\lambda) \approx c - 2d \log \lambda \qquad \text{as } \lambda \to 0_+,$$

where f is the spectral density function of the process. Now, letting $w_x(\lambda_j)$ denote the discrete Fourier transform of the time series $\{x_t\}_{t=1}^T$ evaluated at the Fourier frequency $\lambda_j = 2\pi j/T$, the periodogram of the time series at λ_j is defined as $I_x(\lambda_j) \equiv |w_x(\lambda_j)|^2 = w_x(\lambda_j)w_x(\lambda_j)^*$, where c^* denotes the complex conjugate of any complex number c. I_x can be regarded as a noisy approximation to f. Hence, the LP estimator replaces f in the above relation by I_x, evaluating the relation at Fourier frequencies local to zero to yield the LP regression:

$$\log I_x(\lambda_j) = c + dX_j + e_j, \qquad j = l, \ldots, m,$$

where $X_j = -\log(2 - 2\cos(\lambda_j)) \approx -\log \lambda_j^2$ for $j = l, \ldots, m$. So long as $m/T \to 0$, this approximating relation holds asymptotically. Thus, using the shorthand notation $I_j = I_x(\lambda_j)$, we obtain the LP estimator:

$$\hat{d} = \frac{-0.5 \sum_{j=l}^{m} \left(Y_j - \bar{Y}\right) \log I_j}{\sum_{j=l}^{m} \left(Y_j - \bar{Y}\right)^2},$$

where $Y_j = \log |1 - \exp(-i\lambda_j)|$ and $\bar{Y} = (1/(m-l+1)) \sum_{k=l}^{m} Y_k$.

The standard LP estimator uses $l = 1$ while ours trim out some of the lower Fourier frequencies to obtain consistency and asymptotic normality in the presence of level shifts and deterministic trends. To determine the number of frequencies necessary to trim out to obtain good asymptotic properties for $\hat{d}$, we rely on results concerning the order of the periodogram of a long-memory process contaminated by level shifts or deterministic trends. More specifically, assume the DGP of the time series in question $\{x_t\}$ is given by

$$x_t = k + v_t + u_t, \tag{1}$$

where k is a constant, $\{v_t\}$ is a long or short-memory process and $\{u_t\}$ is an RLS process or a deterministic time trend (we will later make clear exactly what is meant here by imposing specific assumptions on these processes). Then using the results of Robinson (1995) (Theorem 2), a simple extension of arguments presented by Perron and Qu (2010) and Qu (2008) provides that

$$I_x(\lambda_j) = I_v(\lambda_j) + I_u(\lambda_j) + 2I_{vu}(\lambda_j), \tag{2}$$

where $I_v(\lambda_j) = O_p(\lambda_j^{-2d})$, $I_u(\lambda_j) = O_p(T^{-1}\lambda_j^{-2})$ and $I_{vu}(\lambda_j) = O_p(T^{-1/2}\lambda_j^{-(1+d)})$. The component u thus dominates the periodogram for frequencies λ_j such that $j = o(T^{(1-2d)/(2-2d)})$ and the short/long-memory component v dominates when $jT^{(2d-1)/(2-2d)} \to \infty$. In other words, although both a level shift (or deterministic trend) process and a long-memory process tend to have poles in their periodograms at the zero frequency as the sample size increases, these poles essentially taper off differently. The pole induced by level shifts is steeper and tapers off more quickly than the pole induced by a long-memory process.

Not only do these features of the periodogram allow one to distinguish between the two processes, but they also allow one to estimate the memory parameter when both processes are present. Intuitively, if one were to run an LP regression using only frequencies λ_j for which $jT^{(2d-1)/(2-2d)} \to \infty$ but nonetheless $j = o(T)$, one could expect a consistent estimate. Therefore, the intuition behind the estimator is to set $l = O(T^{(1-2d)/(2-2d)+\varepsilon})$ for some $\varepsilon > 0$ in order to extract the memory behavior of the v component. Of course this suggestion is infeasible since d is unknown when one is trying to estimate it. However, for $d \in [0, 1/2)$, the stationary and persistent region, $(1-2d)/(2-2d) \in (0, 1/2]$. Hence, one could obtain an estimate of d by only using frequencies with j growing faster than $T^{1/2}$ but still slower than T.[2] We also introduce an adaptive, data-dependent trimming procedure to reduce finite sample variance in Section 4.

3 Processes of Interest and Their Periodograms

We assume that the process of interest $\{x_t\}$ is given by (1). We shall allow the process $\{u_t\}$ to take a variety of forms to encompass RLS processes, deterministic level shift (DLS) processes and deterministic trends by making the following assumption.

Assumption 1. *The process $\{u_t\}$ is generated according to one of the following DGPs.*

(a) Random Level Shifts (RLS)

$$u_t = \sum_{j=1}^{t} \delta_{T,j}, \qquad \delta_{T,t} = \pi_{T,t}\eta_t,$$

where $\eta_t \sim i.i.d.\left(0, \sigma_\eta^2\right)$ with finite moments of all orders and $\pi_{T,t} \sim i.i.d.Bernoulli(p/T, 1)$ for some $p \geq 0$. The components $\pi_{T,t}$, η_t and v_t are mutually independent.

(b) Deterministic Level Shifts (DLS)

$$u_t = \sum_{i=1}^{B} c_i \mathbb{I}\left(T_{i-1} < t \leq T_i\right),$$

where B is a fixed positive integer (the number of breaks plus one), $\mathbb{I}(\cdot)$ is the indicator function, $T_0 = 0$, $T_B = T$, $T_0 < T_1 < \ldots < T_{B-1} < T_B$ and $T_i/T \to \tau_i \in [0,1]$ for all $i = 1, \ldots, B$.

(c) Deterministic Trends (DT)

$$u_t = h(t/T),$$

where $h(\cdot)$ is a deterministic nonconstant function on $[0,1]$ that is either Lipschitz continuous or monotone and bounded.

It is important to note that the Bernoulli probability of Assumption 1(a) is sample size-dependent. If this were not the case, $\{u_t\}$ would be better construed as a random walk process. This specification allows the average number of level shifts to remain constant (and

[2]For $d \in (-1/2, 1/2)$, $(1-2d)/(2-2d) \in (0, 2/3)$ so that considering frequencies with j growing faster than $T^{2/3}$ but slower than T would produce analogous results. Our focus in later sections of this paper is restricted to the region $d \in [0, 1/2)$ as it seems to be of most practical interest.

equal to p) as the sample size grows. Note that p can be zero in Assumption 1 so that the assumption nests the no level shift, no trend case as well. Perron and Qu (2010) considered the asymptotic properties of the periodogram of this type of process contaminating a short-memory process. Mikosch and Stărică (2004) considered the asymptotic properties of the periodogram of the type of process given by Assumption 1(b), with the addition of a short-memory component. Künsch (1986) considered the asymptotic properties of the periodogram of a short-memory process contaminated by a bounded monotone trend (Lemma 2) and Qu (2008) extended his results to the Lipschitz continuous case (Lemma 1).

We now impose an assumption on the component whose memory parameter we are interested in estimating.

Assumption 2. *The spectral density of the time series $\{v_t\}$ is given by*

$$f(\lambda) = |1 - \exp(-i\lambda)|^{-2d} f^*(\lambda),$$

where $d \in (-1/2, 1/2)$ is the memory parameter, $f^*(\cdot)$ is an even, positive, continuous function on $[-\pi, \pi]$ that is bounded above and away from zero. Moreover, $f^{*\prime}(0) = 0$, $|f^{*\prime\prime}(\lambda)| < B_2 < \infty$ and $|f^{*\prime\prime\prime}(\lambda)| < B_3 < \infty$ for all λ in a neighborhood of zero.

This assumption is identical to that imposed by HDB. The assumption on the spectral density is fairly weak and is satisfied by virtually all processes considered in econometrics, including the popular ARFIMA process. The following theorem is the counterpart to Robinson's (1995) Theorem 2 and also relies on his results. It characterizes the asymptotic behavior of the periodogram of the process $\{x_t\}$ given by (1) under the above assumptions. The theorem is a starting point to proving consistency and asymptotic normality of our estimators as well as being interesting in its own right.

Theorem 1. *Suppose Assumptions 1 and 2 hold. Then for any sequences of positive integers $j = j(T)$ and $k = k(T)$ such that $j > k$ and $j/T \to 0$ as $T \to \infty$, the following bounds hold:*

$$(i)\ E\left[\frac{I_x(\lambda_j)}{f^*(0)\lambda_j^{-2d}}\right] = 1 + O\left[\frac{\log j}{j} + \left(\frac{j}{T}\right)^2 + \frac{T^{1-2d}}{j^{2-2d}}\right]$$

$$(ii)\ E\left[\frac{w_x(\lambda_j)^2}{f^*(0)\lambda_j^{-2d}}\right] = O\left[\frac{\log j}{j} + \frac{T^{1-2d}}{j^{2-2d}}\right],$$

$$(iii)\ E\left[\frac{w_x(\lambda_j)\, w_x(\lambda_k)^*}{f^*(0)\lambda_j^{-d}\lambda_k^{-d}}\right] = O\left[\frac{\log j}{k} + \frac{T^{1-2d}}{j^{1-d}k^{1-d}}\right],$$

$$(iv)\ E\left[\frac{w_x(\lambda_j)\, w_x(\lambda_k)}{f^*(0)\lambda_j^{-d}\lambda_k^{-d}}\right] = O\left[\frac{\log j}{k} + \frac{T^{1-2d}}{j^{1-d}k^{1-d}}\right].$$

Theorem 1(i) implies that, for Fourier frequencies within a certain sample size-dependent range, the periodogram is akin to an asymptotically unbiased estimator of the spectral density function. We say "akin to" as it is not exactly asymptotically unbiased since the spectral density function diverges when evaluated at these frequencies but nevertheless, the expectation of the periodogram mimics this divergent behavior. Theorem 1(ii)-(iv) describes the limiting covariance behavior of the discrete Fourier transform, evaluated at Fourier frequencies within a certain range whose upper bound grows slower than the sample size.

Remark 1. *A result embedded in the proof of Theorem 1, i.e., that under Assumption 1* $E[I_u(\lambda_j)] = O(T/j^2)$, *can be used to show that Iacone's (2010) Assumption 2 (or a slight variant thereof) holds. Hence, all of his ensuing theoretical results also hold under the more general class of processes we consider.*

Remark 2. *It should be noted that under Assumption 1(c), the bounds given in Theorem 1 may not be exact and may therefore overstate the asymptotic orders of the quantities in (i)-(iv), depending on the properties of the deterministic trend function* $h(\cdot)$ *(see the proof for details). However, we wish to impose minimal assumptions on* $h(\cdot)$ *so that our results apply to a wide variety of trends. Nevertheless, if the practitioner is able to impose more structure on* $h(\cdot)$*, he may also be able to use a smaller trimming when estimating* d*.*

Remark 3. *For the processes studied in this paper, given by* (1) *and Assumption 1, the expectation of the periodogram can be decomposed as follows for large sample sizes when* λ_j *is local to zero:*

$$E\left[I_x\left(\lambda_j\right)\right] \approx \lambda_j^{-2d} f^*\left(\lambda_j\right) + T^{-1}\lambda_j^{-2} g(\lambda_j) = \lambda_j^{-2d}\widetilde{f}(\lambda_j),$$

where $\widetilde{f}(\lambda_j) \equiv f^*(\lambda_j) + T^{-1}\lambda_j^{2d-2} g(\lambda_j)$ *and* $g(\cdot)$ *is some nonnegative even function that is bounded at zero. First order Taylor expansions, similar to those leading to the estimator provided by Sun and Phillips (2003), can be used to show*

$$\log \widetilde{f}(\lambda_j) = \log f^*(0) + T^{-1}\lambda_j^{2d-2}\frac{g(0)}{f^*(0)} + o(1)$$

if $jT^{(2d-1)/(2-2d)} \to \infty$ *while* $j = o(T)$*. This motivates the nonlinear pseudo-regression:*

$$\log I_x(\lambda_j) = \alpha - 2d \log \lambda_j + T^{-1}\lambda_j^{2d-2}\gamma + error_j$$

for j *such that* $j = O(T^{(1-2d)/(2-2d)+\varepsilon})$ *for some* $\varepsilon > 0$ *and* $j/T \to 0$*. We have performed some preliminary simulations for this nonlinear pseudo-regression specification using estimators analogous to the adaptive and trimmed ones outlined below to find that they tend* *to underestimate the memory parameter. These (under)estimates still grow in the memory parameter so further examination of this technique could potentially produce a more refined robust estimator.*

Remark 4. *Another potential procedure aimed at reducing the level shift/deterministic trend bias would be to employ the trimmed and adaptive estimators to the tapered periodogram. Tapering is known to decrease biases arising from non-stationary components of a time series in the LP regression (see Velasco, 1999). Initial simulation results show that using our estimators on tapered data (using the cosine bell taper) further reduces bias arising from these components but increases finite sample variance.*

4 Asymptotic Properties of the Robust Estimator

As stated earlier, the lower trimming of our estimators must grow at a certain rate with the sample size. As with all semiparametric estimators of the memory parameter, so too must the bandwidth parameter. The rates at which these two user-chosen parameters must grow also depends on the underlying memory parameter of the process $\{v_t\}$ (though one need not know the true value of d in practice). Assumption 3 makes these rates precise.

Assumption 3. *As $T \to \infty$,*

$$\frac{m \log m}{T} + \frac{l \log^2 m}{m} + \frac{T^{1-2d}}{l^{2-2d}} \to 0.$$

The third term comprises the strongest assumption and is crucial for the trimmed estimators to have good asymptotic properties. Choosing $l = \alpha T^{1/2+\varepsilon}$ for some $\alpha, \varepsilon > 0$ will satisfy this assumption so long as $l \log^2 m/m \to 0$, though efficiency gains can be made by using an adaptive procedure. Finally, we must impose more distributional structure on the processes we consider.

Assumption 4. $\{v_t\}$ *is a Gaussian process.*

We must impose Gaussianity because the existing literature on LP estimators does not cover the non-Gaussian case without necessitating one to "pool" observations across adjacent Fourier frequencies (see Velasco, 2000). The assumption of Gaussianity may appear strong since processes in economics and finance most frequently thought to exhibit long-memory are volatility processes. Nevertheless, simulation evidence presented in Section 5 indicates that this assumption could be relaxed.

We now state results concerning the asymptotic bias and variance of the robust trimmed estimators. The following theorem parallels Theorem 1 of HDB.

Theorem 2. *Under Assumptions 1-4,*

$$(i)\ E\left[\hat{d}-d\right] = \frac{-2\pi^2}{9}\frac{f^{*\prime\prime}(0)}{f^*(0)}\frac{m^2}{T^2} + o\left(\frac{m^2}{T^2}\right) + O\left(\frac{\log^3 m}{m}\right) + O\left(\frac{T^{3/2-3d}\log^2 m}{m^{3-3d}}\right) \tag{3}$$

$$(ii)\ \mathrm{Var}\left(\hat{d}\right) = \frac{\pi^2}{24m} + o\left(\frac{1}{m}\right) + O\left(\frac{T^{3/2-3d}\log^3 m}{m^{3-3d}}\right) \tag{4}$$

$$(iii)\ \mathrm{MSE}\left(\hat{d}\right) = \frac{(4\pi)^4}{81}\left\{\frac{f^{*\prime\prime}(0)}{f^*(0)}\right\}^2\frac{m^4}{T^4} + \frac{\pi^2}{24m} + O\left(\frac{m\log^3 m}{T^2}\right) \tag{5}$$
$$+ O\left(\frac{T^{3/2-3d}\log^3 m}{m^{3-3d}}\right) + o\left(\frac{m^4}{T^4}\right) + o\left(\frac{1}{m}\right).$$

The following corollary is a direct consequence of this theorem.

Corollary. *Under Assumptions 1-4, $\hat{d}$ is a consistent estimator of d.*

Remark 5. *Regarding the rate of convergence of $\hat{d}$ to d under Assumption 1(a), if one is willing to make stronger assumptions on the cumulants of the random variables η_i, then one can obtain faster convergence of the MSE to zero. For example, if one assumed that η_i had zero skewness, the terms $O((T^{3/2-3d}\log^3 m)/(m^{3-3d}))$ could be replaced by $O((T^{2-4d}\log^3 m)/(m^{2-2d}))$, which goes to zero faster under Assumption 3. See Lemma A.3 in the appendix, in conjunction with the proofs of Lemma A.4 and Theorem 2, for details.*

Remark 6. *Neglecting the remainder terms in the mean-squared error* (5) *and minimizing with respect to m yields the same asymptotically optimal choice for m as in HDB:*

$$m^{\mathrm{OPT}} = \left(\frac{27}{128\pi^2}\right)^{1/5}\left\{\frac{f^*(0)}{f^{*\prime\prime}(0)}\right\}^{2/5} T^{4/5}. \tag{6}$$

As noted by HDB, the remainder term $O(m\log^3 m/T^2)$ in (5) is asymptotically negligible as long as $m = KT^\beta$ for $K > 0$ and $2/3 < \beta < 1$. A similar result holds for the remainder term $O(T^{3/2-3d}\log^3 m/m^{3-3d})$, which is asymptotically negligible when $\beta > (11-6d)/(14-6d)$. If we restrict our attention to $d \in [0, 1/2)$, which is likely the most interesting range for d in practice, this merely necessitates that $\beta > 11/14$, the maximum of $(11-6d)/(14-6d)$ within this range. In this case, neither remainder term affects the mean squared error for β in a neighborhood of 4/5 and m^{OPT} is indeed the MSE optimal choice.[3] *Of course in practice, f^* is unknown so the constant in m^{OPT} cannot be found exactly.*

[3]In fact, m^{OPT} is optimal for any $d \in (-1/6, 1/2)$.

With a slightly stronger assumption on the bandwidth parameter m, we can establish asymptotic normality of the estimators.

Theorem 3. *Under Assumptions 1-4, if* $m = o(T^{4/5})$, $m^{1/2}(\hat{d} - d) \xrightarrow{d} N(0, \pi^2/24)$.

Remark 7. *Note that the limiting variance given above is the same as that for the standard LP estimator (see Robinson, 1995 or HDB). Hence, we do not lose asymptotic efficiency by employing a trimming and bandwidth combination that satisfies Assumption 3.*

5 Finite Sample Properties

Before delving into the finite sample properties of our estimators, we first introduce an adaptive procedure that is aimed at decreasing finite sample variance. Focusing on the last term in Assumption 3, suppose that $\hat{d}_0$ is a consistent estimate of d. Then, letting $\hat{d}_1$ denote the trimmed LP estimate of d using the lower trimming $l = \alpha T^{(1-2\hat{d}_0)/(2-2\hat{d}_0)+\varepsilon}$ for some $\alpha > 0$ and some small $\varepsilon > 0$, $\hat{d}_1$ will be consistent and asymptotically normal by Theorems 2 and 3 as long as the remainder of Assumptions 1-4 are satisfied. This adaptive procedure can clearly be repeated. That is, obtain an initial estimate $\hat{d}_0$ using the trimming $l = \alpha T^{1/2+\varepsilon}$. Then, for $i \geq 1$, let $\hat{d}_i$ denote the LP regression estimate obtained using the trimming $l = \alpha T^{(1-2\hat{d}_{i-1})/(2-2\hat{d}_{i-1})+\varepsilon}$. Then $\hat{d}_i$ will retain the asymptotic properties of Theorems 2 and 3 for any finite i. The practitioner can thus choose a convergence criterion that terminates this adaptive procedure to implement it in practice. In the following simulation study, we use the criterion of $|\hat{d}_i - \hat{d}_{i-1}| < 0.01$ (convergence) or $i > 9$ (nonconvergence) to terminate the adaptive procedure.[4] Since we are primarily concerned with the region $[0, 0.5)$ for d, we do not consider trimmings larger than $O(T^{1/2+\varepsilon})$.

We examine the finite sample properties of the trimmed LP estimator using no adaptive procedure, henceforth referred to as the "trimmed" estimator, and the adaptive trimmed estimator, henceforth referred to as the "adaptive" estimator, in comparison to the standard LP estimator. Setting the trimming $l = T^{1/2+\varepsilon}$ for the trimmed estimator, $l = T^{(1-2\hat{d}_{i-1})/(2-2\hat{d}_{i-1})+\varepsilon}$ for the adaptive estimator and the bandwidth $m = T^u$, we look into four different trimming-bandwidth combinations for the trimmed and adaptive estimators ($(\varepsilon, u) = (0.01, 0.7)$, $(0.05, 0.8)$, $(0.1, 0.8)$ and $(0.15, 0.9)$) and four different bandwidths for the standard estimators ($u = 0.5, 0.7, 0.8$ and 0.9). We report the finite sample bias and root

[4] We chose this criterion because, from unreported results, it performs relatively well with very high rates of convergence in larger ($T = 2000$) samples without being too computationally time-consuming.

mean squared error (RMSE) for sample sizes of $T = 500, 1000$ and 2000, though financial time series are often even longer. Each quantity is based on 1000 Monte Carlo replications.

5.1 Comparative Performance in the Presence of Level Shifts or Deterministic Trends

We begin by studying the performance of the different LP estimators under some of the DGPs for which our robust estimators were designed. These are simple Gaussian fractional white noise (ARFIMA$(0, d, 0)$) processes with innovation variance set to unity, contaminated by either RLS's or deterministic time trends. For each contaminating process, we report results for $d = 0, 0.2$ and 0.45. We begin with the RLS DGPs laid out in Assumption 1(a), setting the average number of jumps per sample to $p = 5, 10$ and 20 and the jump distribution $\eta_i \sim i.i.d.N(0, 1)$. The results are recorded in Tables 1-4.

Looking at Tables 1 and 3, begin by noting the substantial upward bias that the level shifts cause in the standard LP estimator. This bias is increasing in the frequency of level shifts p while it is typically decreasing in the sample size and bandwidth used. Nevertheless, the bias remains prevalent for large sample sizes and bandwidths. The fact that the bias is decreasing in the bandwidth can be attributed to the orders given in (2): a larger bandwidth picks up more observations for which the long-memory component dominates. It should be noted that this bias also appears to be decreasing in the true memory parameter d. This can be partially attributed to the orders given in (2), as for a larger d, the long-memory component dominates for more frequencies. However, this is also partially an artifact of scaling. Using $\Gamma(\cdot)$ to denote the gamma function, the variance of a fractional white noise (FWN) process with unit innovation variance is equal to $\Gamma(1 - 2d)/\{\Gamma^2(1 - d)\}$, which is increasing in d. For the quantities we examine, this variance equals 1 when $d = 0$, 1.28 when $d = 0.2$ and 5.89 when $d = 0.45$, quite large differences. Yet for all RLS processes considered in this study, the conditional variance of a RLS is set to unity. Thus the magnitudes of the jumps are relatively smaller when added to processes with higher values of d.

The next feature to note from these tables is that both the trimmed and adaptive estimators remove large portions of this bias. Purely in terms of bias, the trimmed estimator with the largest trimming ($(\varepsilon, u) = (0.15, 0.9)$) performs best for most of the DGPs considered although the adaptive estimator with this trimming performs best when $d = 0$ (see Table 1). This is to be expected again from (2) since the trimmed estimator ignores the most frequencies closest to those for which the RLS process asymptotically dominates the periodogram. By construction, for any fixed (ε, u) combination, the adaptive estimator will "trim out" fewer frequencies than will the trimmed estimator unless $\hat{d} \approx 0$. As expected, the remaining bias is still increasing in p and decreasing in T. Note that for larger sample sizes and $p \leq 10$, both types of our estimators almost entirely eliminate the bias in many cases. The main result that emerges here is the larger the trimming, the less bias due to level shifts will be present. This result may be offset by considerations due to biases arising from short-memory components, as discussed in the next subsection.

In terms of RMSE, the standard estimator is generally dominated by both its trimmed and adaptive counterparts, as can be seen from Tables 2 and 4.[5] This dominance is often quite dramatic, especially for the more frequent level shift cases. When $p = 10$ or 20, the RMSE of our estimators is typically one half to one quarter of its standard counterpart, a major improvement. These cases are hardly extremes, implying samples with an average of 10 or 20 level shifts. In most cases considered, the bias dominates the variance in the RMSE, giving the estimators with the largest trimming ($(\varepsilon, u) = (0.15, 0.9)$) the advantage. In comparison to the numbers for bias, the adaptive estimator essentially "catches up" to its trimmed counterpart, usually surpassing it in performance due to the usual bias-variance tradeoff.[6] In terms of the frequencies it uses in the LP regression, the adaptive estimator lies in between the standard and trimmed ones. This fact makes its bias usually larger (although it remains asymptotically consistent) and its finite sample variance usually smaller than the trimmed one. As d grows larger, the adaptive estimator moves closer to the standard estimator because it trims out fewer frequencies. Contrastingly, it moves closer to the trimmed estimator as d decreases.

In order to work directly with data-driven DGPs, we examine two RLS DGPs that were calibrated to stock market volatility data by Lu and Perron (2010). Using an augmented Kalman filter, Lu and Perron (2010) estimated the parameters of the DGP given by Assumption 1, assuming $d = 0$ and $v_t \sim i.i.d.N\left(0, \sigma_v^2\right)$ for four different stock market volatility series. We examine the DGPs produced by the calibrated parameters of two of these series: the S&P 500 and the NASDAQ. For the S&P 500, Lu and Perron (2010) estimated $\sigma_\eta = 0.75123$, $p = 15.2308$ and $\sigma_v = 0.73995$ while those for the NASDAQ are $\sigma_\eta = 1.43396$, $p = 6.61584$ and $\sigma_v = 0.74255$. The results for these DGPs are presented in the lower portion

[5] This dominance is not as strong and sometimes does not hold when $d = 0.45$ because the RLS component is relatively very small. From unreported results, when we re-scale the variance of η_i to be comparable to the variance of v_t, the standard estimator is RMSE-dominated in all cases considered. Still, using this re-scaling, the reductions in RMSE from using our estimators decrease as d increases but remain quite substantial.

[6] One caveat worth noting is that for the larger value $d = 0.45$, when $p = 20$ or when we increase the variance of η_i to be comparable to that of v_t, the trimmed estimator RMSE dominates the adaptive one.

of Tables 1 and 2. As we can see, there is substantial upward bias in all of the standard LP estimators, bias that our estimators largely removes. Similarly, our estimators clearly dominate the standard one in terms of RMSE, reducing it by up to a factor of four. For these DGPs, the trimmed and adaptive estimators perform somewhat similarly although, in terms of both bias and RMSE, the trimmed one generally performs better.

To conclude this subsection, we examine the estimators' performance in the presence of deterministic trends rather than RLS's. For this exercise, we add two types of trends to the underlying short/long-memory process: a monotonic trend and a seasonal trend. In accord with Assumption 1(c), the monotonic trend is specified as $h(t/T) = 3(t/T + c)^{-0.1}$ for some $c > 0$. We set $c = 0.001$, to correspond closely with that examined by Qu (2008) and Ohanissian et al. (2008). The seasonal trend is specified as $h(t/T) = \sin(3\pi t/T)$. The results for these trends and the three different values $d = 0, 0.2$ and 0.45 are given in Tables 5 and 6. Several interesting results emerge. Apart from those using $(\varepsilon, u) = (0.01, 0.7)$ or for the small sample size $T = 500$, our estimators remove virtually all of the upward bias present in the standard estimator induced by the trends. The trimmed estimator seems to do a marginally better job at bias-reduction. Furthermore, our estimators outperform the standard one in terms of RMSE for $d = 0$. In these cases with deterministic trends, the adaptive estimator performs better than the trimmed one. The adaptive estimator with a lower trimming also tends to outperform the standard one when $d = 0.2$ and performs comparably when $d = 0.45$. Overall, the adaptive estimator RMSE-dominates the other two.

5.2 Comparative Performance without Contaminating Components

It is now time to turn to the cases for which our estimators should not be expected to perform better: those for which there are no contaminating elements present. The purposes of this subsection are to examine what is lost from using our estimators when level shifts or deterministic trends are not an issue and to provide guidance on choosing (ε, u) to appropriately balance the potential for various biases. We first examine the simple case of an uncontaminated FWN process (or the same process we started the previous subsection with, setting $p = 0$). We again assess the estimators' performance for $d = 0, 0.2$ and 0.45. Table 7 displays these results. As should be expected, neither the standard estimator nor the trimmed estimator displays any notable bias. Oddly, the adaptive estimator appears to induce a very small downward bias though the tables show that this bias disappears as the sample grows. In terms of RMSE, the standard estimator is the best, as expected. With the exception of the small sample size and $(\varepsilon, u) = (0.01, 0.7)$, the RMSEs all tend to be so small that very little would be lost from using our estimators.

Similar results to those just discussed hold when the DGP is a short/long-memory process "perturbed" by random noise, as considered by Sun and Phillips (2003). Such a process is empirically relevant as many measures of volatility are known to be noisy. The addition of white noise to the process $\{v_t\}$ is known to bias standard LP estimates downward. Since we are considering an LP estimator with lower frequencies trimmed out, we can only expect this downward bias to be exacerbated. The question that is answered here is, comparatively, how much worse is this downward bias? Table 8 displays the results when a very substantial white noise component is added to the long-memory $\{v_t\}$ process for $d = 0.2$ and 0.45. The white noise is Gaussian with variance set equal to four. When $d = 0.45$, we can see the substantial downward bias in all estimators. The bias is lower for those estimators that use more lower frequencies as a proportion of the total frequencies they use, with the standard estimator with the smallest bandwidth performing best. Making the more fair comparison between the standard estimator and its trimmed/adaptive counterparts by comparing at equal bandwidths, we see that the lower trimming does not usually induce a much larger downward bias. The increase in bias is typically less than or equal to 0.1. Very similar results hold for RMSE. When $d = 0.2$, all estimators are biased fairly close to zero. This enormous bias is likely due to the fact that the variance of the noise is more than three times as large as the variance of the long-memory component, probably exaggerating what would occur in practice. In any event, using the trimmed or adaptive estimators at the same bandwidth increases bias and RMSE by very small amounts.

We now turn to the cases in which our estimators may encounter the most problems: when significant short-memory components are present but contaminating elements are not. We consider various specifications of the following ARFIMA$(1, d, 1)$ process for $\{v_t\}$:

$$(1 - aL)(1 - L)^d v_t = (1 - bL)e_t,$$

where $e_t \sim i.i.d.N(0, 1)$. To begin, we consider a process that is persistent yet $I(0)$ by setting $a = 0.6$, $b = 0$ and $d = 0$. The memory parameter estimates of such a process are known to be upward biased. Indeed, the first blocks of Tables 9 and 10 show this to be the case. In terms of bias and RMSE, the standard estimator with $u = 0.5$ is clearly a strong favorite. Comparing our estimators with the standard one at the same bandwidths, we can see that the trimmed estimators roughly double the bias while the adaptive estimators increase bias to a *much* lesser extent. In many cases, the adaptive estimator hardly increases the bias at all.

Similar results hold for the RMSE of the estimators when compared at the same bandwidths: the trimmed estimator roughly doubles that of the standard estimator while the adaptive estimator increases it, but only slightly. In fact, as d grows, the adaptive estimator and the standard estimator become nearly identical because the adaptive estimator trims out increasingly fewer frequencies. This can be seen in the second blocks of Tables 9 and 10 for which $d = 0.45$. Thus, we can conclude that if we are not concerned about the presence of level shifts or deterministic trends but are concerned about a strong positive autoregressive component, the standard estimator with $u = 0.5$ is preferred. If we are concerned about both these contaminating elements and a strong positive autoregressive component, a version of the adaptive estimator is generally preferred. Somewhat similar results to those just discussed hold when we introduce persistence through the moving average parameter b. However, in this case, biases tend to be a lot smaller so that the standard estimator and our adaptive estimator both perform fairly well and very similarly. See the third blocks of Tables 9 and 10 for which $a = 0$, $b = -0.6$ and $d = 0.45$. The results are similar for other values of d. In this case very little is lost from using the adaptive estimator and much is to gain if level shifts or deterministic trends are suspected.

Now we turn to examining the performance of the estimators when the process is "antipersistent", having a significant negative first order autocorrelation. Our first example of this is obtained by setting $a = -0.6$, $b = 0$ and $d = 0$, given in the fourth blocks of Tables 9 and 10. In this case, the standard estimator with $u = 0.5$ is again preferable. Comparing the standard estimator to our estimators at fixed bandwidths, we can see that both the trimmed and adaptive estimator significantly increase the negative bias and RMSE, doubling to tripling them. The bias of our two estimators are quite similar here because the negative bias causes the adaptive estimator to typically use the maximal trimming, $l = T^{1/2+\varepsilon}$, that used by the trimmed estimator. It is in these cases that our estimators perform the worst. Unreported results show that similar results apply to the trimmed estimator when d is above zero. But as d grows, the adaptive estimator performs progressively better, being nearly comparable to the standard one (at a given bandwidth) when $d = 0.45$. Very similar results to those given in the fourth blocks of Tables 9 and 10 apply to cases for which $a = 0$ and $b = 0.6$, regardless of the value d takes.

Finally, we look into a case directly calibrated to data. Perron and Qu (2010) fitted the ARFIMA$(1, d, 1)$ model to an S&P 500 daily returns volatility series over the period 1973-2002 to obtain $\hat{a} = 0.298$, $\hat{b} = 0.751$ and $\hat{d} = 0.457$. Of course $\hat{d}$ may be upward biased if level shifts are present but we ignore this issue for the time being. We simulated the

ARFIMA series using these calibrated parameters, recording the results in the last blocks of Tables 9 and 10. The standard estimator with a low bandwidth is again clearly preferred and the trimmed and adaptive estimators are nearly identical due to the very strong negative bias. At a given bandwidth, the trimmed and adaptive estimators perform somewhat worse than the standard one in terms of both bias and RMSE although, in some cases, not to a significant extent.

In summary, if level shifts or other types of deterministic components are of concern, which should be the case in most practical applications, our estimators are clearly superior to the standard LP estimator. We recommend the use of our adaptive estimator with $(\varepsilon, u) = (0.05, 0.8)$ or something similar. This is the estimator that appears to best balance the potential biases arising from all of the DGPs studied in this section. Even when such contaminating components are not present, the simulations show that the cost of using this estimator is relatively small although gains could be made by using the standard estimator with a small bandwidth. On the other hand, if short-memory dynamics are of little concern, our trimmed estimator using a large trimming is clearly preferred.

5.3 Robustness to Non-Gaussianity

Given that many of the processes relevant to economics and finance that exhibit long-memory features are volatility processes, the applied econometrician may be concerned with Assumption 4. For example, when absolute or squared returns are used as the volatility process under examination, they are by construction non-Gaussian. As mentioned before, existing theoretical results for LP estimators primarily apply to Gaussian time series (with the exception of Velasco's, 2000, "pooled" estimator). In this section we show, through simulation results, that the Gaussian assumption is hardly critical and can likely be dropped or significantly weakened. We conducted numerous simulation experiments to assess the bias and RMSE of our estimators under a variety of distributional assumptions on the $\{v_t\}$ process. Following the design of Velasco (2000), we simulated series using an ARFIMA$(0, d, 0)$ model with innovation distributions of uniform$(-\sqrt{3}, \sqrt{3})$, re-centered exponential with parameter 1 and student-t with five degrees of freedom (which has only four moments). The results are almost identical in all of these cases so we only present a small subset in Table 11: t_5 distributed innovations with $d = 0.45$, $p = 0$ and $d = 0$, $p = 10$. Comparing these results to those corresponding to Gaussian specifications in Tables 7, 1 and 2, we can see that making the innovation distribution non-Gaussian causes our estimators to perform no worse, and in some cases marginally better. This is a generic result that holds for all combinations of the above distributions with $d = 0, 0.45$ and $p = 0, 10$.

6 Empirical Applications to Stock Market Volatility and Nile River Level Data

In this section, we apply our estimator to various time series that have been typified by the literature as arising from long-memory processes. We examine the extent to which level shifts or other deterministic trends may be playing a role in producing the long-memory features of the data. For each time series under study, we graphed the adaptive and trimmed versions of our estimators against the trimming parameter ε, setting $\alpha = 1$ and $m = T^{0.8}$. We let ε range from 0 to 0.25 in order to display the effects of smaller and larger trimmings on the d estimates, helping us to assess the most likely types of DGPs underlying the data. These graphs are provided in Figures 1-6. For each series, we also provide the standard LP estimate of d using $m = T^{0.5}$ (an estimator that performs relatively well in the absence of level shifts or deterministic components), the trimmed LP estimator using $(l, m) = (T^{0.65}, T^{0.9})$ (an estimator that performs relatively well in presence of these components and the absence of short-memory dynamics) and the adaptive LP estimator using $(l, m) = (T^{(1-2\hat{d}_{i-1})/(2-2\hat{d}_{i-1})+0.05}, T^{0.8})$ (an estimator that balances competing biases relatively well).

The first four time series we study are those examined by Lu and Perron (2010): log-squared daily returns series of the S&P 500, Dow Jones Industrial Average (DJIA), NASDAQ and AMEX stock market indices. Log-squared returns are a common measure of volatility and their empirical distributions more closely resemble Gaussian distributions than do those of absolute or squared returns. The S&P 500 series starts on July 3, 1962 and ends on March 25, 2004 (10504 observations); the DJIA series starts on March 4, 1957 and ends on October 30, 2002 (11534 observations); the NASDAQ series starts on December 15, 1972 and ends on December 31, 2006 (8592 observations); the AMEX series starts on July 3, 1962 and ends on December 31, 2006 (11201 observations). Note that these time series are all much (4 to 6 times) longer than those studied in the previous section. This means that potential biases arising from short-memory dynamics or noise components should be substantially lower as our estimators are consistent under these DGPs.

Starting with the S&P 500 volatility series, the standard LP estimator with a low bandwidth gives a memory parameter estimate of 0.527, indicating a non-stationary long-memory process. On the other hand, our highly trimmed estimator estimates d to be -0.017, indicating the (near) absence of long-memory. These contradicting initial results suggest that level shifts or deterministic trends are playing a major role in biasing the standard LP estimator upward. The adaptive estimator we advocate gives an estimate of 0.056, agreeing with this suggestion. Figure 1 provides a more complete picture. Both the adaptive and trimmed estimates start out by decreasing in the trimming parameter from 0.10/0.072 to 0.02/0.0214 and then increase again to 0.112/0.189. The first feature that emerges here is that all of these estimates are well below that of the standard estimator, re-emphasizing the fact that level shifts or deterministic trends appear to be biasing the standard estimate upwards. Second, the "u-shaped" pattern suggests competing biases are present. Since biases arising from level-shift/deterministic components should be decreasing in ε while those arising from short-memory dynamics should be increasing in ε, it appears that a level-shift/deterministic component slightly biases the estimates upwards for small ε. After this bias dies off for larger ε, an upward bias, likely from a positive AR or MA component, takes over. However, as in the other empirical results following this, it appears that estimates using trimmings above about 0.2 become somewhat erratic and unreliable. This can be attributed to the estimators using less data and the periodogram ordinates used in the LP regressions growing progressively farther from the origin. We also believe the fact that our memory parameter estimators are near zero is not due to the presence of noise. If this were the case, the d estimates should monotonically decrease toward zero as ε increases, not display this u-shaped pattern. In combination, these observations provide strong evidence that the true memory parameter of this time series is somewhere near zero.

Turning to the DJIA series, very similar results emerge. The standard LP estimator we examine gives an estimate of 0.428, this time in the stationary region, while the highly trimmed estimator is -0.034. The overall pattern present in Figure 2 is quite similar to that for the S&P 500 series, given in Figure 1. Here however, it seems that the upward bias arising from level shifts/deterministic components is a bit weaker since our advocated adaptive estimator gives a value of 0.040 and the d estimate curves are somewhat flatter and remain very close to zero until the "erratic range" for ε is reached. This could simply be due to a lower frequency of level shifts, as indicated by the results of Lu and Perron (2010).[7] The results for NASDAQ are again quite similar with the standard estimator being 0.568, the highly trimmed estimator being -0.028, the moderate adaptive estimator being 0.066 and the graph of the adaptive estimator displaying a similar pattern (see Figure 3). For all three of the above series, the results are highly indicative of a short-memory process (with a possible positive AR or MA component) contaminated by level shifts or deterministic components.

[7] This result could also be attributed to level shifts or deterministic trends of smaller magnitude.

The results for the AMEX series are somewhat similar but not as straightforward. The low bandwidth standard estimator is 0.450, the highly trimmed one is 0.038 and our suggested adaptive estimator is 0.137. However, graphing our estimators against ε in Figure 4 no longer produces a u-shaped pattern. Rather, the estimators decrease to values substantially below zero (although the estimators displaying these values seem fairly unreliable). This pattern is consistent with a short-memory process with a negative AR or MA component contaminated by level shifts/deterministic components but is also consistent with a short-memory process with a negative AR or MA component contaminated by noise. We still lean towards the former as being the true DGP since the sample size is so large, noise should not be a major factor and since the other volatility series were constructed identically and their d estimates do not appear to be largely biased by noise.

We now turn to a time series that is often thought to be a low noise measure of volatility: the log of daily realized volatility. The underlying realized volatility series was constructed from five minute returns observations of the S&P 500 futures index from April 21, 1982 to March 2, 2007 (6262 observations).[8] The features of the memory parameter estimates of this series are quite different from those of the other volatility series in that they are broadly in line with a true long-memory process. First, the low bandwidth standard LP estimator is 0.625, indicating a non-stationary long-memory process. Second, the highly trimmed estimator is 0.388, indicating a stationary long-memory process. Third, the adaptive estimator that we advocate gives a value of 0.503, indicating a highly persistent long-memory process on the border of stationarity. Looking at the graph of the adaptive and trimmed estimators in Figure 5 provides a fuller picture. We can see that both estimators are, on average, decreasing slightly in the trimming parameter and that, for a given trimming, the trimmed estimator lies slightly below its adaptive counterpart. These results are in line with a stationary long-memory process with a memory parameter around 0.4 to 0.45 that is contaminated by level shifts or deterministic trends (see Tables 3 and 5).

We finally turn to one of the classic time series for which long-memory was originally posited, the Nile River level series. This time series is the record of the minimum level that the Nile River reached within a given year, from 622 to 1284 A.D. (663 observations).[9] This series also exhibits consistent evidence of long-memory. The low bandwidth standard LP estimator is 0.504, the highly trimmed estimator is 0.515 and our advocated adaptive

[8]We are grateful to Shinsuke Ikeda for kindly providing us this dataset. For more details on its constructions see Ikeda (2009).

[9]These data were taken from Beran (1994).

estimator is 0.370, all being well above zero. Looking at Figure 6, we also see relatively stable memory parameter estimates that are significantly greater than zero although the estimators with higher trimmings should likely be ignored due to the very small sample size. Short-memory dynamics may be playing a minor role in biasing these estimators but they all roughly agree on a long-memory process with a memory parameter in the 0.4 to 0.5 range.

7 Conclusions and Future Research

In this paper, we have shown that very simple modifications to the standard LP estimator of the memory parameter lead to estimators that are robust to a wide range of level shifts and deterministic components. We showed that these trimmed or adaptive estimators have good asymptotic properties and perform well in finite samples. In order to balance potential competing biases, we advocate a particular version of our adaptive estimator, employing the trimming $l = T^{(1-2\hat{d}_{i-1})/(2-2\hat{d}_{i-1})+0.05}$ and bandwidth $m = T^{0.8}$ to balance potential competing biases. However, an automatic data-dependent, theoretically justified procedure for choosing these parameters remains an open question. Applying our estimators to volatility data, we found that for many of these series, level shifts or deterministic trends bias standard memory parameter estimates upwards. Moreover, for many of these time series, our estimators indicate memory parameters near zero. Nevertheless, we found evidence for the presence of long-memory in a log realized volatility series and a classic hydrological time series.

For future research, there appears to be many fruitful avenues. Methods aimed at simultaneously reducing biases arising from (i) level shifts and deterministic trends, (ii) short-memory dynamics (e.g., Andrews and Guggenberger, 2003) and (iii) the presence of noise (e.g., Sun and Phillips, 2003) will prove quite useful in practice. Thus far, we have only seen the emergence of estimators that tackle each of these problems separately, those provided here applying to (i). Apart from this important issue, we believe further improvements can be made to reducing the bias arising from level shifts and deterministic components. Remarks 3 and 4 provide suggestions in this vain. Finally, more work needs to be done regarding the asymptotic properties of standard and non-standard LP estimators in the presence of non-Gaussianity as simulation evidence clearly indicates that these estimators still perform well when the underlying DGP is non-Gaussian.

8 Mathematical Appendix

Proof of Theorem 1: We start with the case of the DGP of Assumption 1(a).
(i) By the decomposition of Perron and Qu (2010),

$$I_x(\lambda_j) = I_v(\lambda_j) + I_u(\lambda_j) + \frac{1}{\pi T} \sum_{t=1}^{T} \sum_{s=1}^{T} v_t u_{t,T} \cos(\lambda_j (t-s)),$$

where $I_v(\cdot)$ and $I_u(\cdot)$ denote the periodograms of $\{v_t\}_{t=1}^T$ and $\{u_{T,t}\}_{t=1}^T$. Hence, since $\{v_t\}$ and $\{u_{T,t}\}$ are independent,

$$EI_x(\lambda_j) = EI_v(\lambda_j) + EI_u(\lambda_j).$$

By Theorem 2(a) of Robinson (1995),

$$E\left[\frac{I_v(\lambda_j)}{f^*(0)\lambda_j^{-2d}}\right] = 1 + O\left[\frac{\log j}{j} + \left(\frac{j}{T}\right)^2\right],$$

and by Assumption 2 and Proposition 3 of Perron and Qu (2010),

$$E\left[\frac{I_u(\lambda_j)}{f^*(0)\lambda_j^{-2d}}\right] = O\left(\frac{T}{f^*(0)\lambda_j^{-2d} j^2}\right) = O\left(\frac{T^{1-2d}}{j^{2-2d}}\right).$$

(ii) We use the simple decomposition

$$w_x(\lambda_j)^2 = w_v(\lambda_j)^2 + w_u(\lambda_j)^2 + \frac{1}{\pi T} \sum_{t=1}^{T} \sum_{s=1}^{T} v_t u_{t,T} \exp(i\lambda_j (t+s)),$$

where $w_v(\cdot)$ and $w_u(\cdot)$ denote the discrete Fourier transforms of $\{v_t\}_{t=1}^T$ and $\{u_{T,t}\}_{t=1}^T$. Hence, again using the independence of $\{v_t\}$ and $\{u_{T,t}\}$,

$$Ew_x(\lambda_j)^2 = Ew_v(\lambda_j)^2 + Ew_u(\lambda_j)^2.$$

By Theorem 2(b) of Robinson (1995),

$$E\left[\frac{w_v(\lambda_j)^2}{f^*(0)\lambda_j^{-2d}}\right] = O\left(\frac{\log j}{j}\right).$$

Using similar techniques as in the proof of Proposition 3 in Perron and Qu (2010), we can show (using results of Georgiev, 2002),

$$\begin{aligned}\frac{1}{T} w_u(\lambda_j)^2 \Rightarrow & \frac{1}{2\pi} \int_0^1 \int_0^1 J(u)J(s)\cos(2\pi j(s+u))dsdu \\ & + \frac{i}{2\pi} \int_0^1 \int_0^1 J(u)J(s)\cos(2\pi j(s+u))dsdu,\end{aligned}$$

where $J(\cdot)$ denotes a compound Poisson process defined as $J(s) = \sum_{j=0}^{N(s)} \eta_j$ with $N(s)$ being a Poisson process with jump intensity p which is independent of η_j for all j. Thus, using the fact that $E[J(u)J(s)] = E[J(\min\{u,s\})^2] = p\sigma_\eta^2 \min\{u,s\}$,

$$
\begin{aligned}
&E\left[\frac{j^2}{2\pi}\int_0^1\int_0^1 J(u)J(s)\cos(2\pi j(s+u))dsdu\right] \\
&= \frac{j^2}{2\pi}\int_0^1\int_0^u E[J(u)J(s)]\cos(2\pi j(s+u))dsdu + \frac{j^2}{2\pi}\int_0^1\int_u^1 E[J(u)J(s)]\cos(2\pi j(s+u))dsdu \\
&= \frac{j^2 p\sigma_\eta^2}{2\pi}\int_0^1\int_0^u s\cos(2\pi j(s+u))dsdu + \frac{j^2 p\sigma_\eta^2}{2\pi}\int_0^1\int_u^1 u\cos(2\pi j(s+u))dsdu \\
&= -\frac{p\sigma_\eta^2}{8\pi^3}.
\end{aligned}
$$

Similarly,

$$
E\left[\frac{j^2}{2\pi}\int_0^1\int_0^1 J(u)J(s)\sin(2\pi j(s+u))dsdu\right] = 0
$$

so that $\lim_T E\left[(j^2/T)w_u(\lambda_j)^2\right] = -p\sigma_\eta^2/8\pi^3$ and

$$
E\left[\frac{w_u(\lambda_j)^2}{f^*(0)\lambda_j^{-2d}}\right] = O\left(\frac{T^{1-2d}}{j^{2-2d}}\right).
$$

(iii) The proof is similar to the proof of (ii), using the fact that

$$
E\left[w_x(\lambda_j)\,w_x(\lambda_j)^*\right] = E\left[w_v(\lambda_j)\,w_v(\lambda_j)^*\right] + E\left[w_u(\lambda_j)\,w_u(\lambda_j)^*\right]
$$

and applying Theorem 2(c) of Robinson (1995) and similar techniques.
(iv) The proof is again similar to the proof of (ii), using the fact that

$$
E\left[w_x(\lambda_j)\,w_x(\lambda_j)\right] = E\left[w_v(\lambda_j)\,w_v(\lambda_j)\right] + E\left[w_u(\lambda_j)\,w_u(\lambda_j)\right]
$$

and applying Theorem 2(d) of Robinson (1995).

We now consider the case of the DGP of Assumption 1 b), and begin by noting that

$$
EI_x(\lambda_j) = EI_v(\lambda_j) + \frac{1}{2\pi T}\sum_{i=1}^{B}\sum_{k=1}^{B} c_i c_k \sum_{t=1}^{T}\sum_{s=1}^{T} \mathbb{I}(T_{i-1} < t \leq T_i)\,\mathbb{I}(T_{k-1} < s \leq T_k)\cos(\lambda_j(t-s)).
$$

The second term in the above expression divided by T is

$$
\begin{aligned}
&\frac{1}{2\pi T^2}\sum_{i=1}^{B}\sum_{k=1}^{B}c_ic_k\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{I}\left(T_{i-1}<t\leq T_i\right)\mathbb{I}\left(T_{k-1}<s\leq T_k\right)\cos\left(\lambda_j(t-s)\right)\\
&\quad\to\frac{1}{2\pi}\sum_{i=1}^{B}\sum_{k=1}^{B}c_ic_k\int_0^1\int_0^1\mathbb{I}\left(\tau_{i-1}<u\leq\tau_i\right)\mathbb{I}\left(\tau_{k-1}<r\leq\tau_k\right)\cos\left(2\pi j(u-r)\right)dudr\\
&\quad=\frac{1}{2\pi}\sum_{i=1}^{B}\sum_{k=1}^{B}c_ic_k\int_{\tau_{k-1}}^{\tau_k}\int_{\tau_{i-1}}^{\tau_i}\cos\left(2\pi j(u-r)\right)dudr\\
&\quad=\frac{2}{(2\pi)^3j^2}\left[\sum_{i=1}^{B}\sum_{k=1}^{B}c_ic_k\left\{\cos\left(2\pi j\left(\tau_i-\tau_k\right)\right)-2\cos\left(2\pi j\left(\tau_{i-1}-\tau_k\right)\right)+\cos\left(2\pi j\left(\tau_{i-1}-\tau_{k-1}\right)\right)\right\}\right]\\
&\quad=O\left(j^{-2}\right).
\end{aligned}
$$

Thus, applying the results of Theorem 2(a) of Robinson (1995) to the above decomposition,

$$
\begin{aligned}
E\left[\frac{I_x\left(\lambda_j\right)}{f^*(0)\lambda_j^{-2d}}\right]=&E\left[\frac{I_v\left(\lambda_j\right)}{f^*(0)\lambda_j^{-2d}}\right]+O\left(\frac{T}{j^2}\right)\frac{\lambda_j^{2d}}{f^*(0)}\\
=&1+O\left[\frac{\log j}{j}+\left(\frac{j}{T}\right)^2+\frac{T^{1-2d}}{j^{2-2d}}\right].
\end{aligned}
$$

The proofs of parts (ii)-(iv) are similar, using the fact that

$$
\begin{aligned}
E\left[w_x\left(\lambda_j\right)^2\right]=&E\left[w_v\left(\lambda_j\right)^2\right]\\
&+\frac{1}{2\pi T}\sum_{i=1}^{B}\sum_{k=1}^{B}c_ic_k\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{I}\left(T_{i-1}<t\leq T_i\right)\mathbb{I}\left(T_{k-1}<s\leq T_k\right)\exp\left(i\lambda_j(t+s)\right)\\
E\left[w_x\left(\lambda_j\right)w_x\left(\lambda_k\right)^*\right]=&E\left[w_v\left(\lambda_j\right)w_v\left(\lambda_k\right)^*\right]\\
&+\frac{1}{2\pi T}\sum_{i=1}^{B}\sum_{k=1}^{B}c_ic_k\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{I}\left(T_{i-1}<t\leq T_i\right)\mathbb{I}\left(T_{k-1}<s\leq T_k\right)\exp\left(i\left(\lambda_jt-\lambda_ks\right)\right)\\
E\left[w_x\left(\lambda_j\right)w_x\left(\lambda_k\right)\right]=&E\left[w_v\left(\lambda_j\right)w_v\left(\lambda_k\right)\right]\\
&+\frac{1}{2\pi T}\sum_{i=1}^{B}\sum_{k=1}^{B}c_ic_k\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{I}\left(T_{i-1}<t\leq T_i\right)\mathbb{I}\left(T_{k-1}<s\leq T_k\right)\exp\left(i\left(\lambda_jt+\lambda_ks\right)\right).
\end{aligned}
$$

We now consider the DGP of Assumption 1 c). First note that Qu (2008) and Künsch (1986) have shown that $|\sum_{t=1}^{T}h(t/T)\cos(\lambda_jt)|$ and $|\sum_{t=1}^{T}h(t/T)\sin(\lambda_jt)|$ are $O(T/j)$ when h is Lipschitz continuous or montonic and bounded (respectively). Note that this bound is not exact, i.e., these quantities may grow at a slower rate than $O(T/j)$, depending on the properties of $h(\cdot)$. Together with Robinson(1995, Theorem 2), this implies the results since

$$\begin{aligned}
E\left[I_x(\lambda_j)\right] =& E\left[I_v(\lambda_j)\right] + \frac{1}{2\pi T}\left|\sum_{t=1}^{T} h(t/T)\cos(\lambda_j t)\right|^2 + \frac{1}{2\pi T}\left|\sum_{t=1}^{T} h(t/T)\sin(\lambda_j t)\right|^2 \\
E\left[w_x(\lambda_j)^2\right] =& E\left[w_v(\lambda_j)^2\right] + \frac{1}{2\pi T}\left(\sum_{t=1}^{T} h(t/T)\exp(i\lambda_j t)\right)^2 \\
E\left[w_x(\lambda_j) w_x(\lambda_k)^*\right] =& E\left[w_v(\lambda_j) w_v(\lambda_k)^*\right] + \frac{1}{2\pi T}\left(\sum_{t=1}^{T} h(t/T)\exp(i\lambda_j t)\right)\left(\sum_{t=1}^{T} h(t/T)\exp(-i\lambda_k t)\right) \\
E\left[w_x(\lambda_j) w_x(\lambda_k)\right] =& E\left[w_v(\lambda_j) w_v(\lambda_k)\right] + \frac{1}{2\pi T}\left(\sum_{t=1}^{T} h(t/T)\exp(i\lambda_j t)\right)\left(\sum_{t=1}^{T} h(t/T)\exp(i\lambda_k t)\right)
\end{aligned}$$

and

$$\begin{aligned}
&\left|\sum_{t=1}^{T} h(t/T)\exp(i\lambda_j t)\right|^2 \leq \left(\left|\sum_{t=1}^{T} h(t/T)\cos(\lambda_j t)\right| + \left|\sum_{t=1}^{T} h(t/T)\sin(\lambda_j t)\right|\right)^2 \\
&\left|\left(\sum_{t=1}^{T} h(t/T)\exp(i\lambda_j t)\right)\left(\sum_{t=1}^{T} h(t/T)\exp(\pm i\lambda_k t)\right)\right| \\
&\quad \leq \left(\left|\sum_{t=1}^{T} h(t/T)\cos(\lambda_j t)\right| + \left|\sum_{t=1}^{T} h(t/T)\sin(\lambda_j t)\right|\right)\left(\left|\sum_{t=1}^{T} h(t/T)\cos(\lambda_k t)\right| + \left|\sum_{t=1}^{T} h(t/T)\sin(\lambda_k t)\right|\right).
\end{aligned}$$

■

The proofs of Theorems 2 and 3 follow those of 1 and 2 in HDB and rely on their results. First, we state and prove a series of lemmas, the majority of which have counterparts in HDB (with the exceptions of Lemmas A.3 and A.8). We try to follow their notation as closely as possible in order to maintain this correspondence. Define $a_j = Y_j - \bar{Y}$ and $S_{YY} = \sum_{k=l}^{m} a_k^2$, so that

$$\hat{d} - d = -\frac{1}{2S_{YY}}\sum_{j=l}^{m} a_j \log f_j^* - \frac{1}{2S_{YY}}\sum_{j=l}^{m} a_j \varepsilon_j, \tag{A.1}$$

where $\varepsilon_j = \log(I_j/f_j) + C$ and C is Euler's constant.

Lemma A.1. *Under Assumption 3, $a_j = O(\log m)$ for all $l \leq j \leq m$ and $S_{YY} = m + o(m)$.*

Proof: The first statement follows directly from a very slight modification of the arguments used by Hurvich and Beltrao (1994, pp. 299-301). For the second statement, note that following the same arguments,

$$a_j = \log j - \frac{1}{m-l+1}\left[\log m! - \log(l-1)!\right] + o(1).$$

By Stirling's formula and Assumption 3,

$$\frac{1}{m-l+1}\left[\log m! - \log(l-1)!\right] = \frac{m}{m-l+1}\log m - 1 + O\left(\frac{\log m}{m}\right).$$

Thus,

$$\begin{aligned}
\frac{1}{m}\sum_{j=l}^{m} a_j^2 &= \frac{1}{m}\sum_{j=l}^{m}\left(\log j - \frac{1}{m-l+1}\left[\log m! - \log(l-1)!\right]\right)^2 \\
&= \frac{1}{m}\sum_{j=l}^{m}\left\{\log j - \frac{m}{m-l+1}\log m + 1 + O\left(\frac{\log m}{m}\right)\right\}^2 \\
&= 1 + O\left(\frac{l}{m}\right) + O\left(\frac{\log^2 m}{m}\right) \qquad\qquad \text{(A.2)} \\
&\quad + \frac{1}{m}\sum_{j=l}^{m}\left(\log j - \frac{m}{m-l+1}\log m\right)^2 + \frac{2}{m}\sum_{j=l}^{m}\left(\log j - \frac{m}{m-l+1}\log m\right).
\end{aligned}$$

Now, by Assumption 3, the last two terms of (A.2) are

$$\begin{aligned}
&\frac{1}{m}\sum_{j=1}^{m}\left(\log j - \frac{m}{m-l+1}\log m\right)^2 - \frac{1}{m}\sum_{j=1}^{l-1}\left(\log j - \frac{m}{m-l+1}\log m\right)^2 \\
&\quad + \frac{2}{m}\sum_{j=l}^{m}\left(\log j - \frac{m}{m-l+1}\log m\right) - \frac{2}{m}\sum_{j=l}^{l-1}\left(\log j - \frac{m}{m-l+1}\log m\right) \\
&= \int_0^1 \log^2(x)dx + 2\int_0^1 \log(x)dx + o(1) + O\left(\frac{1}{m}\sum_{j=1}^{l-1}\log^2 m\right) + O\left(\frac{1}{m}\sum_{j=1}^{l-1}\log m\right) \\
&= O\left(\frac{l\log^2 m}{m}\right) + o(1) = o(1).
\end{aligned}$$

Hence, applying Assumption 3 again, (A.2) is equal to $1 + o(1)$. ■

Lemma A.2. *Under Assumptions 2 and 3,*

$$-\frac{1}{2S_{YY}}\sum_{j=l}^{m} a_j \log f_j^* = \frac{-2\pi^2}{9}\frac{f^{*\prime\prime}(0)}{f^*(0)}\frac{m^2}{T^2} + o\left(\frac{m^2}{T^2}\right).$$

Proof: Using results in the proof of Lemma 1 of HDB (pages 37-38) and Assumption 2,

$$\sum_{j=l}^{m} a_j \log f_j^* = \frac{f^{*\prime\prime}(0)}{2f^*(0)}\sum_{j=l}^{m} a_j \lambda_j^2 + R,$$

where $R = O\left(T^{-3}m^4\log m\right)$. From Hurvich and Beltrao (1994, pp. 299-301),

$$a_j = \log j - \frac{1}{m-l+1}\sum_{k=l}^{m}\log k + O\left(\frac{m^2}{T^2}\right).$$

Also note that

$$\begin{aligned}\sum_{j=l}^{m} j^2 \log j =& \frac{1}{6} m(m+1)(2m+1)\log m - \frac{m^3}{9} + o\left(m^3\right) \\ & - \frac{1}{6} l(l-1)(2l-1)\log(l-1) - \frac{(l-1)^3}{9} + o\left(l^3\right) \\ =& \frac{1}{6} m(m+1)(2m+1)\log m - \frac{m^3}{9} + o\left(m^3\right)\end{aligned}$$

by HDB (page 38) and Assumption 3,

$$\sum_{j=l}^{m} j^2 = \frac{1}{6} m(m+1)(2m+1) - \frac{1}{6} l(l-1)(2l-1),$$

and

$$\begin{aligned}\frac{1}{m-l+1}\sum_{k=l}^{m}\log k =& \frac{m}{m-l+1}\frac{1}{m}\log m! - \frac{1}{m-l+1}\sum_{k=1}^{l-1}\log k \\ =& \frac{m}{m-l+1}\left(\log m - 1 + o(1)\right) + O\left(\frac{l \log l}{m}\right) \\ =& \frac{m}{m-l+1}\left(\log m - 1\right) + o(1),\end{aligned}$$

by Stirling's formula and Assumption 3. Using these results, we have

$$\begin{aligned}\sum_{j=l}^{m} a_j \log f_j^* =& \frac{2\pi^2}{T^2}\frac{f^{*\prime\prime}(0)}{f^*(0)}\left\{\sum_{j=l}^{m} j^2 \log j - \frac{1}{m-l+1}\sum_{k=l}^{m}\log k \sum_{j=l}^{m} j^2 + O\left(\frac{m^2}{T^2}\right)\sum_{j=l}^{m} j^2\right\} \\ & + O\left(T^{-3} m^4 \log m\right) \\ =& \frac{2\pi^2}{T^2}\frac{f^{*\prime\prime}(0)}{f^*(0)}\left\{-\frac{m^3}{9} + \frac{1}{6} m(m+1)(2m+1) + o\left(m^3\right)\right\} + O\left(T^{-3} m^4 \log m\right)\end{aligned}$$

by Assumption 3. Thus Lemma A.1 and Assumption 3 provide,

$$\begin{aligned}-\frac{1}{2S_{YY}}\sum_{j=l}^{m} a_j \log f_j^* =& -\frac{\pi^2}{mT^2}\frac{f^{*\prime\prime}(0)}{f^*(0)}\left\{\frac{2m^3}{9} + o\left(m^3\right)\right\} + O\left(T^{-3} m^3 \log m\right) \\ =& \frac{-2\pi^2}{9}\frac{f^{*\prime\prime}(0)}{f^*(0)}\frac{m^2}{T^2} + o\left(\frac{m^2}{T^2}\right). \quad \blacksquare\end{aligned}$$

Note that the normalized periodogram can be expressed as follows:

$$\frac{I_x(\lambda_j)}{f(\lambda_j)} = \left(\frac{A_j}{f(\lambda_j)^{1/2}}\right)^2 + \left(\frac{B_j}{f(\lambda_j)^{1/2}}\right)^2,$$

where

$$A_j \equiv \frac{1}{(2\pi T)^{1/2}} \sum_{t=1}^{T} x_t \cos(\lambda_j t) \qquad B_j \equiv \frac{1}{(2\pi T)^{1/2}} \sum_{t=1}^{T} x_t \sin(\lambda_j t).$$

Now, define the vector

$$\gamma = \left(\frac{A_j}{f_j^{1/2}}, \frac{B_j}{f_j^{1/2}}, \frac{A_k}{f_k^{1/2}}, \frac{B_k}{f_k^{1/2}} \right)'.$$

The next Lemma details how well γ can be approximated by a multivariate Gaussian random variable when considering Assumption 1(a). Its results will be used by a later lemma to derive Edgeworth approximations.

Lemma A.3. *Under Assumptions 1(a) and 4, for any sequences of positive integers, $j = j(T)$ and $k = k(T)$ such that $j > k$ and $j/T \to 0$ as $T \to \infty$, the following result holds: for $n > 2$, the n^{th} cumulants of γ are $O(T^{n/2-nd}/k^{n-nd})$.*

Proof: Before proceeding, note that

$$A_j = \frac{1}{(2\pi T)^{1/2}} \sum_{t=1}^{T} v_t \cos(\lambda_j t) + \frac{1}{(2\pi T)^{1/2}} \sum_{t=1}^{T} u_{T,t} \cos(\lambda_j t) \equiv A_j^v + A_j^u$$

and similarly for A_k, B_j and B_k. Let γ_i denote the i^{th} entry of vector γ and $\kappa(X_1, \ldots, X_n)$ denote the joint cumulant of random variables $X_1, \ldots, X_n$ so that we have

$$\kappa(X_1, \ldots, X_n) = \sum_{\Pi} (|\Pi| - 1)!(-1)^{|\Pi|-1} \prod_{B \in \Pi} E\left(\prod_{i \in B} X_i \right), \tag{A.3}$$

where Π runs through the list of all partitions of $\{1, \ldots, n\}$, B runs through the list of all blocks of the partition Π and $|\cdot|$ denotes the number of elements in a set. Using the independence of $\{v_t\}$ and $\{u_{T,t}\}$ and the properties of cumulants, the following holds for any n^{th} joint cumulant of γ:

$$\begin{aligned} \kappa\left(\gamma_1^{n_1}, \gamma_2^{n_2}, \gamma_3^{n_3}, \gamma_4^{n_4}\right) =& \kappa\left(\left[\frac{A_j^v}{f_j^{1/2}}\right]^{n_1}, \left[\frac{A_k^v}{f_k^{1/2}}\right]^{n_2}, \left[\frac{B_j^v}{f_j^{1/2}}\right]^{n_3}, \left[\frac{B_k^v}{f_k^{1/2}}\right]^{n_4} \right) \\ &+ \kappa\left(\left[\frac{A_j^u}{f_j^{1/2}}\right]^{n_1}, \left[\frac{A_k^u}{f_k^{1/2}}\right]^{n_2}, \left[\frac{B_j^u}{f_j^{1/2}}\right]^{n_3}, \left[\frac{B_k^u}{f_k^{1/2}}\right]^{n_4} \right) \\ =& f_j^{-\frac{n_1+n_3}{2}} f_k^{-\frac{n_2+n_4}{2}} \kappa\left(\left[A_j^u\right]^{n_1}, \left[A_k^u\right]^{n_2}, \left[B_j^u\right]^{n_3}, \left[B_k^u\right]^{n_4} \right), \end{aligned} \tag{A.4}$$

where n_1, n_2, n_3 and n_4 are nonnegative integers that sum to n. The second equality follows from Assumption 4. Upon inspection of (A.3), it becomes apparent that what is of concern are sums of products over $B \in \Pi$ of terms of the form

$$E\left[\left(A_j^u\right)^{r_1} \left(A_k^u\right)^{r_2} \left(B_j^u\right)^{r_3} \left(B_k^u\right)^{r_4} \right],$$

where r_1, r_2, r_3 and r_4 are nonnegative integers that sum to $|B|$, $\sum_{B\in\Pi} r_i = n_i$ and $\sum_{B\in\Pi} |B| = n$. Now consider as given block B of a fixed partition $|\Pi|$. Using the results of Georgiev (2002) and similar reasoning to that used in the proof of Theorem 1(ii),

$$\begin{aligned}\frac{j^{r_1+r_3}k^{r_2+r_4}}{T^{|B|/2}}\left(A_j^u\right)^{r_1}\left(A_k^u\right)^{r_2}\left(B_j^u\right)^{r_3}\left(B_k^u\right)^{r_4} \Rightarrow & \frac{j^{r_1+r_3}k^{r_2+r_4}}{(2\pi)^{|B|/2}}\int_0^1 \dots \int_0^1 \left(\prod_{i=1}^{|B|} J\left(u_i\right)\right)\left(\prod_{i=1}^{r_1}\cos\left(2\pi j u_i\right)\right)\times \\ & \times\left(\prod_{i=r_1+1}^{r_1+r_2}\cos\left(2\pi k u_i\right)\right)\left(\prod_{i=r_1+r_2+1}^{|B|-r_4}\sin\left(2\pi j u_i\right)\right)\times \\ & \times\left(\prod_{i=|B|-r_4+1}^{|B|}\sin\left(2\pi k u_i\right)\right)du_1\dots du_{|B|}. \qquad \text{(A.5)}\end{aligned}$$

Now, the expectation of the first term inside the integral can be decomposed as follows. Without loss of generality (WLOG), suppose $u_1 \leq u_2 \leq \ldots \leq u_{|B|}$. Then

$$\begin{aligned}E\left[\prod_{i=1}^{|B|} J\left(u_i\right)\right] =& E\left[\prod_{i=1}^{|B|-1} J\left(u_i\right)\sum_{k=1}^{|B|}\left\{J\left(u_k\right)-J\left(u_{k-1}\right)\right\}\right] \\ =& \ldots = E\left[\prod_{i=1}^{|B|}\sum_{k=1}^{i}\left\{J\left(u_k\right)-J\left(u_{k-1}\right)\right\}\right] \\ =& \sum_{i_{|B|}=1}^{|B|}\sum_{i_{|B|-1}=1}^{|B|-1}\dots\sum_{i_2=1}^{2}\sum_{i_1=1}^{1} E\left[\prod_{k=1}^{|B|}\left\{J\left(u_{i_k}\right)-J\left(u_{i_k-1}\right)\right\}\right]. \qquad \text{(A.6)}\end{aligned}$$

Since $J(\cdot)$ is a zero mean process with independent increments, the terms $E[\prod_{k=1}^{|B|}\{J(u_{i_k}) - J(u_{i_k-1})\}]$ are only nonzero when for each $k = 1, \ldots, |B|$, there is some $j \neq k$ with $j \in \{1, \ldots, |B|\}$ such that $i_k = i_j$. This implies that (A.6) is a sum of terms of the form $\prod_{k=1}^{|B|} E[\{J(u_k) - J(u_{k-1})\}^{\alpha_k}]$, where the α_k's are nonnegative integers such that $\sum_{k=1}^{|B|} \alpha_k = |B|$.

The moment generating function of $J(t)$ is given as follows for any $t \in [0, 1]$:

$$M_{J(t)}(s) \equiv E\left[\exp\left(J(t)s\right)\right] = \exp\left(pt\left\{M_\eta(s) - 1\right\}\right),$$

where $M_\eta(s) \equiv E\left[\exp\left(\eta_i s\right)\right]$. The increments of $J(\cdot)$ are stationary, implying that $J\left(u_k\right) - J\left(u_{k-1}\right)$ is distributed identically to $J\left(u_k - u_{k-1}\right)$ so that

$$E\left[\left\{J\left(u_k\right) - J\left(u_{k-1}\right)\right\}^{\alpha_k}\right] = E\left[\left\{J\left(u_k - u_{k-1}\right)\right\}^{\alpha_k}\right].$$

Since all of the moments of η_i exist, it can be seen from repeated differentiation of the moment generating function of $J(u_k - u_{k-1})$ and evaluating it at zero, that this is a polynomial of degree α_k in $(u_k - u_{k-1})$, where the polynomial coefficients are products of the moments of

η_i and finite positive constants. In turn, (A.6) is a sum of polynomials in the increments $(u_k - u_{k-1})$, each of degree less than or equal to $|B|$. Thus we can deduce that the term $E[\prod_{i=1}^{|B|} J(u_i)]$ is a function of $u_1, \ldots, u_{|B|}$ that is bounded for all $u_i \in [0,1]$, $i = 1, \ldots, |B|$. Denote this function as $g(u_1, \ldots, u_{|B|})$.

The expectation of the right hand side of (A.5) is thus

$$\frac{j^{r_1+r_3}k^{r_2+r_4}}{(2\pi)^{|B|/2}} \int_0^1 \cdots \int_0^1 g\left(u_1, \ldots, u_{|B|}\right) \left(\prod_{i=1}^{r_1} \cos\left(2\pi j u_i\right)\right) \left(\prod_{i=r_1+1}^{r_1+r_2} \cos\left(2\pi k u_i\right)\right) \times$$
$$\times \left(\prod_{i=r_1+r_2+1}^{|B|-r_4} \sin\left(2\pi j u_i\right)\right) \left(\prod_{i=|B|-r_4+1}^{|B|} \sin\left(2\pi k u_i\right)\right) du_1 \ldots du_{|B|} \tag{A.7}$$

Since $g\left(u_1, \ldots, u_{|B|}\right)$ is a sum of polynomials in the increments $(u_k - u_{k-1})$, for any constant k, $f\left(k, \ldots, k, u_i, k, \ldots, k\right)$, is a polynomial in u_i of degree less than or equal to $|B|$. Using this together with the facts that for $\alpha > 0$,

$$\int_0^1 x^\alpha \cos\left(2\pi j x\right) dx = \frac{\sin\left(2\pi j\right)}{2\pi j} - \frac{\alpha}{2\pi j} \int_0^1 x^{\alpha-1} \sin\left(2\pi j x\right) dx = O\left(j^{-1}\right)$$
$$\int_0^1 x^\alpha \sin\left(2\pi j x\right) dx = -\frac{\cos\left(2\pi j\right)}{2\pi j} + \frac{\alpha}{2\pi j} \int_0^1 x^{\alpha-1} \cos\left(2\pi j x\right) dx = O\left(j^{-1}\right),$$

we obtain that (A.7) is

$$\frac{j^{r_1+r_3}k^{r_2+r_4}}{(2\pi)^{|B|/2}} \prod_{i=1}^{r_1} O\left(j^{-1}\right) \prod_{i=r_1+1}^{r_1+r_2} O\left(k^{-1}\right) \prod_{i=r_1+r_2+1}^{|B|-r_4} O\left(j^{-1}\right) \prod_{i=|B|-r_4+1}^{|B|} O\left(k^{-1}\right) = O(1).$$

This implies that

$$E\left[\left(A_j^u\right)^{r_1} \left(A_k^u\right)^{r_2} \left(B_j^u\right)^{r_3} \left(B_k^u\right)^{r_4}\right] = O\left(\frac{T^{|B|/2}}{j^{r_1+r_3}k^{r_2+r_4}}\right).$$

Referring back to earlier reasoning derived from (A.3), this means that

$$\begin{aligned}
\kappa\left(\left[A_j^u\right]^{n_1}, \left[A_k^u\right]^{n_2}, \left[B_j^u\right]^{n_3}, \left[B_k^u\right]^{n_4}\right) &= \sum_{\Pi} (|\Pi|-1)!(-1)^{|\Pi|-1} \prod_{B\in\Pi} O\left(\frac{T^{|B|/2}}{j^{r_1+r_3}k^{r_2+r_4}}\right) \\
&= \sum_{\Pi} (|\Pi|-1)!(-1)^{|\Pi|-1} O\left(\frac{T^{\sum_{B\in\Pi}|B|/2}}{j^{\sum_{B\in\Pi}(r_1+r_3)}k^{\sum_{B\in\Pi}(r_2+r_4)}}\right) \\
&= \sum_{\Pi} (|\Pi|-1)!(-1)^{|\Pi|-1} O\left(\frac{T^{n/2}}{j^{n_1+n_3}k^{n_2+n_4}}\right) \\
&= O\left(\frac{T^{n/2}}{j^{n_1+n_3}k^{n_2+n_4}}\right).
\end{aligned}$$

Finally, from (A.4) we then have

$$\begin{aligned}\kappa\left(\gamma_1^{n_1},\gamma_2^{n_2},\gamma_3^{n_3},\gamma_4^{n_4}\right)=&f_j^{-\frac{n_1+n_3}{2}}f_k^{-\frac{n_2+n_4}{2}}O\left(\frac{T^{n/2}}{j^{n_1+n_3}k^{n_2+n_4}}\right)\\=&O\left(\lambda_j^{d(n_1+n_3)}\lambda_k^{d(n_2+n_4)}\right)O\left(\frac{T^{n/2}}{j^{n_1+n_3}k^{n_2+n_4}}\right)\\=&O\left(\frac{T^{n/2-nd}}{j^{(1-d)(n_1+n_3)}k^{(1-d)(n_2+n_4)}}\right)\\=&O\left(T^{n/2-nd}/k^{n-nd}\right).\quad\blacksquare\end{aligned}$$

We now present a lemma that applies to all three DGPs in Assumption 1. Recall that ε_t is defined by (A.1).

Lemma A.4. *Under Assmptions 1-3,* $\mathrm{Cov}(\varepsilon_j,\varepsilon_k)=O(\log^2 j/k^2+T^{3/2-3d}/k^{3-3d})$ *uniformly for* $l\le k<j\le m$.

Proof: The proof for the DGP of Assumption 1(a) involves the use of Lemma A.3 via an Edgeworth expansion. We begin with this. Define $\chi_j=\log(I_j/f_j)-E[\log(I_j/f_j)]$ and $\psi=\Sigma^{-1}$, where $\Sigma=\mathrm{Cov}(\gamma)$. Note that in what follows, any mention of uniformity is taken to mean the property holds uniformly for $l\le k<j\le m$.

The results of Lemma A.3 allow us to make an asymptotic multivariate Edgeworth expansion of the density of γ in terms of a Gaussian distribution since, under Assumption 3, the higher order cumulants of γ tend to zero more rapidly, the higher the order of the cumulant. More specifically, letting $f_\gamma(\cdot)$ denote the density of γ, the second order expansion provides (see page 172 of Skovgaard, 1986)

$$\begin{aligned}f_\gamma(g)=&(2\pi)^{-2}\left|\psi\right|^{1/2}\exp\left(-\frac{g'\psi g}{2}\right)\\&+O\left(\frac{T^{3/2-3d}}{k^{3-3d}}\right)(2\pi)^{-2}\left|\psi\right|^{1/2}\exp\left(-\frac{g'\psi g}{2}\right)\left(\sum_{1\le m,n,r\le 4}g_mg_ng_r+\sum_{1\le m\le 4}g_m\right),\end{aligned}$$

where g_i denotes the i^{th} entry of vector g. From this Edgeworth expansion, we obtain

$$\begin{aligned}E\left[\chi_j\chi_k\right]=&(2\pi)^{-2}\left|\psi\right|^{1/2}\int\int\int\int\chi_j\chi_k\exp\left(-\frac{g'\psi g}{2}\right)dg\\&+O\left(\frac{T^{3/2-3d}}{k^{3-3d}}\right)(2\pi)^{-2}\left|\psi\right|^{1/2}\int\int\int\int\chi_j\chi_k\exp\left(-\frac{g'\psi g}{2}\right)\left(\sum_{1\le m,n,r\le 4}g_mg_ng_r\right)dg\\&+O\left(\frac{T^{3/2-3d}}{k^{3-3d}}\right)(2\pi)^{-2}\left|\psi\right|^{1/2}\int\int\int\int\chi_j\chi_k\exp\left(-\frac{g'\psi g}{2}\right)\left(\sum_{1\le m,n\le 4}g_mg_n\right)dg\end{aligned}\tag{A.8}$$

uniformly. Applying Theorem 1, by Lemmas 2 and 3 of HDB (with minor modification), the first term of (A.8) is $O(\log^2 j/k^2+T^{2-4d}/j^{2-2d}k^{2-2d})$. Consider now the terms arising from the sums in the last two terms of (A.8). Each will be of the form

$$(2\pi)^{-2}|\psi|^{1/2}\int\int\int\int \chi_j\chi_k g_1^{n_1}g_2^{n_2}g_3^{n_3}g_4^{n_4}\exp\left(-\frac{g'\psi g}{2}\right)dg, \tag{A.9}$$

where n_1, n_2, n_3, n_4 are nonnegative integers that sum to 3 in the first sum and 1 in the second. Given Assumption 3 and Theorem 1, $\Sigma = (1/2)I_4 + o(1)$ uniformly, where I_4 is the 4×4 identity matrix. It follows that $\psi = 4I_4 + o(1)$ and $|\psi| = O(1)$ uniformly so we can conclude that (A.9) is $O(1)$ uniformly. Hence, we can conclude that (A.8) is $O(\log^2 j/k^2) + O(T^{3/2-3d}/k^{3-3d})$ uniformly.

The proofs for the DGPs given by Assumption 1(b) and (c) follow straightforwardly from the proof techniques of Lemmas 2 and 3 of HDB (using our Theorem 1) since the random vector γ remains Gaussian in these cases and an Edgeworth expansion is unnecessary. ■

Lemma A.5. *Under Assumptions 1-3,*

$$E[\varepsilon_j] = O(\log j/j) + O\left(\frac{T^{3/2-3d}}{j^{3-3d}}\right)$$

uniformly for $l \le j \le m$.

Proof: Very similar to the proof of Lemmas A.3 and A.4. ■

Lemma A.6. *Under Assumptions 1-3,*

$$\mathrm{Var}(\varepsilon_j) = \pi^2/6 + O(\log j/j) + O\left(\frac{T^{3/2-3d}}{j^{3-3d}}\right)$$

uniformly for $l \le j \le m$.

Proof: Very similar to the proof of Lemmas A.3 and A.4. ■

Lemma A.7. *Under Assumptions 1-3,*

$$-\frac{1}{2S_{YY}}\sum_{j=l}^{m} a_j E(\varepsilon_j) = O\left(\frac{\log^3 m}{m} + \frac{T^{3/2-3d}\log^2 m}{m^{3-3d}}\right).$$

Proof: By Lemmas A.1 and A.5,

$$\begin{aligned}\left|\frac{1}{2S_{YY}}\sum_{j=l}^{m} a_j E(\varepsilon_j)\right| &\le \frac{1}{2S_{YY}}\sum_{j=l}^{m} a_j E|\varepsilon_j| = O\left(m^{-1}\log m\left[\sum_{j=l}^{m}\frac{\log j}{j} + T^{3/2-3d}\sum_{j=l}^{m}\frac{1}{j^{3-3d}}\right]\right)\\ &= O\left(\frac{\log^3 m}{m} + \frac{T^{3/2-3d}\log^2 m}{m^{3-3d}}\right). \quad ■\end{aligned}$$

Using the results of these seven lemmas, we can now prove Theorem 2.

Proof of Theorem 2: This proof follows that of Theorem 1 of HDB with appropriate modifications. Part (i) follows directly from Lemmas A.2 and A.7. For part (ii), note first that

$$\mathrm{Var}(\hat{d}) = \frac{1}{4S_{YY}^2}\sum_{j=l}^{m} a_j^2 \,\mathrm{Var}\left(\varepsilon_j\right) + \frac{1}{2S_{YY}^2}\sum_{k=l}^{m}\sum_{j=k+1}^{m} a_j a_k \,\mathrm{Cov}\left(\varepsilon_j, \varepsilon_k\right). \tag{A.10}$$

Now, applying Lemmas A.6 and A.4, then Lemma A.1,

$$\begin{aligned}
&\sum_{j=l}^{m} a_j^2 \,\mathrm{Var}\left(\varepsilon_j\right) + 2\sum_{k=l}^{m}\sum_{j=k+1}^{m} a_j a_k \,\mathrm{Cov}\left(\varepsilon_j, \varepsilon_k\right) \\
&= \sum_{j=l}^{m} a_j^2 \left\{\frac{\pi^2}{6} + O\left(\frac{\log j}{j}\right) + O\left(\frac{T^{3/2-3d}}{j^{3-3d}}\right)\right\} + O\left(\log^2 m \sum_{k=l}^{m}\sum_{j=k+1}^{m}\left[\frac{\log^2 j}{k^2} + \frac{T^{3/2-3d}}{k^{3-3d}}\right]\right) \\
&= \frac{\pi^2 m}{6} + o(m) + O\left(\log^2 m \sum_{j=l}^{m}\frac{\log j}{j}\right) + O\left(T^{3/2-3d} m^{1-3(1-d)}\log^3 m\right) \\
&\quad + O\left(m\log^4 m \sum_{k=l}^{m}\frac{1}{k^2}\right) + O\left(T^{3/2-3d} m\log^2 m\sum_{k=l}^{m}\frac{1}{k^{3-3d}}\right) \\
&= \frac{\pi^2 m}{6} + o(m) + O\left(\log^4 m\right) + O\left(\frac{m\log^4 m}{l}\right) + O\left(T^{3/2-3d} m^{2-3(1-d)}\log^3 m\right) \\
&= \frac{\pi^2 m}{6} + o(m) + O\left(T^{3/2-3d} m^{2-3(1-d)}\log^3 m\right).
\end{aligned}$$

Then again using Lemma A.1 together with (A.10) shows part (ii). Part (iii) is a direct consequence of parts (i) and (ii). ■

We must now introduce two additional lemmas to prove Theorem 3.

Lemma A.8. *Under Assumption 3, the sequence $\{a_j\}_{j=l}^{m}$ satisfies (5.15) of Robinson (1995).*

Proof: The first two parts of (5.15) follow directly from Lemma A.1 and the third part follows from nearly identical expressions to those leading to (A18) of HDB. ■

Lemma A.9. *Let $\chi_j = a_j U_j/m^{1/2}$ and $\chi_j^v = a_j U_j^v/m^{1/2}$ where*

$$\begin{aligned}
U_j =& \varepsilon_j + \log\left[\frac{f^*\left(\lambda_j\right)}{f^*(0)}\right] - 2d\log\left[\frac{\left|1-\exp\left(-i\lambda_j\right)\right|}{\lambda_j}\right], \\
U_j^v =& \varepsilon_j^v + \log\left[\frac{f^*\left(\lambda_j\right)}{f^*(0)}\right] - 2d\log\left[\frac{\left|1-\exp\left(-i\lambda_j\right)\right|}{\lambda_j}\right]
\end{aligned}$$

and $\varepsilon_j^v \equiv \log(I^v(\lambda_j)/f_j) - E[\log(I^v(\lambda_j)/f_j)]$. For any nonnegative integer N, under Assumptions 1-3,

$$E\left[\left(\sum_{k=l}^{m}\chi_k\right)^N\right] = E\left[\left(\sum_{k=l}^{m}\chi_k^v\right)^N\right] + O\left(\frac{T^{3/2-3d}}{l^{3-3d}}\right).$$

Proof: Very similar (but much more cumbersome) arguments to those made in Lemmas A.3 and A.4 (using a multivariate Edgeworth expansion) provide the result. ■

Proof of Theorem 3: Begin by noting that

$$m^{1/2}\left(\hat{d}-d\right)=-\frac{m^{1/2}}{2S_{YY}}\sum_{j=l}^{m}a_j\log f_j^*-\frac{m}{2S_{YY}}\frac{1}{m^{1/2}}\sum_{j=l}^{m}a_j\varepsilon_j.$$

By Lemma A.2,

$$-\frac{m^{1/2}}{2S_{YY}}\sum_{j=l}^{m}a_j\log f_j^*=o(1)$$

since $m=o(T^{4/5})$. Let U_j be defined as in Lemma A.9. Then we have the following decomposition:

$$\begin{aligned}\frac{1}{m^{1/2}}\sum_{j=l}^{m}a_j\varepsilon_j=&\frac{1}{m^{1/2}}\sum_{j=l}^{m}a_jU_j-\frac{1}{m^{1/2}}\sum_{j=l}^{m}a_j\log\left\{\frac{f^*(\lambda_j)}{f^*(0)}\right\}+\frac{2d}{m^{1/2}}\sum_{j=l}^{m}a_j\log\left\{\frac{|1-\exp(-i\lambda_j)|}{\lambda_j}\right\}\\ \equiv&T_1+T_2+T_3.\end{aligned}$$

Now,

$$T_2=-\frac{1}{m^{1/2}}\sum_{j=l}^{m}a_j\log f^*(\lambda_j)+\frac{f^*(0)}{m^{1/2}}\sum_{j=l}^{m}a_j,$$

which is $o(1)$ since the first term is $o(1)$ by Lemmas A.1 and A.2 and the second term is zero. As shown by HDB (p. 44),

$$\log\left\{\frac{|1-\exp(-i\lambda_j)|}{\lambda_j}\right\}=O\left(\frac{m^2}{T^2}\right)$$

uniformly in $1\le j\le m$. Thus,

$$\begin{aligned}T_3^2\le&\frac{4d^2}{m}\sum_{j=1}^{m}a_j^2\sum_{j=1}^{m}\left[\log\left\{\frac{|1-\exp(-i\lambda_j)|}{\lambda_j}\right\}\right]^2\\ =&O(1/m)O(m)O\left(m^5/T^4\right)=o(1)\end{aligned}$$

by Lemma 1 of Hurvich and Beltrao (1994) since $m=o\left(T^{4/5}\right)$.

Finally note that the N^{th} moment of T_1 is precisely $E[(\sum_{k=l}^{m}\chi_k)^N]$ of Lemma A.9. Hence, Lemma A.9, Assumption 3 and the results of Robinson (1995, pp. 1067-1070) provide that the moments of T_1 converge to the corresponding moments of a variate that converges in distribution to $N(0,\pi^2/6)$. Using the same method of moments argument used by Robinson (1995, pp. 1067-1070), this provides that

$$T_1\xrightarrow{d}N\left(0,\pi^2/6\right).$$

In summary,

$$
\begin{aligned}
m^{1/2}\left(\hat{d}-d\right) =& o(1) - \frac{m}{2S_{xx}}\left(T_1+T_2+T_3\right) \\
=& o(1) - \frac{m}{2[m+o(m)]}\left[T_1+o(1)\right] \xrightarrow{d} N\left(0, \pi^2/24\right). \quad \blacksquare
\end{aligned}
$$

References

Andrews, D., Guggenberger, P., 2003. A bias-reduced log-periodogram regression estimator for the long-memory parameter. Econometrica 71, 675–712.

Beran, J., 1994. Statistics for Long-Memory Processes (Monographs on Statistics and Applied Probability). London: Chapman and Hall.

Bhattacharya, R., Gupta, V., Waymire, E., 1983. The Hurst effect under trends. Journal of Applied Probability 20, 649–662.

Chen, C., Tiao, G., 1990. Random level-shift time series models, ARIMA approximations and level-shift detection. Journal of Business and Economic Statistics 8, 83–97.

Dahlhaus, R., 1989. Efficient parameter estimation for self similar processes. The Annals of Statistics 17, 1749–1766.

Diebold, F., Inoue, A., 2001. Long memory and regime switching. Journal of Econometrics 105, 131–159.

Dolado, J., Gonzalo, J., Mayoral, L., 2005. What is what?: a simple test of long-memory versus structural breaks in the time domain, Unpublished Manuscript, Department of Economics, Universidad Carlos III de Madrid and Department of Economics Universidat Pompeu Fabra.

Fox, R., Taqqu, M., 1986. Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. The Annals of Statistics 14, 517–532.

Georgiev, I., 2002. Functional weak limit theory for rare outlying events, Unpublished Manuscript, European University Institute.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221–238.

Granger, C., Joyeux, R., 1980. An introduction to long memory time series models and fractional differencing. Journal of Time Series Analysis 1, 15–29.

Granger, C. W. J., Hyung, N., 2004. Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. Journal of Empirical Finance 11, 399–421.

Hosking, J., 1981. Fractional differencing. Biometrika 68, 165–176.

Hurst, H., 1951. Long-term storage capacity of reservoirs. Transactions of the American Society of Civil Engineers 116, 770–799.

Hurvich, C., Beltrao, K., 1994. Automatic semiparametric estimation of the memory parameter of a long-memory time series. Journal of Time Series Analysis 15, 285–302.

Hurvich, C., Deo, R., Brodsky, J., 1998. The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. Journal of Time Series Analysis 19, 19–46.

Hurvich, C., Lang, G., Soulier, P., 2005. Estimation of long memory in the presence of a smooth nonparametric trend. Journal of the American Statistical Association 100, 853–871.

Iacone, F., 2010. Local Whittle estimation of the memory parameter in presence of deterministic components. Journal of Time Series Analysis 31, 37–49.

Ikeda, S., 2009. Two scale realized kernels: A univariate case, Unpublished Manuscript, Department of Economics, Boston University.

Künsch, H., 1986. Discriminating between monotonic trends and long-range dependence. Journal of Applied Probability 23, 1025–1030.

Künsch, H., 1987. Statistical aspects of self-similar processes. In: Prohorov, Y., Sazarov, V. (Eds.), Proceedings of the First World Congress of the Bernoulli Society. Vol. 1. VNU Science Press, Utrecht, pp. 67–74.

Lu, Y., Perron, P., 2010. Modeling and forecasting stock return volatility using a random level shift model. Journal of Empirical Finance 17, 138–156.

Mikosch, T., Stărică, C., 2004. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. Review of Economics and Statistics 86, 378–390.

Ohanissian, A., Russell, J., Tsay, R., 2004. True or spurious long memory in volatility: Does it matter for pricing options?, Unpublished Manuscript, Booth School of Business, University of Chicago.

Ohanissian, A., Russell, J., Tsay, R., 2008. True or spurious long memory? a new test. Journal of Business and Economic Statistics 26, 161–175.

Perron, P., 1989. The great crash, the oil price shock, and the unit root hypothesis. Econometrica 57, 1361–1401.

Perron, P., Qu, Z., 2010. Long-memory and level shifts in the volatility of stock market return indices. Journal of Business and Economic Statistics 28, 275–290.

Qu, Z., 2008. A test against spurious long memory, Unpublished Manuscript, Department of Economics, Boston University.

Robinson, P., 1995. Log-periodogram regression of time series with long range dependence. The Annals of Statistics 23, 1048–1072.

Robinson, P., 1997. Large-sample inference for nonparametric regression with dependent errors. The Annals of Statistics 25, 2054–2083.

Shimotsu, K., 2006. Simple (but effective) tests of long memory versus structural breaks, Working Paper No. 1101, Department of Economics, Queen's University.

Skovgaard, I., 1986. On multivariate Edgeworth expansions. International Statistical Review 54, 169–186.

Smith, A., 2005. Level shifts and the illusion of long memory in economic time series. Journal of Business and Economic Statistics 23, 321–335.

Sun, Y., Phillips, P., 2003. Nonlinear log-periodogram regression for perturbed fractional processes. Journal of Econometrics 115, 355–389.

Taylor, S., 2000. Consequences for option pricing of a long memory in volatility, Unpublished Manuscript, Department of Accounting and Finance, Lancaster University.

Velasco, C., 1999. Non-stationary log-periodogram regression. Journal of Econometrics 91, 325–371.

Velasco, C., 2000. Non-Gaussian log periodogram regression. Econometric Theory 16, 44–79.

Table 1: Bias for Short-Memory Processes with Random Level Shifts

	Standard; u				Trimmed; ε, u				Adaptive; ε, u			
T	0.5	0.7	0.8	0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9
						$p = 5$						
500	0.656	0.362	0.254	0.193	0.063	0.032	0.025	0.024	0.065	0.027	0.002	0.001
1000	0.658	0.335	0.221	0.155	0.044	0.021	0.015	0.016	0.057	0.020	0.000	0.005
2000	0.685	0.323	0.200	0.129	0.045	0.015	0.010	0.004	0.068	0.016	0.004	0.001
						$p = 10$						
500	0.778	0.471	0.338	0.257	0.127	0.052	0.037	0.025	0.160	0.075	0.039	0.017
1000	0.792	0.449	0.306	0.219	0.095	0.053	0.039	0.022	0.163	0.058	0.027	0.017
2000	0.801	0.430	0.277	0.182	0.098	0.040	0.027	0.008	0.168	0.045	0.022	0.007
						$p = 20$						
500	0.866	0.584	0.438	0.343	0.208	0.113	0.087	0.059	0.338	0.189	0.124	0.055
1000	0.875	0.555	0.391	0.285	0.173	0.083	0.063	0.040	0.343	0.146	0.088	0.038
2000	0.885	0.526	0.349	0.236	0.158	0.064	0.046	0.020	0.339	0.117	0.070	0.020
						S&P 500						
500	0.836	0.541	0.399	0.310	0.163	0.087	0.067	0.047	0.269	0.130	0.071	0.035
1000	0.850	0.519	0.362	0.262	0.157	0.071	0.050	0.028	0.294	0.117	0.070	0.026
2000	0.863	0.493	0.324	0.218	0.143	0.060	0.039	0.016	0.288	0.098	0.057	0.015
						NASDAQ						
500	0.859	0.597	0.454	0.359	0.240	0.133	0.104	0.068	0.389	0.234	0.167	0.080
1000	0.863	0.559	0.398	0.293	0.194	0.097	0.074	0.044	0.377	0.180	0.126	0.050
2000	0.881	0.539	0.362	0.247	0.179	0.078	0.057	0.024	0.384	0.152	0.104	0.027

Table 2: RMSE for Short-Memory Processes with Random Level Shifts

	Standard; u				Trimmed; ε, u				Adaptive; ε, u			
T	0.5	0.7	0.8	0.9	0.01, 0.7	0.0.5,0.8	0.1,0.8	0.15,0.9	0.01, 0.7	0.0.5,0.8	0.1,0.8	0.15,0.9
						$p = 5$						
500	0.699	0.394	0.281	0.215	0.274	0.160	0.203	0.166	0.289	0.161	0.179	0.137
1000	0.697	0.365	0.243	0.173	0.198	0.106	0.138	0.097	0.232	0.107	0.119	0.084
2000	0.714	0.348	0.218	0.143	0.150	0.072	0.089	0.057	0.204	0.076	0.081	0.053
						$p = 10$						
500	0.802	0.491	0.355	0.273	0.295	0.172	0.206	0.173	0.370	0.211	0.213	0.147
1000	0.810	0.467	0.323	0.232	0.226	0.119	0.139	0.095	0.331	0.151	0.141	0.090
2000	0.816	0.447	0.292	0.193	0.179	0.087	0.093	0.058	0.299	0.112	0.103	0.055
						$p = 20$						
500	0.883	0.600	0.453	0.358	0.356	0.201	0.225	0.179	0.500	0.316	0.290	0.182
1000	0.888	0.568	0.403	0.297	0.275	0.145	0.154	0.105	0.471	0.247	0.211	0.110
2000	0.894	0.538	0.360	0.245	0.227	0.105	0.104	0.062	0.443	0.200	0.162	0.061
						S&P 500						
500	0.854	0.557	0.414	0.324	0.330	0.188	0.221	0.170	0.452	0.260	0.240	0.157
1000	0.864	0.535	0.376	0.275	0.262	0.135	0.145	0.097	0.431	0.219	0.194	0.103
2000	0.872	0.507	0.337	0.228	0.213	0.101	0.102	0.060	0.400	0.177	0.150	0.060
						NASDAQ						
500	0.884	0.621	0.477	0.381	0.381	0.226	0.241	0.187	0.540	0.365	0.338	0.221
1000	0.883	0.581	0.419	0.311	0.300	0.161	0.166	0.109	0.506	0.291	0.261	0.132
2000	0.893	0.557	0.380	0.261	0.257	0.121	0.114	0.066	0.489	0.244	0.212	0.073

Table 3: Bias for Long-Memory Processes with Random Level Shifts

	Standard; u				Trimmed; ε, u				Adaptive; ε, u			
T	0.5	0.7	0.8	0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9
						$d = 0.2, p = 5$						
500	0.404	0.214	0.150	0.115	0.035	0.024	0.019	0.021	0.027	0.025	-0.007	-0.002
1000	0.384	0.185	0.121	0.086	0.020	0.012	0.010	0.013	0.066	0.040	0.018	0.011
2000	0.387	0.169	0.103	0.067	0.019	0.007	0.006	0.003	0.064	0.027	0.013	0.007
						$d = 0.2, p = 10$						
500	0.507	0.295	0.212	0.164	0.058	0.041	0.029	0.037	0.141	0.087	0.041	0.018
1000	0.513	0.275	0.188	0.135	0.064	0.034	0.027	0.021	0.146	0.076	0.049	0.027
2000	0.506	0.248	0.157	0.104	0.058	0.022	0.015	0.006	0.144	0.063	0.045	0.018
						$d = 0.2, p = 20$						
500	0.606	0.392	0.294	0.232	0.135	0.084	0.067	0.052	0.254	0.255	0.249	0.254
1000	0.605	0.355	0.248	0.182	0.096	0.054	0.045	0.034	0.191	0.148	0.117	0.191
2000	0.601	0.323	0.211	0.142	0.081	0.035	0.028	0.015	0.141	0.115	0.089	0.141
						$d = 0.45, p = 5$						
500	0.139	0.077	0.058	0.045	0.013	0.020	0.016	0.015	0.023	0.045	0.035	0.029
1000	0.133	0.067	0.047	0.034	0.017	0.013	0.011	0.007	0.049	0.037	0.034	0.022
2000	0.114	0.051	0.033	0.022	0.015	0.008	0.009	0.002	0.043	0.027	0.025	0.014
						$d = 0.45, p = 10$						
500	0.223	0.133	0.098	0.079	0.047	0.025	0.021	0.019	0.094	0.085	0.075	0.062
1000	0.200	0.107	0.072	0.053	0.041	0.013	0.005	0.011	0.094	0.063	0.059	0.041
2000	0.180	0.082	0.053	0.036	0.026	0.012	0.007	0.005	0.074	0.046	0.043	0.026
						$d = 0.45, p = 20$						
500	0.301	0.191	0.148	0.118	0.085	0.055	0.052	0.030	0.157	0.140	0.134	0.105
1000	0.284	0.158	0.113	0.086	0.047	0.035	0.029	0.025	0.150	0.108	0.105	0.076
2000	0.256	0.126	0.085	0.059	0.036	0.025	0.021	0.012	0.122	0.080	0.077	0.049

Table 4: RMSE for Long-Memory Processes with Random Level Shifts

	Standard; u				Trimmed; ε, u				Adaptive; ε, u			
T	0.5	0.7	0.8	0.9	0.01, 0.7	0.0.5,0.8	0.1,0.8	0.15,0.9	0.01, 0.7	0.0.5,0.8	0.1,0.8	0.15,0.9
						$d = 0.2, p = 5$						
500	0.462	0.248	0.178	0.139	0.270	0.157	0.202	0.165	0.257	0.149	0.171	0.125
1000	0.434	0.215	0.144	0.104	0.192	0.105	0.138	0.096	0.206	0.110	0.117	0.070
2000	0.427	0.194	0.122	0.080	0.138	0.069	0.088	0.058	0.150	0.069	0.071	0.041
						$d = 0.2, p = 10$						
500	0.546	0.323	0.235	0.183	0.284	0.158	0.204	0.166	0.313	0.197	0.211	0.142
1000	0.542	0.298	0.206	0.150	0.199	0.111	0.135	0.096	0.260	0.144	0.144	0.080
2000	0.529	0.266	0.172	0.115	0.148	0.075	0.089	0.058	0.223	0.107	0.100	0.052
						$d = 0.2, p = 20$						
500	0.633	0.411	0.311	0.249	0.313	0.181	0.215	0.176	0.392	0.273	0.271	0.192
1000	0.624	0.372	0.263	0.195	0.224	0.124	0.146	0.102	0.349	0.213	0.200	0.115
2000	0.616	0.337	0.223	0.152	0.168	0.084	0.095	0.060	0.308	0.164	0.151	0.070
						$d = 0.45, p = 5$						
500	0.235	0.126	0.094	0.076	0.281	0.157	0.207	0.163	0.193	0.102	0.115	0.088
1000	0.208	0.103	0.073	0.055	0.186	0.105	0.137	0.092	0.121	0.075	0.076	0.056
2000	0.177	0.081	0.053	0.039	0.136	0.070	0.089	0.059	0.082	0.052	0.053	0.037
						$d = 0.45, p = 10$						
500	0.290	0.172	0.128	0.105	0.273	0.162	0.209	0.166	0.206	0.134	0.145	0.110
1000	0.256	0.136	0.097	0.072	0.197	0.107	0.138	0.094	0.145	0.097	0.098	0.071
2000	0.227	0.106	0.072	0.051	0.135	0.073	0.094	0.058	0.107	0.071	0.070	0.048
						$d = 0.45, p = 20$						
500	0.350	0.219	0.171	0.140	0.292	0.166	0.208	0.173	0.244	0.172	0.176	0.142
1000	0.321	0.180	0.131	0.100	0.199	0.108	0.130	0.097	0.183	0.130	0.130	0.098
2000	0.288	0.146	0.101	0.072	0.140	0.075	0.088	0.059	0.146	0.100	0.099	0.068

Table 5: Bias for Short/Long-Memory Processes with Deterministic Trends

	Standard; u				Trimmed; ε, u				Adaptive; ε, u			
T	0.5	0.7	0.8	0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9
					Monotone Trend, $d = 0$							
500	0.435	0.229	0.158	0.119	0.035	0.016	0.016	0.017	-0.007	0.002	-0.020	-0.007
1000	0.498	0.245	0.162	0.113	0.032	0.017	0.014	0.011	0.010	0.010	-0.002	0.001
2000	0.543	0.252	0.158	0.103	0.031	0.018	0.012	0.007	0.020	0.014	0.005	0.004
					Monotone Trend, $d = 0.2$							
500	0.177	0.096	0.064	0.050	0.015	-0.005	-0.013	0.008	-0.038	-0.017	-0.047	-0.022
1000	0.202	0.100	0.065	0.048	0.032	0.005	0.000	0.012	0.005	0.006	-0.008	0.004
2000	0.223	0.100	0.062	0.040	0.017	0.007	0.006	0.002	0.019	0.012	0.004	0.003
					Monotone Trend, $d = 0.45$							
500	0.043	0.029	0.021	0.021	0.023	0.005	0.002	0.017	-0.005	0.011	-0.001	0.010
1000	0.036	0.018	0.016	0.014	0.006	0.015	0.015	0.011	0.003	0.012	0.010	0.010
2000	0.025	0.016	0.013	0.009	0.007	0.009	0.009	0.000	0.011	0.011	0.010	0.007
					Seasonal Trend, $d = 0$							
500	0.683	0.267	0.169	0.118	0.009	0.004	0.004	0.000	-0.036	-0.014	-0.027	-0.022
1000	0.634	0.219	0.127	0.083	0.007	-0.005	-0.010	0.000	-0.020	-0.012	-0.025	-0.009
2000	0.580	0.176	0.097	0.058	0.002	-0.001	-0.003	-0.002	-0.014	-0.005	-0.011	-0.005
					Seasonal Trend, $d = 0.2$							
500	0.409	0.157	0.100	0.073	0.000	0.004	0.003	0.013	-0.054	-0.021	-0.043	-0.021
1000	0.370	0.125	0.073	0.049	0.005	0.001	-0.002	0.001	-0.032	-0.009	-0.020	-0.006
2000	0.329	0.097	0.054	0.032	-0.008	0.001	0.004	0.002	-0.025	-0.005	-0.008	-0.003
					Seasonal Trend, $d = 0.45$							
500	0.156	0.058	0.037	0.028	-0.005	0.002	0.002	0.014	-0.014	0.009	-0.004	0.001
1000	0.122	0.041	0.025	0.017	0.000	0.003	0.003	0.004	0.009	0.007	0.002	0.002
2000	0.094	0.031	0.018	0.011	0.003	-0.001	0.001	0.000	0.012	0.005	0.004	0.002

Table 6: RMSE for Short/Long-Memory Processes with Deterministic Trends

	Standard; u				Trimmed; ε, u				Adaptive; ε, u			
T	0.5	0.7	0.8	0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9
					Monotone Trend, $d = 0$							
500	0.452	0.239	0.166	0.127	0.273	0.153	0.208	0.163	0.220	0.134	0.167	0.136
1000	0.507	0.249	0.166	0.117	0.189	0.105	0.132	0.091	0.162	0.094	0.111	0.078
2000	0.548	0.255	0.161	0.105	0.137	0.072	0.089	0.056	0.122	0.066	0.078	0.051
					Monotone Trend, $d = 0.2$							
500	0.238	0.124	0.086	0.070	0.271	0.155	0.208	0.164	0.211	0.127	0.169	0.121
1000	0.233	0.117	0.078	0.059	0.195	0.102	0.130	0.091	0.145	0.081	0.100	0.061
2000	0.242	0.110	0.069	0.047	0.129	0.067	0.086	0.055	0.098	0.052	0.061	0.039
					Monotone Trend, $d = 0.45$							
500	0.172	0.089	0.066	0.057	0.259	0.152	0.199	0.167	0.152	0.080	0.109	0.072
1000	0.143	0.068	0.051	0.040	0.183	0.108	0.137	0.087	0.097	0.054	0.057	0.043
2000	0.120	0.055	0.039	0.029	0.136	0.072	0.090	0.057	0.062	0.041	0.042	0.030
					Seasonal Trend, $d = 0$							
500	0.693	0.274	0.177	0.126	0.269	0.155	0.204	0.164	0.220	0.135	0.171	0.138
1000	0.642	0.225	0.132	0.088	0.187	0.101	0.131	0.092	0.155	0.093	0.114	0.082
2000	0.585	0.181	0.102	0.063	0.124	0.068	0.089	0.057	0.106	0.063	0.080	0.053
					Seasonal Trend, $d = 0.2$							
500	0.429	0.172	0.113	0.085	0.262	0.151	0.204	0.166	0.205	0.118	0.156	0.119
1000	0.384	0.136	0.084	0.058	0.180	0.101	0.131	0.091	0.133	0.075	0.098	0.063
2000	0.341	0.106	0.062	0.040	0.137	0.069	0.088	0.056	0.105	0.050	0.059	0.038
					Seasonal Trend, $d = 0.45$							
500	0.226	0.097	0.069	0.055	0.269	0.160	0.208	0.166	0.188	0.091	0.114	0.077
1000	0.177	0.074	0.051	0.039	0.187	0.103	0.132	0.090	0.104	0.059	0.064	0.043
2000	0.143	0.057	0.040	0.030	0.133	0.072	0.089	0.057	0.066	0.042	0.042	0.031

Table 7: Bias and RMSE for Uncontaminated Processes

	Standard; u				Trimmed; ε, u				Adaptive; ε, u			
T	0.5	0.7	0.8	0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9
Bias						$d = 0$						
500	0.009	-0.004	-0.003	0.001	-0.011	-0.002	0.005	0.017	-0.056	-0.024	-0.032	-0.018
1000	-0.001	-0.001	-0.002	0.002	-0.013	-0.005	-0.003	0.013	-0.027	-0.012	-0.020	-0.008
2000	0.004	0.005	0.000	0.001	-0.001	-0.008	-0.006	0.001	-0.011	-0.006	-0.008	-0.007
						$d = 0.2$						
500	0.003	0.001	0.001	0.003	-0.013	0.000	-0.002	0.014	-0.070	-0.025	-0.046	-0.021
1000	-0.004	0.001	0.001	0.002	-0.003	-0.002	-0.003	0.007	-0.037	-0.010	-0.020	-0.006
2000	0.006	0.003	0.001	0.001	0.007	-0.001	-0.003	-0.002	-0.015	-0.004	-0.009	-0.003
						$d = 0.45$						
500	0.016	0.006	0.003	0.004	-0.004	-0.002	-0.004	0.010	-0.041	-0.010	-0.017	-0.006
1000	0.014	0.005	0.004	0.003	-0.002	0.003	0.005	0.003	-0.012	0.000	-0.002	0.000
2000	0.010	0.005	0.002	0.002	0.001	0.000	-0.002	0.000	0.001	0.000	0.000	0.000
RMSE						$d = 0$						
500	0.167	0.080	0.060	0.051	0.278	0.155	0.200	0.166	0.227	0.135	0.173	0.139
1000	0.132	0.060	0.042	0.033	0.189	0.100	0.135	0.095	0.152	0.093	0.115	0.079
2000	0.110	0.045	0.031	0.023	0.131	0.066	0.087	0.055	0.113	0.064	0.077	0.053
						$d = 0.2$						
500	0.163	0.084	0.060	0.049	0.271	0.153	0.204	0.161	0.213	0.120	0.155	0.115
1000	0.136	0.061	0.043	0.033	0.186	0.102	0.129	0.093	0.142	0.079	0.098	0.064
2000	0.107	0.046	0.032	0.024	0.130	0.070	0.087	0.055	0.093	0.050	0.062	0.039
						$d = 0.45$						
500	0.171	0.082	0.059	0.050	0.272	0.155	0.199	0.161	0.179	0.081	0.101	0.070
1000	0.138	0.065	0.047	0.037	0.191	0.102	0.137	0.092	0.102	0.051	0.056	0.041
2000	0.115	0.050	0.036	0.028	0.138	0.072	0.093	0.059	0.054	0.038	0.039	0.030

Table 8: Bias and RMSE for Long-Memory Processes Contaminated by Noise

	Standard; u				Trimmed; ε, u				Adaptive; ε, u			
T	0.5	0.7	0.8	0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9
Bias						$d = 0.2$						
500	-0.111	-0.132	-0.142	-0.145	-0.154	-0.160	-0.164	-0.156	-0.209	-0.181	-0.198	-0.181
1000	-0.105	-0.126	-0.137	-0.145	-0.150	-0.160	-0.160	-0.166	-0.179	-0.171	-0.182	-0.178
2000	-0.092	-0.123	-0.133	-0.143	-0.147	-0.150	-0.151	-0.159	-0.166	-0.156	-0.163	-0.164
						$d = 0.45$						
500	-0.121	-0.208	-0.253	-0.281	-0.309	-0.349	-0.353	-0.368	-0.352	-0.364	-0.387	-0.393
1000	-0.096	-0.187	-0.239	-0.275	-0.293	-0.336	-0.347	-0.368	-0.305	-0.336	-0.358	-0.375
2000	-0.070	-0.167	-0.224	-0.270	-0.261	-0.316	-0.329	-0.366	-0.253	-0.300	-0.322	-0.360
RMSE						$d = 0.2$						
500	0.202	0.154	0.154	0.153	0.316	0.225	0.264	0.224	0.300	0.225	0.259	0.224
1000	0.173	0.141	0.144	0.149	0.240	0.193	0.214	0.192	0.236	0.195	0.216	0.196
2000	0.145	0.132	0.137	0.144	0.197	0.166	0.175	0.168	0.198	0.169	0.179	0.171
						$d = 0.45$						
500	0.211	0.224	0.260	0.286	0.406	0.382	0.409	0.401	0.412	0.389	0.421	0.414
1000	0.168	0.197	0.243	0.277	0.347	0.351	0.370	0.381	0.343	0.349	0.376	0.384
2000	0.134	0.174	0.227	0.271	0.293	0.323	0.341	0.370	0.279	0.307	0.332	0.364

Table 9: Bias for ARFIMA Processes

	Standard; u				Trimmed; ε, u				Adaptive; ε, u			
T	0.5	0.7	0.8	0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9
					$a = 0.6, b = 0, d = 0$							
500	0.048	0.221	0.369	0.484	0.533	0.717	0.766	0.868	0.311	0.430	0.443	0.511
1000	0.019	0.165	0.322	0.463	0.414	0.633	0.696	0.833	0.257	0.389	0.402	0.493
2000	0.008	0.123	0.283	0.444	0.305	0.552	0.619	0.790	0.203	0.349	0.363	0.477
					$a = 0.6, b = 0, d = 0.45$							
500	0.063	0.228	0.370	0.482	0.518	0.703	0.753	0.851	0.229	0.370	0.370	0.482
1000	0.033	0.176	0.329	0.464	0.423	0.632	0.690	0.822	0.176	0.329	0.329	0.464
2000	0.018	0.130	0.287	0.447	0.312	0.553	0.617	0.790	0.130	0.287	0.287	0.447
					$a = 0, b = -0.6, d = 0.45$							
500	0.010	0.036	0.110	0.386	0.097	0.288	0.354	1.336	0.017	0.111	0.112	0.386
1000	0.015	0.026	0.079	0.340	0.058	0.184	0.224	1.042	0.019	0.080	0.080	0.340
2000	0.009	0.016	0.058	0.291	0.035	0.130	0.158	0.790	0.014	0.059	0.059	0.291
					$a = -0.6, b = 0, d = 0$							
500	-0.009	-0.030	-0.104	-0.382	-0.090	-0.284	-0.350	-1.342	-0.121	-0.284	-0.351	-1.342
1000	-0.003	-0.019	-0.075	-0.338	-0.049	-0.187	-0.234	-1.050	-0.068	-0.187	-0.235	-1.050
2000	-0.003	-0.014	-0.055	-0.290	-0.040	-0.126	-0.151	-0.790	-0.052	-0.126	-0.152	-0.790
					$a = 0.298, b = 0.751, d = 0.457$							
500	-0.092	-0.328	-0.405	-0.419	-0.614	-0.565	-0.540	-0.422	-0.636	-0.575	-0.567	-0.459
1000	-0.046	-0.271	-0.380	-0.410	-0.568	-0.584	-0.583	-0.463	-0.575	-0.585	-0.588	-0.478
2000	-0.022	-0.223	-0.356	-0.405	-0.490	-0.577	-0.593	-0.497	-0.487	-0.577	-0.593	-0.500

Table 10: RMSE for ARFIMA Processes

	Standard; u				Trimmed; ε, u				Adaptive; ε, u			
T	0.5	0.7	0.8	0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9
					$a = 0.6, b = 0, d = 0$							
500	0.177	0.235	0.373	0.487	0.597	0.734	0.793	0.882	0.330	0.432	0.445	0.512
1000	0.143	0.176	0.325	0.464	0.455	0.642	0.709	0.838	0.269	0.390	0.403	0.494
2000	0.107	0.131	0.285	0.445	0.332	0.557	0.626	0.792	0.213	0.350	0.364	0.478
					$a = 0.6, b = 0, d = 0.45$							
500	0.179	0.243	0.376	0.485	0.586	0.722	0.780	0.868	0.243	0.376	0.376	0.485
1000	0.143	0.187	0.332	0.466	0.464	0.641	0.704	0.829	0.187	0.332	0.332	0.466
2000	0.110	0.139	0.290	0.449	0.338	0.558	0.624	0.793	0.139	0.290	0.290	0.449
					$a = 0, b = -0.6, d = 0.45$							
500	0.172	0.091	0.127	0.390	0.292	0.329	0.409	1.347	0.138	0.127	0.127	0.390
1000	0.140	0.069	0.092	0.342	0.203	0.211	0.262	1.047	0.089	0.092	0.092	0.342
2000	0.110	0.054	0.069	0.293	0.137	0.148	0.182	0.793	0.057	0.069	0.069	0.293
					$a = -0.6, b = 0, d = 0$							
500	0.171	0.090	0.119	0.386	0.277	0.322	0.406	1.352	0.251	0.322	0.406	1.352
1000	0.134	0.065	0.086	0.340	0.187	0.212	0.268	1.054	0.169	0.212	0.268	1.054
2000	0.114	0.049	0.063	0.291	0.142	0.144	0.175	0.793	0.130	0.144	0.175	0.793
					$a = 0.298, b = 0.751, d = 0.457$							
500	0.199	0.339	0.409	0.421	0.671	0.587	0.579	0.451	0.680	0.592	0.593	0.474
1000	0.147	0.278	0.383	0.412	0.599	0.592	0.598	0.472	0.602	0.593	0.601	0.484
2000	0.118	0.228	0.357	0.405	0.508	0.581	0.599	0.500	0.506	0.581	0.599	0.502

Table 11: Bias and RMSE for t_5 Innovation Distributed Short/Long-Memory Processes

	Standard; u				Trimmed; ε, u				Adaptive; ε, u			
T	0.5	0.7	0.8	0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9	$0.01, 0.7$	0.0.5,0.8	0.1,0.8	0.15,0.9
Bias						$d = 0.45, p = 0$						
500	0.007	0.011	0.009	0.007	-0.004	0.003	0.010	0.002	-0.043	0.001	-0.009	-0.003
1000	0.011	0.008	0.005	0.004	0.000	-0.001	-0.003	0.000	-0.005	0.001	0.001	0.001
2000	0.009	0.006	0.004	0.003	0.004	0.002	0.002	0.000	0.002	0.003	0.002	0.001
						$d = 0, p = 10$						
500	0.705	0.396	0.279	0.210	0.061	0.034	0.027	0.020	0.071	0.028	0.000	-0.001
1000	0.726	0.384	0.256	0.179	0.066	0.032	0.025	0.010	0.090	0.034	0.015	0.003
2000	0.732	0.358	0.225	0.147	0.063	0.024	0.012	0.007	0.103	0.027	0.008	0.004
RMSE						$d = 0.45, p = 0$						
500	0.176	0.084	0.063	0.050	0.281	0.160	0.209	0.158	0.191	0.081	0.108	0.074
1000	0.139	0.065	0.046	0.036	0.190	0.103	0.136	0.092	0.095	0.051	0.054	0.041
2000	0.115	0.050	0.035	0.028	0.132	0.070	0.087	0.058	0.059	0.038	0.039	0.030
						$d = 0, p = 10$						
500	0.734	0.419	0.299	0.227	0.290	0.163	0.213	0.163	0.307	0.159	0.182	0.138
1000	0.749	0.404	0.272	0.193	0.210	0.108	0.130	0.092	0.255	0.121	0.126	0.084
2000	0.750	0.376	0.239	0.157	0.149	0.077	0.091	0.056	0.226	0.086	0.088	0.051

Figure 1: S&P 500

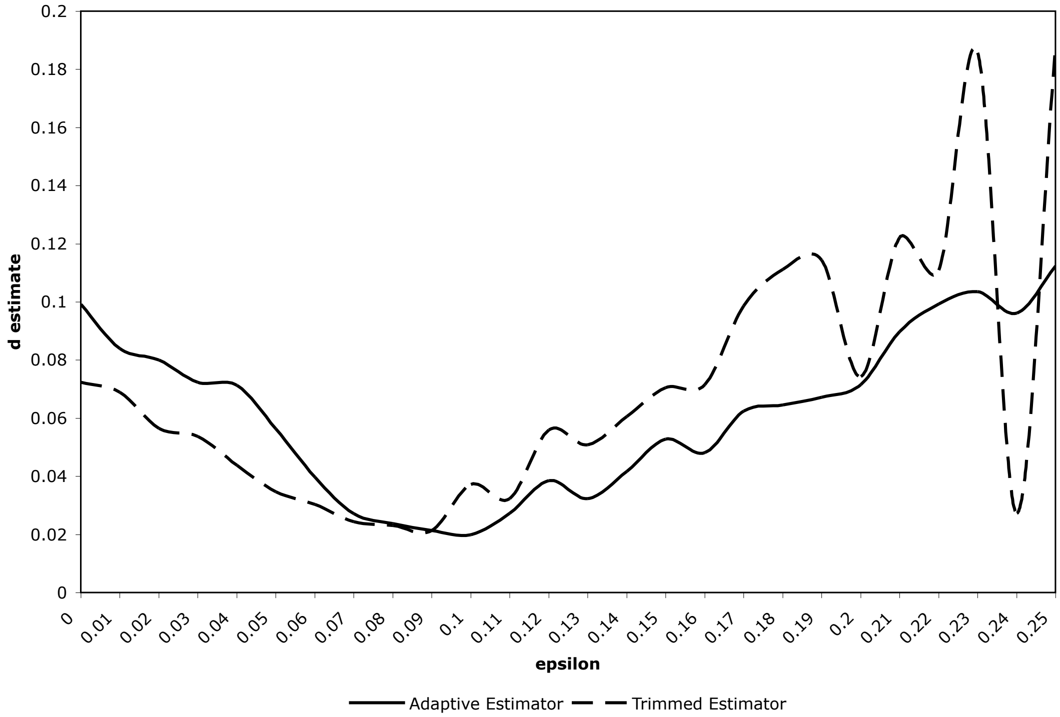

Figure 2: Dow Jones Industrial Average

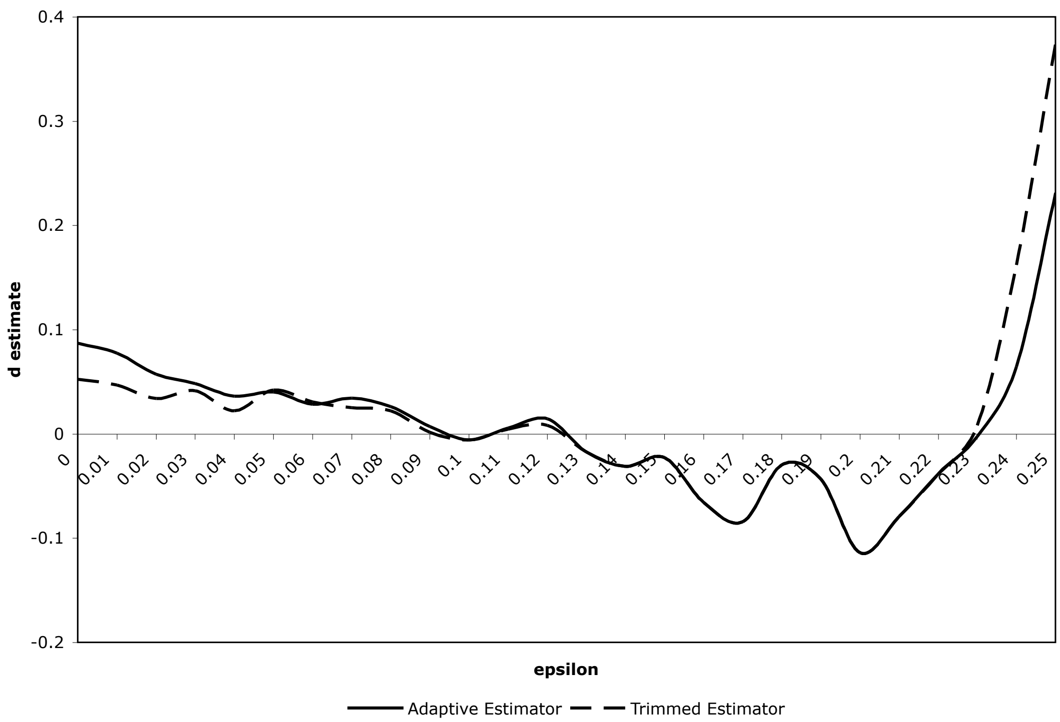

Figure 3: NASDAQ

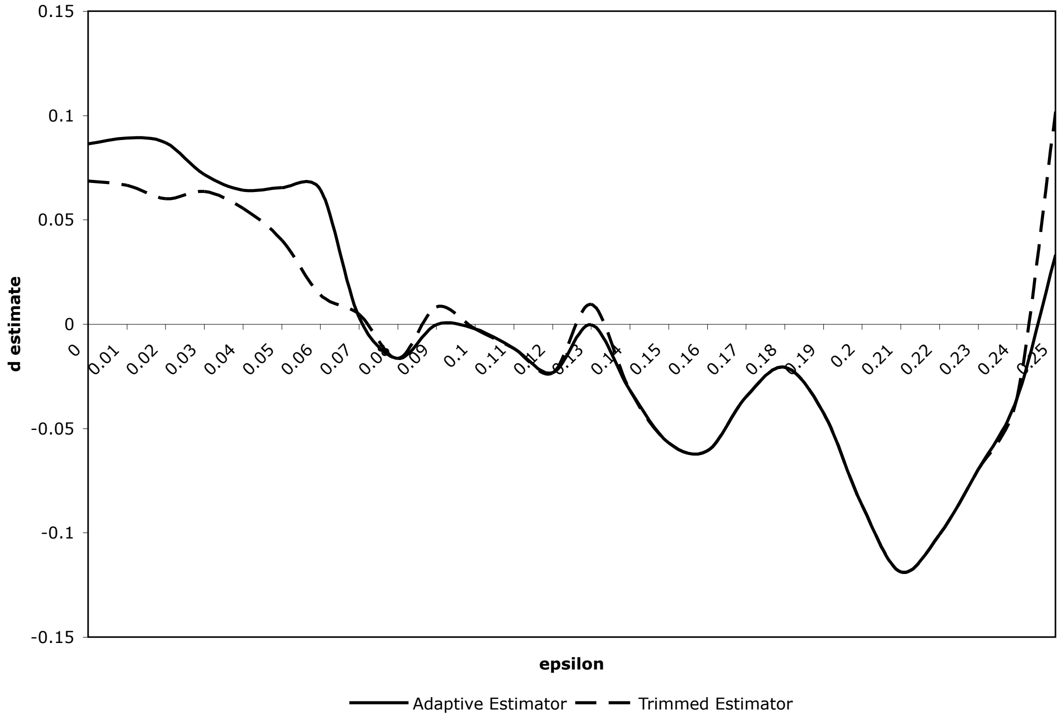

Figure 4: AMEX

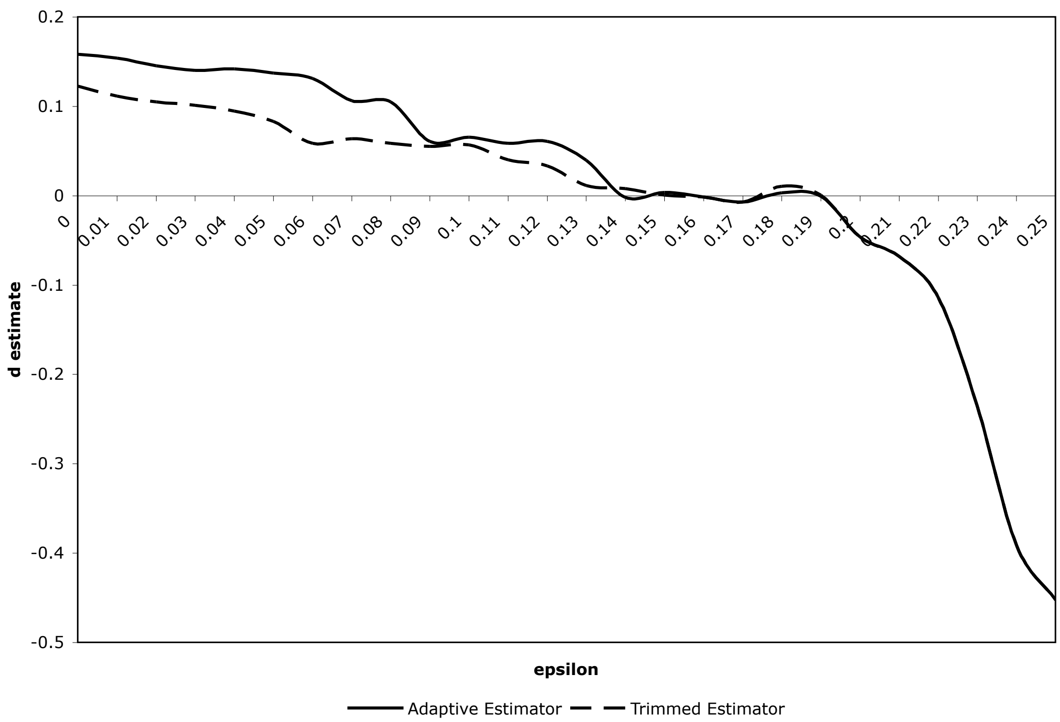

Figure 5: S&P 500 Realized Volatility

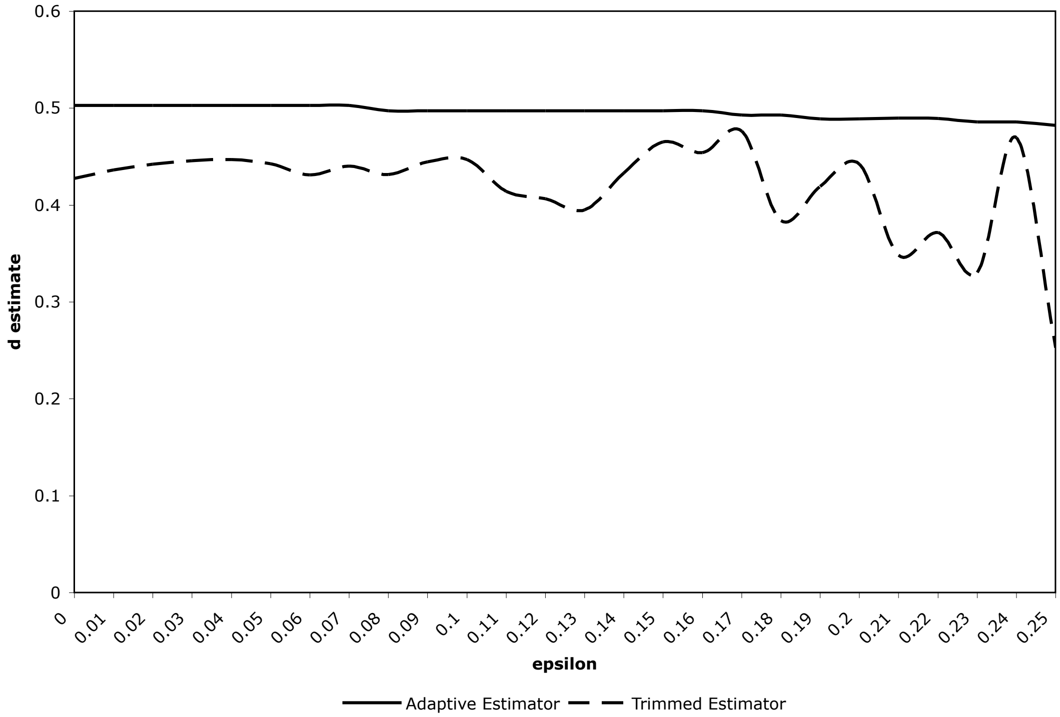

Figure 6: Nile River Level