Econometric Theory, **29**, 2013, 1196–1237.
doi:10.1017/S0266466613000042

MEMORY PARAMETER ESTIMATION IN THE PRESENCE OF LEVEL SHIFTS AND DETERMINISTIC TRENDS

ADAM MCCLOSKEY
Brown University

PIERRE PERRON
Boston University

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 that discard low-frequency components. We also show that a previously developed trimmed local Whittle estimator is robust to the same forms of data contamination. Regardless of whether the underlying long- or short-memory process is contaminated by level shifts or deterministic trends, the estimators are consistent and asymptotically normal with the same limiting variance as their standard untrimmed counterparts. Simulations show that the trimmed 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 trade-offs 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 adaptive version of the trimmed log-periodogram estimator that performs well in a wide variety of circumstances. We apply the estimators to stock market volatility data to find that various time series typically thought to be long-memory processes actually appear to be short- or very weak long-memory processes contaminated by level shifts or deterministic trends.

1. INTRODUCTION

There has long been interest in time series that are stationary yet exhibit persistence beyond that of short-memory or $I(0)$ variates, the so-called long-memory processes, starting in the 1950s 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

The authors are grateful to Shinsuke Ikeda for kindly sharing S&P 500 futures realized volatility data. We also thank Morten Nielsen, the co-editor Robert Taylor, and two referees for useful comments. Address correspondence to Adam McCloskey, Department of Economics, Brown University, Box B, 64 Waterman St., Providence, RI 02912, USA; e-mail: adam_mccloskey@brown.edu.

 © Cambridge University Press 2013

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

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 from the right, where d is the "memory parameter" of the process. It is stationary when $d \in \left(\frac{-1}{2}, \frac{1}{2}\right)$, 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" component. The most widely used are the log-periodogram (LP) estimator of Geweke and Porter-Hudak (1983) and the local Whittle (LW) estimator proposed by Künsch (1987).

The fact that the presence of level shifts or deterministic time trends affects the apparent persistence properties of time series has long been recognized. Perron (1989) showed that the presence of shifts in a time trend will often induce spurious nonrejection of the unit root hypothesis. Bhattacharya, Gupta, and Waymire (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 in addition to 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. Relatedly, Haldrup and Nielsen (2007) show via Monte Carlo simulation that when the mean of a time series exhibits random level shifts, both parametric and semiparametric estimates of the memory parameter are upward biased for a wide range of d values. These biases can indeed be quite large. 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, and Ohanissian, Russell, and Tsay (2004) have shown its considerable impact on the pricing of call options.

The potential presence of level shifts and deterministic trends in economic and financial time series is not merely a theoretical curiosity. For example, Granger and Hyung (2004), Mikosch and Stărică (2004), and Perron and Qu (2010) have reported strong evidence that these forms of data contamination are in fact a very real feature of stock market volatility data. Garcia and Perron (1996) find the presence of large level shifts in U.S. real interest rate series. Qu (2011) rejects the null hypothesis that a U.S. inflation rate series is a stationary short- or long-memory process in a test directed against the presence of level shifts or

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

deterministic trends. The preceding references represent a small subset of examples within a large and growing body of empirical evidence of this type of phenomena.

Presumably due to the empirical evidence, a handful of papers aimed at distinguishing true from spurious long memory has emerged in recent years. Tests in both the time and frequency domains have been proposed by Dolado, Gonzalo, and Mayoral (2005), Shimotsu (2006), Ohanissian, Russell, and Tsay (2008), Perron and Qu (2010), and Qu (2011). Many have argued that the long-memory properties of various economic time series are indeed spurious. However, scant attention has been paid to *estimation* of the memory parameter in the presence of contaminating elements.[1] 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, researchers have not adequately explored the implications of *long-memory* processes contaminated by these elements, especially level shifts. As we will show later (and as noted in a particular context by Haldrup and Nielsen, 2007), 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, providing further motivation for robust estimation. 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. It is based upon the assumption of a short-memory process contaminated by a stationary random level shift (RLS) process and is not necessarily valid when the process is contaminated long memory. Assuming Gaussianity, the standard LP estimator is already consistent under the data generating processes (DGPs) he considers (Hurvich, Deo, and Brodsky, 1998) (henceforth HDB) 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 nonstationary RLS processes and smooth or monotonic trends. In terms of RLS processes, those we consider are arguably more practically relevant as they imply periodograms that diverge in expectation at the zero frequency rather than flattening out and converging to some constant. These nonstationary RLS processes have received considerable attention in the econometrics literature (see, e.g., Chen and Tiao, 1990; Diebold and Inoue, 2001; Granger and Hyung, 2004; Perron and Qu, 2010).

The trimmed LP 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 and

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

also a trimming parameter to determine the lowest frequency.[2] 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. Their limiting variance 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 shifts or deterministic components, often nearly eliminating it entirely. On the other hand, trimming can increase the finite-sample variance so that in the assured absence of contamination, standard estimation is preferable. Nevertheless, when contamination is present, the bias reduction often significantly outweighs the variance inflation, yielding substantial mean squared error (MSE) reduction so that the trimmed estimators dominate their standard counterparts when contaminating elements are present. The finite-sample properties of the estimators depend upon their trimming and bandwidth parameters much as the standard LP estimator depends upon its bandwidth parameter. Through Monte Carlo simulation, we explore their finite-sample properties for different combinations of these parameters, providing practical suggestions for choosing them. Moreover, we provide the asymptotic MSE-minimizing choice of the bandwidth parameter that requires the trimming parameter to grow with the sample size in a certain unrestrictive manner.

After the work of this paper was completed, we became aware of related and complementary research by Iacone (2010), who also uses trimming but in the context of the LW estimator. The class of processes he considers is, however, somewhat more restrictive (with the exclusion of the "single impulse"). We add to his results by showing that the robustness of the trimmed LW estimators he considers extends to the very broad class of processes we consider when proper trimming is used. In conjunction with Iacone (2010), the results of this paper thus provide the practitioner with the choice of using either the trimmed LP or LW estimator in the potential presence of contaminating elements. There are benefits and drawbacks to both approaches, as discussed in some detail in Remark 6 and Section 5.4.

The structure of the paper is as follows. Section 2 introduces the new robust estimators. Section 3 describes the DGPs we consider, imposing specific assumptions; details a crucial asymptotic property of the periodogram of these processes; and extends the Iacone (2010) results on the trimmed LW estimator to these processes. Section 4 explores the asymptotic properties of the new trimmed LP estimators. Section 5 provides a Monte Carlo study of the finite-sample properties of the trimmed LP and LW estimators, in comparison with each other and their untrimmed counterparts under a variety of DGPs. We also show through simulations that the assumption of Gaussianity on the contaminated process can likely be relaxed. Section 6 focuses on empirical application of the new estimators to stock market volatility data. We find that some of the series that have been previously typified as long-memory processes appear actually to be contaminated short-memory processes with robust estimators of their memory parameters being

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

very near zero. Section 7 contains concluding remarks, and the proofs of our theoretical results are given in the Mathematical Appendix.

2. ROBUST MEMORY PARAMETER ESTIMATION

LP estimators are quite popular memory parameter estimators among empiricists because of their simplicity, intuitiveness, and ease of use. As mentioned, we examine "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 that implies $\log f(\lambda) \approx c - 2d \log \lambda$ as $\lambda \to 0_+$, where f is the spectral density function of the process. 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 is $I_x(\lambda_j) \equiv |w_x(\lambda_j)|^2 = w_x(\lambda_j) w_x(\lambda_j)^*$, where * denotes the complex conjugate value. The LP estimator replaces f in the preceding relation by I_x, evaluating it at frequencies local to zero to yield the LP regression:

$$\log I_x\left(\lambda_j\right) = c + d X_j + e_j, \qquad j = l, \ldots, m,$$

where $X_j = -\log(2 - 2\cos(\lambda_j)) \approx -\log \lambda_j^2$ for $j = l, \ldots, m$. If $m/T \to 0$, this approximation holds asymptotically. Thus, letting $I_j = I_x(\lambda_j)$, the LP estimator is

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

where $Y_j = \log\left|1 - \exp\left(-i\lambda_j\right)\right|$ and $\bar{Y} = (m - l + 1)^{-1} \sum_{k=l}^m Y_k$. The standard LP estimator uses $l = 1$, whereas we trim some of the lower frequencies to obtain consistency and asymptotic normality in the presence of level shifts and deterministic trends. To determine the number of frequencies to trim out, we rely on results about the order of the periodogram of a contaminated long-memory process. We assume the DGP of the series in question $\{x_t\}$ is

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

where k is a constant, $\{v_t\}$ is a mean-zero long- or short-memory process, and $\{u_t\}$ is a level shift process or a deterministic time trend (specified subsequently). Using Theorem 2 of Robinson (1995), an extension of arguments in Perron and Qu (2010) and Qu (2011), provides

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

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_t thus dominates the periodogram for frequencies λ_j such that $j = o(T^{(1-2d)/(2-2d)})$, and the component v_t dominates

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

when $jT^{(2d-1)/(2-2d)} \to \infty$. Although both a level shift (or deterministic trend) process and a long-memory process have poles in their periodograms at the zero frequency as T increases, they 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. These features of the periodogram allow one to distinguish between the processes $\{v_t\}$ and $\{u_t\}$ and also allow estimation of the memory parameter of the former when both are present. Intuitively, with an LP regression using only frequencies λ_j for which $jT^{(2d-1)/(2-2d)} \to \infty$ but nonetheless $j = o(T)$, one can expect a consistent estimate. Therefore, the motivation behind the estimator is to set l proportional to $T^{(1-2d)/(2-2d)+\varepsilon}$ for some $\varepsilon > 0$ to extract the memory behavior of the v_t component. Of course this suggestion is infeasible because d is unknown. However, for $d \in [0, \frac{1}{2})$, $(1-2d)/(2-2d) \in (0, \frac{1}{2}]$. Hence, one could obtain an estimate of d using only frequencies with j growing faster than $T^{1/2}$ but slower than T. For $d \in (\frac{-1}{2}, 0]$, $(1-2d)/(2-2d) \in [\frac{1}{2}, \frac{2}{3})$, so that considering frequencies with j growing faster than $T^{2/3}$ but slower than T would produce analogous results. This latter range is relevant to estimating the memory parameter of an underlying nonstationary long-memory process by examining its first differences. 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 observed process $\{x_t\}$ 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.

A1. The process $\{u_t\}$ is generated according to one of the following DGPs:

(a) Random level shifts (RLS): $u_t = \sum_{j=1}^{t} \delta_{T,j}$, $\delta_{T,t} = \pi_{T,t}\eta_t$, 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): $u_t = \sum_{i=1}^{B} c_i \mathbb{I}(T_{i-1} < t \leq T_i)$, where B is a fixed positive integer (the number of breaks plus one), $|c_i| < \infty$ for $i = 1, \ldots, B$, $\mathbb{I}(\cdot)$ is the indicator function, $T_0 = 0$, $T_B = T$, $T_0 < T_1 < \cdots < T_{B-1} < T_B$, and $T_i/T \to \tau_i \in (0,1)$ for $i = 1, \ldots, B$.

(c) Deterministic trends: $u_t = h(t/T)$, where $h(\cdot)$ is a deterministic nonconstant function on $[0,1]$ that is either Lipschitz continuous or monotone with $h(1) = 0$.[3]

(d) Fractional trends: $u_t = O((t+1)^{\phi-1/2})$ with $u_0 = 0$, $|u_{t+1} - u_t| = O(|u_t|/t)$, where $\phi \in (\frac{-1}{2}, \frac{1}{2})$.

It is important to note that the Bernoulli probability of A1(a) is sample-size-dependent; otherwise $\{u_t\}$ would be better construed as a random walk process.

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

This specification allows the average number of level shifts to remain constant (and equal to p) as the sample size grows. Note that p can be zero in A1 so that the assumption nests the no level shift, no trend case also. Perron and Qu (2010) considered the asymptotic properties of the periodogram of this type of process. Mikosch and Stărică (2004) and Iacone (2010) considered the asymptotic properties of the periodogram of the type of process given by A1(b) when $B = 2$ (one level shift). Künsch (1986) considered the asymptotic properties of the periodogram of a short-memory process contaminated by a bounded monotone trend (Lemma 2), and Qu (2011) extended his results to the Lipschitz continuous case (Lemma 1). Iacone (2010) also discussed the order of the periodogram of fractional trends. We now impose an assumption on the component whose memory parameter we are interested in estimating.

A2. The spectral density of $\{v_t\}$ is given by $f(\lambda) = |1 - \exp(-i\lambda)|^{-2d} f^*(\lambda)$, where $d \in (\frac{-1}{2}, \frac{1}{2})$ is the memory parameter, and $f^*(\cdot)$ 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, e.g., an autoregressive fractionally integrated moving average (ARFIMA) process. The following theorem is the counterpart to Theorem 2 of Robinson (1995) and also relies on his results. It characterizes the asymptotic behavior of the periodogram of $\{x_t\}$ given by (1) under A1 and A2. It is a key ingredient to proving consistency and asymptotic normality of the trimmed estimators in addition to being interesting in its own right.

THEOREM 1. *Suppose A1 and A2 hold. For any sequences of positive integers $j = j(T)$ and $k = k(T)$ such that $j > k$ and $j/T \to 0$ as $T \to \infty$:*

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

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

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

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

Theorem 1(i) implies that, for frequencies within a certain sample-size-dependent range, the periodogram is akin to an asymptotically unbiased estimator of the spectral density function in the sense that although the latter diverges at these frequencies, the former mimics this divergent behavior. Parts (ii)–(iv) describe the limiting covariance of the discrete Fourier transform, evaluated at frequencies within a certain range whose upper bound grows more slowly than the sample size. The following corollary provides that the broad class of level shift and deterministic trend contamination under scrutiny satisfies a high-level assumption of Iacone (2010). Together, with Theorem 2 (Theorem 3) of Iacone (2010), this provides the consistency (asymptotic normality) of the trimmed LW

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

estimator when the trimming assumption, Assumption 3 (Assumption 3′) of that paper is satisfied (along with the other relevant assumptions). The proof follows from results in the proof of Theorem 1 and is hence omitted.

COROLLARY 1. *Under A1,* $\{u_t\}$ *satisfies Assumption 2 of Iacone (2010) with* $\phi = \frac{1}{2}$.

Remark 1. Note that under A1(c) and (d), the bounds in Theorem 1 may not be exact and may overstate the asymptotic orders of the quantities, depending on the properties of the trend function $h(\cdot)$ or the value of ϕ (see the proof of Theorem 1 for details). However, we wish to impose minimal assumptions so that our results apply to a wide variety of trends. If one can impose more structure on $h(\cdot)$, it may be possible to use a smaller trimming.

Remark 2. For the processes studied, given by (1) and A1(a) or (b) (level shifts), the expectation of the periodogram can be decomposed as follows for large samples when λ_j is local to zero: $\mathbb{E}\left[I_x\left(\lambda_j\right)\right] \approx \lambda_j^{-2d} f^*\left(\lambda_j\right) + T^{-1}\lambda_j^{-2} g(\lambda_j) = \lambda_j^{-2d}\widetilde{f}(\lambda_j)$, where $\widetilde{f}(\lambda_j) \equiv f^*(\lambda_j) + T^{-1}\lambda_j^{2d-2} g(\lambda_j)$ and $g(\cdot)$ is a nonnegative even function bounded at zero. Taylor expansions yield

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

if $j = o(T)$. This motivates the nonlinear pseudo-regression:

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

The drawback to this approach is that, because the order $T^{-1}\lambda_j^{-2}$ is explicitly used in the formation of the "dependent variables" of the pseudo-regression, it may not work well for contamination of forms (c) and (d). This is because $O(T^{-1}\lambda_j^{-2})$ is a conservative bound for the periodogram of trend functions and may not well approximate its behavior local to the zero frequency, depending on the trend. Because the goal is to provide robustness to the wide variety of mean specifications given by A1, we do not pursue this approach here.

Remark 3. Another procedure to reduce 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 nonstationary components in the LP regression (see Velasco, 1999). Initial simulations show that using the robust estimators on tapered data (with the cosine bell taper) further reduces bias but increases finite-sample variance.

4. ASYMPTOTIC PROPERTIES OF THE ROBUST ESTIMATORS

As stated earlier, the lower trimming of the trimmed LP 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

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

two user-chosen parameters must grow also depend on the underlying memory parameter of the process $\{v_t\}$ (though one need not know the true value of d in practice). Assumption A3 makes these rates precise.

A3. As $T \to \infty$, $(m \log m)/T + (l \log^2 m)/m + (T^{1-2d} \log^4 m)/l^{2-2d} \to 0$.

The third term is the strongest part and is crucial to showing that the trimmed LP estimator has good asymptotic properties. Set $l = \max\{1, \lfloor K_l T^{\alpha} \rfloor\}$ and $m = \max\{l+1, \lfloor K_m T^{\beta} \rfloor\}$ for some $K_l, K_m \in (0, \infty)$ and $0 < \alpha < \beta < 1$. For estimation conducted on the levels of a stationary persistent long-memory process ($0 \le d < \frac{1}{2}$), these will satisfy A3 if $\alpha = \frac{1}{2} + \varepsilon$ for some $\varepsilon > 0$. For estimation on the first differences of a nonstationary long-memory process ($\frac{-1}{2} < d \le 0$), they will satisfy A3 if $\alpha = \frac{2}{3} + \varepsilon$. However, efficiency gains can be made by using an adaptive procedure that will be exploited in the following section. Finally, we impose more distributional structure on the processes we consider.

A4. $\{v_t\}$ is a Gaussian process.

We rely on this assumption because the existing literature on LP estimators of stationary processes does not cover the non-Gaussian case without necessitating "pooling" of observations across adjacent frequencies (see Velasco, 2000). The assumption of Gaussianity may appear strong, but simulation evidence presented in Section 5 indicates that it could be relaxed. There is also evidence that some economic and financial time series are (approximately) Gaussian, e.g., log-absolute returns (Anderson, Bollerslev, Diebold, and Labys, 2001). Furthermore, the presence of level shifts can induce a Gaussian series to appear non-Gaussian by, e.g., increasing the appearance of excess skewness; A4 applies to v_t, not the observed x_t. We now state results concerning the asymptotic bias and variance of the robust trimmed LP estimators. The following theorem parallels Theorem 1 of HDB.

THEOREM 2. *Under A1–A4,*

$$(i)\ \mathbb{E}[\hat{d} - d] = \frac{-2\pi^2}{9} \frac{f^{*\prime\prime}(0)}{f^*(0)} \frac{m^2}{T^2} + o\left(\frac{m^2}{T^2}\right) + O\left(\frac{\log^3 m}{m}\right) + O\left(\frac{T^{1-2d} \log^2 m}{m l^{1-2d}}\right), \tag{3}$$

$$(ii)\ \mathrm{Var}(\hat{d}) = \frac{\pi^2}{24m} + o\left(\frac{1}{m}\right) + O\left(\frac{T^{2-4d} \log^4 m \log l}{m^2 l^{2-4d}}\right) + O\left(\frac{T^{2-4d} \log^4 m}{m l^{4-4d}}\right) + O\left(\frac{T^{3-6d} \log^3 m}{m l^{5-6d}}\right), \tag{4}$$

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

$$(iii)\ \mathrm{MSE}(\hat{d}) = \frac{4\pi^4}{81}\left\{\frac{f^{*\prime\prime}(0)}{f^*(0)}\right\}^2 \frac{m^4}{T^4} + \frac{\pi^2}{24m} + O\left(\frac{m\log^2 m}{T^{1+2d}l^{1-2d}}\right) + O\left(\frac{T^{2-4d}\log^4 m\log l}{m^2 l^{2-4d}}\right) + O\left(\frac{T^{2-4d}\log^4 m}{ml^{4-4d}}\right) + O\left(\frac{T^{3-6d}\log^3 m}{ml^{5-6d}}\right) + o\left(\frac{m^4}{T^4}\right) + o\left(\frac{1}{m}\right). \quad \textbf{(5)}$$

The following corollary is a direct consequence of this theorem.

COROLLARY 2. *Under A1–A4, $\hat{d}$ is a consistent estimator of d.*

Remark 4. The expression for the MSE of $\hat{d}$ is interesting for what it implies about the asymptotically optimal bandwidth as this now depends upon the growth rate of the trimming parameter. Neglecting the remainder terms in the MSE (5) and minimizing with respect to m yields the same asymptotically optimal choice for m as in HDB:

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

However, it is no longer necessarily the case that the other remainder terms are asymptotically negligible in a neighborhood of this value of m. As before, set $l = \max\{1, \lfloor K_l T^{\alpha}\rfloor\}$ and $m = \max\{l+1, \lfloor K_m T^{\beta}\rfloor\}$, where $K_l, K_m \in (0,\infty)$ and $0 < \alpha < \beta < 1$. When $\beta = 4/5$, $O\left(m\log^2 m/(T^{1+2d}l^{1-2d})\right)$ and $O\left(T^{2-4d}\log^4 m/(m^2 l^{2-4d})\right)$ are asymptotically negligible compared to the first two terms of (iii) as long as $\alpha > (3-10d)/(5-10d)$, whereas $O\left(T^{2-4d}\log^4 m/(ml^{4-4d})\right)$ and $O\left(T^{3-6d}\log^3 m/(ml^{5-6d})\right)$ are asymptotically negligible compared to the first two terms of (iii) as long as $\alpha > (3-6d)/(5-6d)$. Hence, m^{OPT} is asymptotically optimal as long as α is larger than both $(3-10d)/(5-10d)$ and $(3-6d)/(5-6d)$. When estimating on the levels of persistent stationary data, $d \in [0, \frac{1}{2})$, so that if $\alpha > (3-6d)/(5-6d)$, m^{OPT} is asymptotically optimal, making the largest lower bound for α 3/5 and the smallest lower bound 0. When estimating on the first differences of nonstationary data, $d \in (\frac{-1}{2}, 0]$, so that if $\alpha > (3-10d)/(5-10d)$, m^{OPT} is asymptotically optimal. For this range, the largest lower bound is arbitrarily close to 4/5, and the smallest is 3/5. The asymptotically optimal bandwidth with these restrictions on α could provide guidance for simultaneously choosing the trimming and bandwidth parameters in large samples.

With the following stronger assumption on the trimming and bandwidth parameters, we can establish asymptotic normality of the estimators.

A3*. As $T \to \infty$, $m^5/T^4 + (l\log^2 m)/m + (T^{2-4d}\log^4 m)/(ml^{2-4d}) + (T^{3-6d}\log m)/l^{5-6d} \to 0$.

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

A feature of A3* is its connection to the optimal choice for the bandwidth in (6) and the lower bound required on the trimming for this bandwidth to be optimal. A3* enforces this lower bound when $d \in [0, \frac{1}{2})$ and makes the bandwidth grow at a slower rate than that of m^{OPT}. We now present the asymptotic distribution of the trimmed LP estimator.

THEOREM 3. *Under A1, A2, A3*, and A4,* $\sqrt{m}(\hat{d} - d) \xrightarrow{d} N(0, \pi^2/24)$.

Remark 5. Note that the limiting variance is the same as that for the standard LP estimator (see Robinson, 1995; HDB). Hence, we do not lose asymptotic efficiency by employing a trimming and bandwidth combination that satisfies A3*.

Remark 6. Corollaries 1 and 2 and Theorem 3 of this paper, along with Theorems 2 and 3 of Iacone (2010), provide one with a choice between trimmed LP and LW estimators of the memory parameter in the presence of level shifts and deterministic trends. Both types of estimation have their own benefits and drawbacks. The trimmed LP estimator is easier to use. The asymptotic results for the trimmed LP estimator impose Gaussianity, whereas those of the trimmed LW estimator do not. Simulation evidence indicates, however, that the LP estimators are robust to non-Gaussianity. Though the asymptotic variance of the LP estimator is 64% higher than that of the LW estimator, simulations indicate that the LP estimator is somewhat better at removing the bias arising from level shifts or trends in finite samples. In terms of MSE, the results are mixed. See Secion 5.4 for additional discussion.

Remark 7. Unreported simulations indicate that the finite-sample distribution of $\hat{d}$ is not well approximated by an $N(d, \pi^2/(24m))$. Approximations by an $N(d, \pi^2/(24S(l,m)))$, where $S(l,m) \equiv \sum_{j=l}^{m} v_j^2$ with $v_j = \log j - (m-l+1)^{-1} \sum_{k=l}^{m} \log k$, as suggested by Geweke and Porter-Hudak (1983), work much better. (Note that $S(l,m)/m \to 1$.) They deliver confidence intervals with broadly adequate coverage probabilities although, in the presence of strong level shifts or deterministic trends, larger samples ($T = 2{,}000+$) may be necessary.

We now briefly comment on how our results can lead to a test of the null hypothesis of a pure long-memory process vs. one contaminated by level shifts or deterministic trends. The idea is to look at the difference between trimmed LP estimators using different trimmings. This leads to a different type of test than those considered in Perron and Qu (2010), which are based upon the differences of standard LP estimators constructed at different bandwidths. The null hypothesis of no level shifts or deterministic trends can be written as $H_0 : u_t = 0$. Suppose $\hat{d}_1$ is the trimmed LP estimator using bandwidth m and trimming l_1 and $\hat{d}_2$ is the trimmed LP estimator using bandwidth m and trimming $l_2 > l_1$. Then under H_0, if $l_1(\log^2 l_2)/l_2 + l_2^5/T^4 \to 0$ and A2 and A4 hold, we have by (A.2) (in the

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

Appendix), Lemmas A.1 and A.2, and results in the proof of Theorem 3

$$H(l_1,l_2) \equiv \frac{\sqrt{24}m}{\pi\sqrt{l_2}}(\hat{d}_1-\hat{d}_2)$$
$$= -(1+o(1))\frac{\sqrt{24}}{2\pi}\frac{1}{\sqrt{l_2}}\left[\sum_{j=l_1}^{l_2} a_j \log f_j^* - \sum_{j=l_1}^{l_2} a_j \varepsilon_j\right] \xrightarrow{d} N(0,1). \quad (7)$$

To have high power one should set $l_1 = 1$ and l_2 to satisfy $T^{1-2d}(\log^4 m)/l_2^{2-2d} \to 0$ so that with level shifts or trends, the estimator using l_2 is consistent whereas that using l_1 is not. Ideally, one would consider the supremum over values of l_2 over some range, an extension outside the scope of this paper. We simply note that the resulting test would be similar to that of Qu (2011), namely, the LP analog of his statistic based upon the LW score. In the applications, we set $l_2 = 1$, $l_2 = [T^{0.51}]$, and $m = [T^{0.79}]$ to provide a test with high power.

5. FINITE-SAMPLE PROPERTIES OF ROBUST ESTIMATORS

We first introduce an adaptive procedure aimed at decreasing finite-sample variance. Focusing on the last term in A3, suppose that $\hat{d}_0$ is a consistent estimate of d. Then, letting $\hat{d}_1$ denote the trimmed LP estimate using the trimming $l = [K_l T^{(1-2\hat{d}_0)/(2-2\hat{d}_0)+\varepsilon}]$ for some $K_l > 0$ and some small $\varepsilon > 0$, $\hat{d}_1$ is consistent (asymptotically normal) by Theorem 2 (Theorem 3) if the remainder of A1–A4 (A1, A2, A3*, and A4) are satisfied. This adaptive procedure can be repeated from an initial estimate $\hat{d}_0$ using the trimming $l = [K_l T^{1/2+\varepsilon}]$. For $i \geq 1$, let $\hat{d}_i$ denote the LP estimate using the trimming $l = [K_l T^{(1-2\hat{d}_{i-1})/(2-2\hat{d}_{i-1})+\varepsilon}]$. Then $\hat{d}_i$ should satisfy Theorems 2 and 3 for any finite i. One can thus choose a convergence criterion that terminates this adaptive procedure, with the terminal value labeled as $\hat{d}^a$.[4] In the simulations, we use the criterion of $|\hat{d}_i - \hat{d}_{i-1}| < 0.01$ (convergence) or $i > 9$ (nonconvergence) to terminate the procedure and retain the final value of $\hat{d}$.[5] We concentrate on the region $[0, 1/2)$ for d because it is of most practical interest, and so we do not consider trimmings larger than $O(T^{1/2+\varepsilon})$.

We compare the finite-sample properties of the trimmed LP estimator using no adaptive procedure, labeled "trimmed," the adaptive trimmed estimator, and the standard LP estimator. Setting the trimming $l = [T^{1/2+\varepsilon}]$ for the trimmed estimator, $l = [T^{(1-2\hat{d}^a)/(2-2\hat{d}^a)+\varepsilon}]$ for the adaptive estimator, and the bandwidth $m = [T^u]$, we look into three different trimming-bandwidth combinations for the trimmed and adaptive estimators ($(\varepsilon, u) = (0.01, 0.7)$, $(0.05, 0.8)$, and $(0.1, 0.8)$) and three different bandwidths for the standard estimators ($u = 0.5, 0.7$, and 0.8). Based on 1,000 Monte Carlo replications, we report the finite-sample bias and root mean squared error (RMSE) for sample sizes of $T = 500, 1{,}000$, and $2{,}000$, though financial time series are often longer.[6] We focus on the robust LP estimators. The results for the robust LW estimators are qualitatively similar. Section 5.4 contains a comparison of their relative performance.

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

5.1. Comparative Performance in the Presence of Level Shifts

We begin with some DGPs for which the robust estimators were designed. These are simple Gaussian fractional white noise (ARFIMA$(0, d, 0)$) processes with unit innovation variance, contaminated by RLSs. We report results for $d = 0, 0.2$, and 0.45. The RLS DGPs are generated as in A1(a), setting $p = 5, 10$, and 20 and $\eta_t \sim i.i.d.N(0, 1)$. These are practically relevant values. For example, Lu and Perron (2010) estimated 7–28 level shifts in the volatility of some daily stock market returns series. The results are recorded in Tables 1–3.

Looking at Tables 1 and 2, begin by noting the substantial upward bias that the level shifts cause in the standard LP estimator. This bias is increasing in p, whereas it is typically decreasing in the sample size and bandwidth used.

TABLE 1. Bias and RMSE for short-memory processes with random level shifts

	Standard; u			Trimmed; ε, u			Adaptive; ε, u		
T	0.5	0.7	0.8	0.01, 0.7	0.05, 0.8	0.1, 0.8	0.01, 0.7	0.05, 0.8	0.1, 0.8
Bias					$p = 5$				
500	0.656	0.362	0.254	0.063	0.032	0.025	0.065	0.027	0.002
1,000	0.658	0.335	0.221	0.044	0.021	0.015	0.057	0.020	0.000
2,000	0.685	0.323	0.200	0.045	0.015	0.010	0.068	0.016	0.004
					$p = 10$				
500	0.778	0.471	0.338	0.127	0.052	0.037	0.160	0.075	0.039
1,000	0.792	0.449	0.306	0.095	0.053	0.039	0.163	0.058	0.027
2,000	0.801	0.430	0.277	0.098	0.040	0.027	0.168	0.045	0.022
					$p = 20$				
500	0.866	0.584	0.438	0.208	0.113	0.087	0.338	0.189	0.124
1,000	0.875	0.555	0.391	0.173	0.083	0.063	0.343	0.146	0.088
2,000	0.885	0.526	0.349	0.158	0.064	0.046	0.339	0.117	0.070
RMSE					$p = 5$				
500	0.699	0.394	0.281	0.274	0.160	0.203	0.289	0.161	0.179
1,000	0.697	0.365	0.243	0.198	0.106	0.138	0.232	0.107	0.119
2,000	0.714	0.348	0.218	0.150	0.072	0.089	0.204	0.076	0.081
					$p = 10$				
500	0.802	0.491	0.355	0.295	0.172	0.206	0.370	0.211	0.213
1,000	0.810	0.467	0.323	0.226	0.119	0.139	0.331	0.151	0.141
2,000	0.816	0.447	0.292	0.179	0.087	0.093	0.299	0.112	0.103
					$p = 20$				
500	0.883	0.600	0.453	0.356	0.201	0.225	0.500	0.316	0.290
1,000	0.888	0.568	0.403	0.275	0.145	0.154	0.471	0.247	0.211
2,000	0.894	0.538	0.360	0.227	0.105	0.104	0.443	0.200	0.162

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

TABLE 2. Bias for long-memory processes with random level shifts

	Standard; u			Trimmed; ε, u			Adaptive; ε, u		
T	0.5	0.7	0.8	0.01, 0.7	0.05, 0.8	0.1, 0.8	0.01, 0.7	0.05, 0.8	0.1, 0.8
					$d = 0.2, p = 5$				
500	0.404	0.214	0.150	0.035	0.024	0.019	0.027	0.025	−0.007
1,000	0.384	0.185	0.121	0.020	0.012	0.010	0.066	0.040	0.018
2,000	0.387	0.169	0.103	0.019	0.007	0.006	0.064	0.027	0.013
					$d = 0.2, p = 10$				
500	0.507	0.295	0.212	0.058	0.041	0.029	0.141	0.087	0.041
1,000	0.513	0.275	0.188	0.064	0.034	0.027	0.146	0.076	0.049
2,000	0.506	0.248	0.157	0.058	0.022	0.015	0.144	0.063	0.045
					$d = 0.2, p = 20$				
500	0.606	0.392	0.294	0.135	0.084	0.067	0.254	0.255	0.249
1,000	0.605	0.355	0.248	0.096	0.054	0.045	0.191	0.148	0.117
2,000	0.601	0.323	0.211	0.081	0.035	0.028	0.141	0.115	0.089
					$d = 0.45, p = 5$				
500	0.139	0.077	0.058	0.013	0.020	0.016	0.023	0.045	0.035
1,000	0.133	0.067	0.047	0.017	0.013	0.011	0.049	0.037	0.034
2,000	0.114	0.051	0.033	0.015	0.008	0.009	0.043	0.027	0.025
					$d = 0.45, p = 10$				
500	0.223	0.133	0.098	0.047	0.025	0.021	0.094	0.085	0.075
1,000	0.200	0.107	0.072	0.041	0.013	0.005	0.094	0.063	0.059
2,000	0.180	0.082	0.053	0.026	0.012	0.007	0.074	0.046	0.043
					$d = 0.45, p = 20$				
500	0.301	0.191	0.148	0.085	0.055	0.052	0.157	0.140	0.134
1,000	0.284	0.158	0.113	0.047	0.035	0.029	0.150	0.108	0.105
2,000	0.256	0.126	0.085	0.036	0.025	0.021	0.122	0.080	0.077

Nevertheless, it 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. Note that this bias is decreasing in d, which 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. With $\Gamma(\cdot)$ the gamma function, the variance of a fractional white noise (FWN) process with unit innovation variance is $\Gamma(1-2d)/\Gamma^2(1-d)$, which is increasing in d. It is 1 when $d = 0$, 1.1 when $d = 0.2$, and 3.64 when $d = 0.45$, quite large differences. Yet for all RLS processes considered, its conditional variance is set to unity. Thus the magnitudes of the jumps are relatively smaller when added to processes with higher values of d.

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

TABLE 3. RMSE for long-memory processes with random level shifts

	Standard; u			Trimmed; ε, u			Adaptive; ε, u		
T	0.5	0.7	0.8	0.01, 0.7	0.05, 0.8	0.1, 0.8	0.01, 0.7	0.05, 0.8	0.1, 0.8
					$d = 0.2, p = 5$				
500	0.462	0.248	0.178	0.270	0.157	0.202	0.257	0.149	0.171
1,000	0.434	0.215	0.144	0.192	0.105	0.138	0.206	0.110	0.117
2,000	0.427	0.194	0.122	0.138	0.069	0.088	0.150	0.069	0.071
					$d = 0.2, p = 10$				
500	0.546	0.323	0.235	0.284	0.158	0.204	0.313	0.197	0.211
1,000	0.542	0.298	0.206	0.199	0.111	0.135	0.260	0.144	0.144
2,000	0.529	0.266	0.172	0.148	0.075	0.089	0.223	0.107	0.100
					$d = 0.2, p = 20$				
500	0.633	0.411	0.311	0.313	0.181	0.215	0.392	0.273	0.271
1,000	0.624	0.372	0.263	0.224	0.124	0.146	0.349	0.213	0.200
2,000	0.616	0.337	0.223	0.168	0.084	0.095	0.308	0.164	0.151
					$d = 0.45, p = 5$				
500	0.235	0.126	0.094	0.281	0.157	0.207	0.193	0.102	0.115
1,000	0.208	0.103	0.073	0.186	0.105	0.137	0.121	0.075	0.076
2,000	0.177	0.081	0.053	0.136	0.070	0.089	0.082	0.052	0.053
					$d = 0.45, p = 10$				
500	0.290	0.172	0.128	0.273	0.162	0.209	0.206	0.134	0.145
1,000	0.256	0.136	0.097	0.197	0.107	0.138	0.145	0.097	0.098
2,000	0.227	0.106	0.072	0.135	0.073	0.094	0.107	0.071	0.070
					$d = 0.45, p = 20$				
500	0.350	0.219	0.171	0.292	0.166	0.208	0.244	0.172	0.176
1,000	0.321	0.180	0.131	0.199	0.108	0.130	0.183	0.130	0.130
2,000	0.288	0.146	0.101	0.140	0.075	0.088	0.146	0.100	0.099

The next feature to note is that both the trimmed and adaptive estimators remove large portions of this bias. In terms of bias, the trimmed estimator with the largest trimming ($(\varepsilon, u) = (0.1, 0.8)$) performs best for most of the DGPs considered although the adaptive estimator with this trimming tends to perform better when $d = 0$ and p is not too high. This is expected again from (2) because 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 $\hat{d}^a \approx 0$. 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 estimators almost entirely eliminate the bias in many cases. The main result that emerges is the larger the trimming, the less bias due to level shifts will be present.

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

In terms of RMSE, the standard estimator is generally dominated by both its trimmed and adaptive counterparts, as can be seen from Tables 1 and 3.[7] This dominance is often quite dramatic, especially for the more frequent level shift cases. When $p = 10$ or 20, the RMSE of the robust estimators is typically one-half to one-quarter of its standard counterpart, a major improvement. In most cases considered, the trimmed estimator with moderate trimming ($(\varepsilon, u) = (0.05, 0.8)$) best balances the bias-variance trade-off, but, as the sample size grows, the adaptive estimator essentially "catches up" to its trimmed counterpart by reducing finite-sample variance. 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 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. We also examined DGPs calibrated to stock market volatility data by Lu and Perron (2010) with results not reported as they are very similar to those discussed earlier. Finally, we examined the robust estimators' performance in the presence of a variety of deterministic trends and found similar results, though the upward bias did not tend to be quite as large as in the level shifts cases, favoring adaptive estimation for RMSE minimization.

5.2. Comparative Performance without Contaminating Components

We now turn to cases with no contaminating elements present to examine what is lost from using robust estimation when level shifts or trends are not an issue and to provide guidance on choosing (ε, u) that achieves low RMSE in a variety of circumstances. We first examine the simple case of an uncontaminated FWN process, again for $d = 0, 0.2$, and 0.45. Table 4 displays these results. As expected, neither the standard estimator nor the trimmed one displays any notable bias. Oddly, the adaptive estimator appears to induce a very small downward bias though this bias disappears as the sample grows. In terms of RMSE, the standard estimator is the best, as expected. The RMSEs of the trimmed estimators can be 2–3 times the value of their untrimmed counterparts. For the smaller samples, relative to the differences in bias and RMSE when level shifts are present, these differences in RMSE tend to be similar in value. As the sample size grows, the differences in bias and RMSE between the trimmed and standard estimators shrink in the absence of contamination but grow in its presence. Hence, depending on the sample size, one has about the same or less to lose from using an adaptive estimator with, say $(\varepsilon, u) = (0.05, 0.8)$, in the absence of level shifts than from using a standard estimator in their presence.

Similar results hold when the DGP is a short-long-memory process "perturbed" by random noise, as considered by, e.g., Sun and Phillips (2003). This is empirically relevant as many measures of volatility are known to be noisy. Note that when the process is perturbed by noise, A2 fails to hold. However, Deo and Hurvich (2001) have shown that the standard LP estimator is also consistent and

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

TABLE 4. Bias and RMSE for uncontaminated processes

	Standard; u			Trimmed; ε, u			Adaptive; ε, u		
T	0.5	0.7	0.8	0.01, 0.7	0.05, 0.8	0.1, 0.8	0.01, 0.7	0.05, 0.8	0.1, 0.8
Bias					$d = 0$				
500	0.009	−0.004	−0.003	−0.011	−0.002	0.005	−0.056	−0.024	−0.032
1,000	−0.001	−0.001	−0.002	−0.013	−0.005	−0.003	−0.027	−0.012	−0.020
2,000	0.004	0.005	0.000	−0.001	−0.008	−0.006	−0.011	−0.006	−0.008
					$d = 0.2$				
500	0.003	0.001	0.001	−0.013	0.000	−0.002	−0.070	−0.025	−0.046
1,000	−0.004	0.001	0.001	−0.003	−0.002	−0.003	−0.037	−0.010	−0.020
2,000	0.006	0.003	0.001	0.007	−0.001	−0.003	−0.015	−0.004	−0.009
					$d = 0.45$				
500	0.016	0.006	0.003	−0.004	−0.002	−0.004	−0.041	−0.010	−0.017
1,000	0.014	0.005	0.004	−0.002	0.003	0.005	−0.012	0.000	−0.002
2,000	0.010	0.005	0.002	0.001	0.000	−0.002	0.001	0.000	0.000
RMSE					$d = 0$				
500	0.167	0.080	0.060	0.278	0.155	0.200	0.227	0.135	0.173
1,000	0.132	0.060	0.042	0.189	0.100	0.135	0.152	0.093	0.115
2,000	0.110	0.045	0.031	0.131	0.066	0.087	0.113	0.064	0.077
					$d = 0.2$				
500	0.163	0.084	0.060	0.271	0.153	0.204	0.213	0.120	0.155
1,000	0.136	0.061	0.043	0.186	0.102	0.129	0.142	0.079	0.098
2,000	0.107	0.046	0.032	0.130	0.070	0.087	0.093	0.050	0.062
					$d = 0.45$				
500	0.171	0.082	0.059	0.272	0.155	0.199	0.179	0.081	0.101
1,000	0.138	0.065	0.047	0.191	0.102	0.137	0.102	0.051	0.056
2,000	0.115	0.050	0.036	0.138	0.072	0.093	0.054	0.038	0.039

asymptotically normal in the presence of noise and absence of contamination. Because proving these results requires an entirely separate strategy, the proof for the trimmed or adaptive estimators in the potential presence of noise is beyond the scope of this paper. We conjecture that, under proper trimming and bandwidth choice, the robust estimators remain consistent and asymptotically normal when noise is added to the $\{v_t\}$ component of (1). Though consistency still holds, the addition of white noise to the process $\{v_t\}$ is known to bias standard LP estimates downward in finite samples. Because we are considering an LP estimator with lower frequencies trimmed out, we can expect this downward bias to be exacerbated. The question addressed subsequently is, comparatively, how much worse is this downward bias and the corresponding RMSE?

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

TABLE 5. 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	0.01, 0.7	0.05, 0.8	0.1, 0.8	0.01, 0.7	0.05, 0.8	0.1, 0.8
Bias					$d = 0.2$				
500	−0.111	−0.132	−0.142	−0.154	−0.160	−0.164	−0.209	−0.181	−0.198
1,000	−0.105	−0.126	−0.137	−0.150	−0.160	−0.160	−0.179	−0.171	−0.182
2,000	−0.092	−0.123	−0.133	−0.147	−0.150	−0.151	−0.166	−0.156	−0.163
4,000	−0.092	−0.118	−0.132	−0.144	−0.152	−0.154	−0.153	−0.154	−0.160
8,000	−0.082	−0.113	−0.129	−0.134	−0.149	−0.151	−0.139	−0.148	−0.154
					$d = 0.45$				
500	−0.121	−0.208	−0.253	−0.309	−0.349	−0.353	−0.352	−0.364	−0.387
1,000	−0.096	−0.187	−0.239	−0.293	−0.336	−0.347	−0.305	−0.336	−0.358
2,000	−0.070	−0.167	−0.224	−0.261	−0.316	−0.329	−0.253	−0.300	−0.322
4,000	−0.051	−0.150	−0.212	−0.242	−0.298	−0.311	−0.209	−0.274	−0.292
8,000	−0.035	−0.133	−0.201	−0.218	−0.282	−0.297	−0.172	−0.252	−0.268
RMSE					$d = 0.2$				
500	0.202	0.154	0.154	0.316	0.225	0.264	0.300	0.225	0.259
1,000	0.173	0.141	0.144	0.240	0.193	0.214	0.236	0.195	0.216
2,000	0.145	0.132	0.137	0.197	0.166	0.175	0.198	0.169	0.179
4,000	0.130	0.124	0.134	0.171	0.159	0.165	0.173	0.160	0.168
8,000	0.110	0.116	0.131	0.148	0.152	0.157	0.150	0.151	0.159
					$d = 0.45$				
500	0.211	0.224	0.260	0.406	0.382	0.409	0.412	0.389	0.421
1,000	0.168	0.197	0.243	0.347	0.351	0.370	0.343	0.349	0.376
2,000	0.134	0.174	0.227	0.293	0.323	0.341	0.279	0.307	0.332
4,000	0.104	0.155	0.214	0.259	0.301	0.316	0.222	0.277	0.297
8,000	0.084	0.137	0.202	0.228	0.284	0.300	0.179	0.254	0.271

Table 5 displays the results when a very substantial white noise component is added to the long-memory process $\{v_t\}$ for $d = 0.2$ and 0.45. The white noise is Gaussian with variance set equal to 4. We added the larger sample sizes of $T = 4{,}000$ and $8{,}000$ to the analysis here to obtain a more detailed picture of how the estimators behave for sample sizes that are typical of daily volatility series, like those in Section 6. When $d = 0.45$, we see a substantial downward bias in all estimators. The bias is lower for those estimators that use proportionately more lower frequencies, e.g., the standard estimator with the smallest bandwidth performs best. Nevertheless, consistent with our conjecture, the biases of all estimators decrease with T. Making the more fair comparison between the standard estimator and its trimmed/adaptive counterparts by comparing at equal bandwidths, the increase in bias is typically less than or equal to 0.1. Very similar results hold for RMSE as bias is the dominant component for these DGPs.

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

Similar results hold for $d = 0.2$ although differences between standard and robust estimators are smaller. Nevertheless, the differences in bias and RMSE between the standard and robust estimators in the presence of noise tend to be smaller than they are with level shifts and trends.

We now turn to cases in which a significant short-memory component is present but contaminating elements are not. We consider various specifications of an ARFIMA$(1, d, 1)$ process for $\{v_t\}$, $(1-aL)(1-L)^d v_t = (1-bL)e_t$, where $e_t \sim i.i.d.N(0,1)$. We examine specifications that are persistent as these are more empirically relevant. To begin, we consider a process that is persistent yet $I(0)$ setting $a = 0.6$, $b = 0$, and $d = 0$. The memory parameter estimates of such a process are known to be upward biased. The first and fourth panels of Table 6 show this to be the case. In terms of bias and RMSE, the standard estimator with

TABLE 6. Bias and RMSE for ARFIMA processes

	Standard; u			Trimmed; ε, u			Adaptive; ε, u		
T	0.5	0.7	0.8	0.01, 0.7	0.05, 0.8	0.1, 0.8	0.01, 0.7	0.05, 0.8	0.1, 0.8
Bias				$a = 0.6, b = 0, d = 0$					
500	0.048	0.221	0.369	0.533	0.717	0.766	0.311	0.430	0.443
1,000	0.019	0.165	0.322	0.414	0.633	0.696	0.257	0.389	0.402
2,000	0.008	0.123	0.283	0.305	0.552	0.619	0.203	0.349	0.363
				$a = 0.6, b = 0, d = 0.45$					
500	0.063	0.228	0.370	0.518	0.703	0.753	0.229	0.370	0.370
1,000	0.033	0.176	0.329	0.423	0.632	0.690	0.176	0.329	0.329
2,000	0.018	0.130	0.287	0.312	0.553	0.617	0.130	0.287	0.287
				$a = 0, b = -0.6, d = 0.45$					
500	0.010	0.036	0.110	0.097	0.288	0.354	0.017	0.111	0.112
1,000	0.015	0.026	0.079	0.058	0.184	0.224	0.019	0.080	0.080
2,000	0.009	0.016	0.058	0.035	0.130	0.158	0.014	0.059	0.059
RMSE				$a = 0.6, b = 0, d = 0$					
500	0.177	0.235	0.373	0.597	0.734	0.793	0.330	0.432	0.445
1,000	0.143	0.176	0.325	0.455	0.642	0.709	0.269	0.390	0.403
2,000	0.107	0.131	0.285	0.332	0.557	0.626	0.213	0.350	0.364
				$a = 0.6, b = 0, d = 0.45$					
500	0.179	0.243	0.376	0.586	0.722	0.780	0.243	0.376	0.376
1,000	0.143	0.187	0.332	0.464	0.641	0.704	0.187	0.332	0.332
2,000	0.110	0.139	0.290	0.338	0.558	0.624	0.139	0.290	0.290
				$a = 0, b = -0.6, d = 0.45$					
500	0.172	0.091	0.127	0.292	0.329	0.409	0.138	0.127	0.127
1,000	0.140	0.069	0.092	0.203	0.211	0.262	0.089	0.092	0.092
2,000	0.110	0.054	0.069	0.137	0.148	0.182	0.057	0.069	0.069

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

$u = 0.5$ is clearly the favorite. Comparing robust estimators to the standard one at the same bandwidths, the trimmed estimators roughly double the bias, whereas the adaptive estimators increase bias to a *much* lesser extent. In many cases, the adaptive estimator hardly increases the bias. 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, whereas 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 (see the second and fifth panels of Table 6 for $d = 0.45$). Thus, if we are not concerned about level shifts or deterministic trends but about a strong autoregressive component, the standard estimator with $u = 0.5$ is preferred. If we are concerned about both, a version of the adaptive estimator is generally preferred. Similar results hold with persistence induced by a moving average component, though biases and RMSEs tend to be much smaller so that the standard estimator and our adaptive estimator both perform well and similarly (e.g., the third and sixth panels of Table 6 for $a = 0$, $b = -0.6$, and $d = 0.45$). The results are similar for other values of d.

In summary, if level shifts or trend components are present, the robust estimation is clearly superior. If one is agnostic about the DGP, our adaptive estimator with $(\varepsilon, u) = (0.05, 0.8)$ is recommended, as it best balances different potential biases. Even without contaminating components, the cost of using it is relatively small, although gains could be made using the standard estimator with a small bandwidth. If short-memory dynamics or noise is of little concern, the trimmed estimator is clearly preferred.

5.3. Robustness to Non-Gaussianity

Given that volatility series are relevant processes exhibiting long-memory features, one may be concerned with A4. For example, when absolute or squared returns are used as proxies, they are by construction non-Gaussian. In this section we illustrate, through simulations, that the Gaussianity assumption is hardly critical and can likely be dropped. We conducted numerous simulation experiments to assess the bias and RMSE of the robust estimators under a variety of distributional assumptions on the $\{v_t\}$ process following the design of Velasco (2000). The results are almost identical in all cases, and so we only present a small subset in Table 7: an ARFIMA$(0, d, 0)$ process with t_5 distributed innovations, $d = 0.45$, $p = 0$ and $d = 0$, $p = 10$. Comparing these results to those corresponding to the Gaussian specifications in Tables 1 and 4, we can see that changing the innovation distribution to be non-Gaussian does not hinder the performance of the robust estimators. This is a generic result that holds for many distributions with $d = 0, 0.45$ and $p = 0, 10$. Moreover, unreported results on finite-sample coverage probabilities suggest that the asymptotic distributional approximation $\hat{d} \sim N(d, \pi^2/24S(l, m))$ works just as well when the underlying process is non-Gaussian.

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

TABLE 7. Bias and RMSE for t_5 innovation distributed short- and long-memory processes

	Standard; u			Trimmed; ε, u			Adaptive; ε, u		
T	0.5	0.7	0.8	0.01, 0.7	0.05, 0.8	0.1, 0.8	0.01, 0.7	0.05, 0.8	0.1, 0.8
Bias					$d = 0.45, p = 0$				
500	0.007	0.011	0.009	−0.004	0.003	0.010	−0.043	0.001	−0.009
1,000	0.011	0.008	0.005	0.000	−0.001	−0.003	−0.005	0.001	0.001
2,000	0.009	0.006	0.004	0.004	0.002	0.002	0.002	0.003	0.002
					$d = 0, p = 10$				
500	0.705	0.396	0.279	0.061	0.034	0.027	0.071	0.028	0.000
1,000	0.726	0.384	0.256	0.066	0.032	0.025	0.090	0.034	0.015
2,000	0.732	0.358	0.225	0.063	0.024	0.012	0.103	0.027	0.008
RMSE					$d = 0.45, p = 0$				
500	0.176	0.084	0.063	0.281	0.160	0.209	0.191	0.081	0.108
1,000	0.139	0.065	0.046	0.190	0.103	0.136	0.095	0.051	0.054
2,000	0.115	0.050	0.035	0.132	0.070	0.087	0.059	0.038	0.039
					$d = 0, p = 10$				
500	0.734	0.419	0.299	0.290	0.163	0.213	0.307	0.159	0.182
1,000	0.749	0.404	0.272	0.210	0.108	0.130	0.255	0.121	0.126
2,000	0.750	0.376	0.239	0.149	0.077	0.091	0.226	0.086	0.088

5.4. Comparison of Robust LP and LW Estimators

We compared the performance of the trimmed and adaptive LP estimators with trimmed and adaptive versions of the LW estimator across a wide variety of DGPs. We examined trimmed and adaptive versions of the LW estimator based on the procedures given earlier in this section (arising from A3) and also those based on Assumption 3 of Iacone (2010), i.e., $l = [T^{1-u+\varepsilon}]$ and $l = [T^{(1-2\hat{d}^a-u)/(1-\hat{d}^a)+\varepsilon}]$. The results show that the former versions of the LW estimator perform *significantly* better in the presence of level shifts and deterministic trends. This makes sense in light of the discussion of orders following (2). Comparing between the LP and LW versions based on the trimming and adaptive procedures of this section, a few general features emerged: (i) the RMSE performance of the two estimators is similar in the presence of contamination, (ii) the robust LP estimators tend to better remove biases from level shifts and deterministic trends, and (iii) the robust LW estimators tend to be less variable. Table 8 displays the bias and RMSE for the robust LW estimator under two DGPs that are illustrative of these general patterns: a pure ARFIMA$(0, 0.45, 0)$ process and a white noise process with an average of 10 RLS per sample. Comparing the second panel of Table 8 with the second panel of Table 1 illustrates (ii). The robust LP estimators exhibit lower bias in 16 of the 18 cases, up to 45% lower. Comparing the first and third panels

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

TABLE 8. Bias and RMSE for trimmed LW estimator

	Standard; u			Trimmed; ε, u			Adaptive; ε, u		
T	0.5	0.7	0.8	0.01, 0.7	0.05, 0.8	0.1, 0.8	0.01, 0.7	0.05, 0.8	0.1, 0.8
Bias				$d = 0.45, p = 0$					
500	−0.006	−0.001	−0.011	−0.022	−0.038	−0.045	−0.024	−0.019	−0.023
1,000	−0.003	0.000	−0.008	−0.008	−0.025	−0.032	−0.008	−0.012	−0.013
2,000	0.002	0.001	−0.006	−0.004	−0.019	−0.025	−0.002	−0.009	−0.010
				$d = 0, p = 10$					
500	0.771	0.515	0.401	0.138	0.064	0.044	0.220	0.091	0.045
1,000	0.796	0.510	0.390	0.121	0.052	0.036	0.229	0.081	0.046
2,000	0.811	0.507	0.384	0.107	0.046	0.032	0.231	0.068	0.040
RMSE				$d = 0.45, p = 0$					
500	0.147	0.069	0.050	0.208	0.116	0.147	0.119	0.062	0.069
1,000	0.108	0.052	0.038	0.149	0.082	0.106	0.069	0.042	0.045
2,000	0.090	0.041	0.029	0.102	0.057	0.073	0.046	0.032	0.034
				$d = 0, p = 10$					
500	0.790	0.531	0.415	0.267	0.137	0.160	0.381	0.193	0.175
1,000	0.808	0.523	0.402	0.205	0.099	0.106	0.364	0.164	0.139
2,000	0.820	0.517	0.393	0.167	0.079	0.078	0.348	0.134	0.109

of Table 8 with the third and sixth panels of Table 4 illustrates (iii). In the absence of contamination, the RMSE of the robust LW estimators is around 78% of that for the robust LP estimators, as suggested by their relative asymptotic variance.

6. EMPIRICAL APPLICATIONS TO STOCK MARKET VOLATILITY DATA

We apply the robust estimation techniques to various time series that have been typified as arising from long-memory processes. We examine the extent to which level shifts or deterministic trends may produce the long-memory features of the data. For each time series, we graphed the adaptive and trimmed LP estimators against the trimming parameter ε, setting $K_l = 1$ and $m = [T^{0.79}]$ to be congruent with A3* but close to the MSE-optimal bandwidth rate. We let ε range from 0 to 0.25 to display the effects of the trimming on the d estimates. Each graph also contains pointwise 95% confidence bands for the trimmed estimators using the $S(l, m)$ scaling (with $d \geq 0$, the trimmings with $\varepsilon > 0.1$ are guaranteed to satisfy A3*). These graphs are provided in Figures 1–4. The results for robust LW estimators are quite similar. For each series, we also provide the standard LP estimator of d using $m = [T^{0.5}]$, the nonlinear LP (NLP) estimator of Sun and Phillips (2003) using $m = [T^{0.5}]$ (which reduces the bias from noise), and the

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

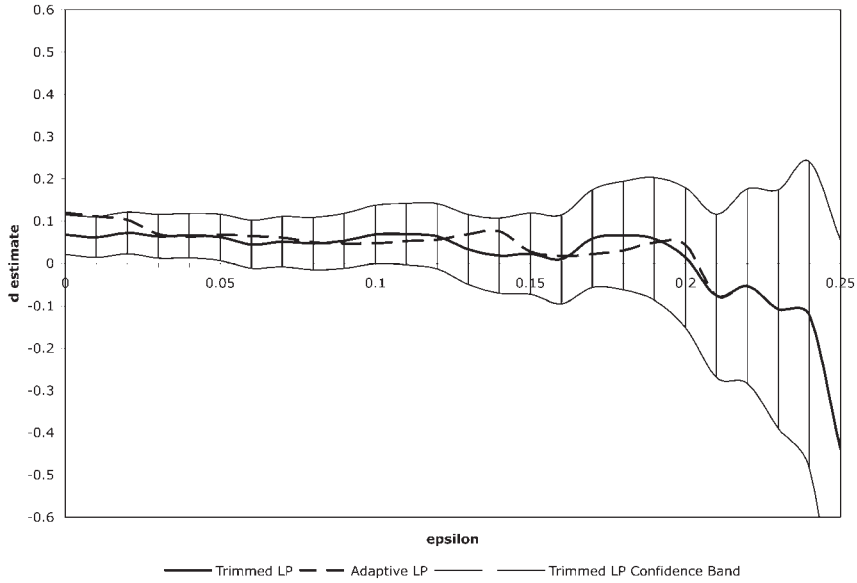


FIGURE 1. S&P 500 daily volatility.

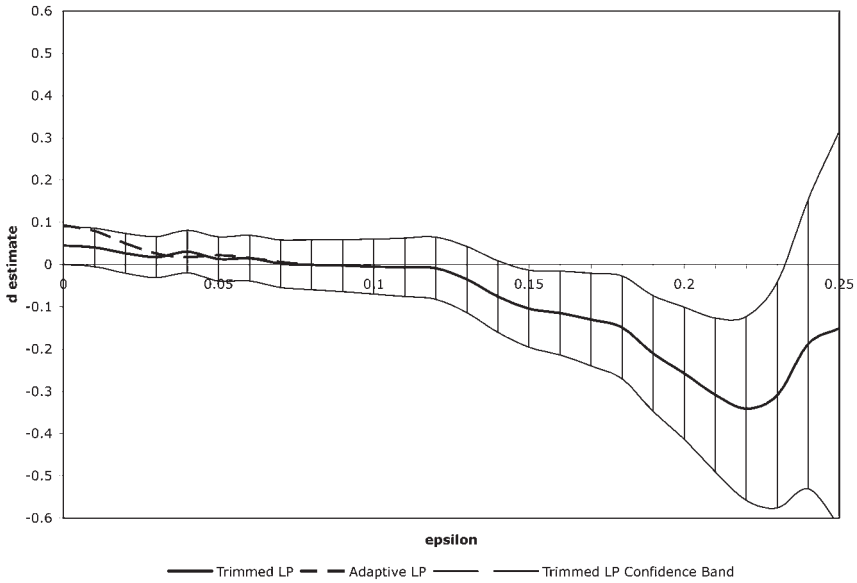


FIGURE 2. DJIA daily volatility.

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

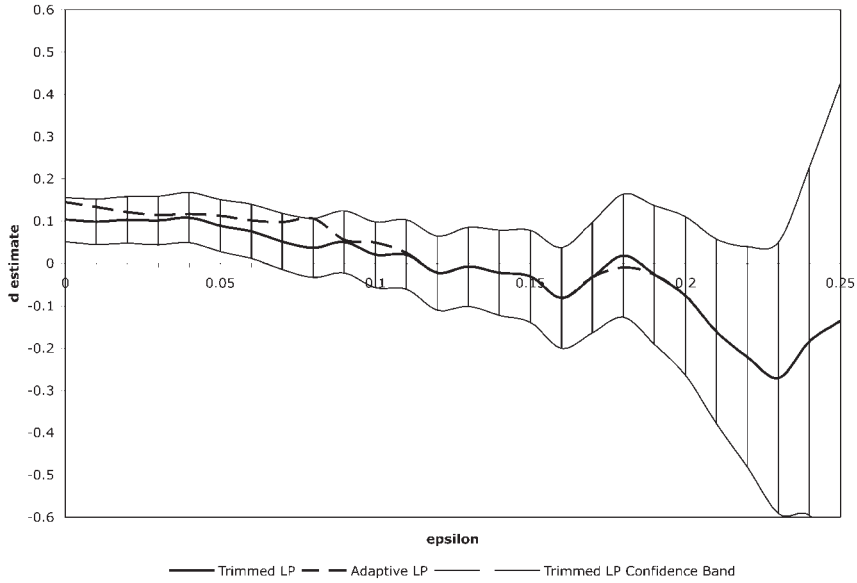


FIGURE 3. NASDAQ daily volatility.

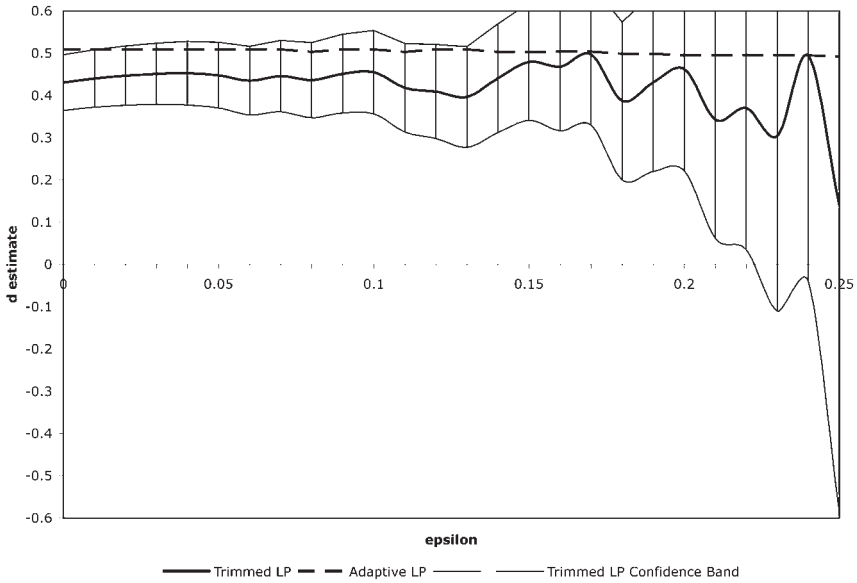


FIGURE 4. S&P 500 realized volatility.

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

adaptive LP estimator using $(l, m) = ([T^{(1-2\hat{d}^a)/(2-2\hat{d}^a)+0.05}], [T^{0.79}])$ and their standard errors (in parentheses).

The first three time series we study are the log-absolute daily returns series of the S&P 500 (7/3/1962–7/23/2012, 12,603 observations), the Dow Jones Industrial Average (DJIA, 3/4/1957–7/23/2012, 13,982 observations), and the NASDAQ (12/15/1972–7/23/2012, 9,991 observations) stock market indexes. Log-absolute returns are a common measure of volatility. Note that these series are much longer than those studied in the previous section; hence biases from short-memory dynamics or noise components should be much lower than reported.

Starting with the S&P 500 volatility series, the standard LP and NLP estimates are 0.505 (0.061) and 0.622 (0.109), indicating a nonstationary long-memory process. On the other hand, the adaptive estimator gives an estimate of 0.068 (0.025), indicating very weak long-memory. Figure 1 provides a more complete picture. All robust estimates are well below the standard ones, strongly suggesting that level shifts or deterministic trends appear to be biasing the standard estimate upward. Moreover, the adaptive and trimmed estimators are similar and the 95% confidence band covers zero for most trimmings. However, estimates using trimmings above about 0.2 become somewhat erratic and unreliable, which can be attributed to the estimators using less data and the periodogram ordinates used in the LP regressions growing progressively farther from the origin.

Turning to the other two daily volatility series, very similar results emerge. The standard LP estimator for the DJIA series is 0.470 (0.054), this time in the stationary region, whereas the NLP estimator is 0.540 (0.109). The adaptive estimator is 0.022 (0.026), indicative of short memory. The overall pattern present in Figure 2 is quite similar to that for the S&P 500 series. The results for the NASDAQ are again quite similar with the standard estimator being 0.601 (0.064), the NLP estimator being 0.655 (0.114), and the adaptive estimator being 0.112 (0.026) and the graph of the robust estimators displaying a similar pattern (see Figure 3). For all three of the preceding series, the results are highly indicative of a short-memory or very weak long-memory process contaminated by level shifts or deterministic components.

We now turn to a less noisy measure of volatility: the log of daily realized volatility constructed from 5-minute returns of the S&P 500 futures index from 4/21/1982 to 3/2/2007 (6,262 observations).[8] The features of the memory parameter estimates of this series are quite different in that they are broadly in line with a true long-memory process. First, the standard LP estimator is 0.625 (0.072), and the NLP estimator is also 0.625 (0.130), indicating a nonstationary long-memory process. Second, the adaptive estimator gives a value of 0.509 (0.021), indicating a highly persistent long-memory process. The robust estimators in Figure 4 are consistent with a long-memory process near the border of stationarity.

Finally, the values of the test statistic $H(1, [T^{0.51}])$ proposed in Section 4 are as follows: 38.9 for S&P 500, 41.7 for DJIA, 31.3 for NASDAQ, and 9.4 for the S&P 500 futures log realized volatility. These provide strong evidence that level

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

shifts or trends do indeed contaminate all series, including the realized volatility series for which long memory is still present.

7. CONCLUSIONS AND FUTURE RESEARCH

We have shown that simple modifications to the standard estimators of the memory parameter lead to estimators that are robust to a wide range of level shifts and deterministic components. The trimmed and adaptive estimators have good asymptotic properties and perform well in finite samples. To balance potential competing biases, we advocate a particular version of adaptive LP estimation, employing the trimming $l = [T^{(1-2\hat{d}^a)/(2-2\hat{d}^a)+0.05}]$ 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 the robust estimators to volatility data, we found that for many series, level shifts or deterministic trends appear to bias standard memory parameter estimates upward as the robust estimators indicate memory parameters near zero. Nevertheless, we found evidence of long memory in a log realized volatility series.

For future research, there appear to be many fruitful avenues. Recent years have seen the emergence of estimators that reduce the biases arising from (i) level shifts and deterministic trends (e.g., the present paper), (ii) short-memory dynamics (e.g., Andrews and Guggenberger, 2003), and (iii) the presence of noise (e.g., Sun and Phillips, 2003). Very recently, Frederiksen, Nielsen, and Nielsen (2010) produced an estimator that simultaneously reduces the biases arising from (ii) and (iii), even allowing for dynamics in the noise component. Given the abundant evidence of the presence of (i) in economic and financial data, methods aimed at simultaneously reducing the biases arising from (i) and (ii), (i) and (iii), or (i)–(iii) would also prove quite useful in practice. Apart from this important issue, we believe further improvements can be made to reducing the bias arising from level shifts and deterministic components. Remarks 2 and 3 provide suggestions in this vein.

NOTES

1. Some authors have addressed memory parameter estimation in the presence of specific types of deterministic trends. See Robinson (1997) and Hurvich, Lang, and Soulier (2005) for examples. In this paper we aim for a more general treatment that applies to a wide variety of contaminating elements.

2. Robinson (1995) also employed a trimming for the LP estimator but for an entirely different purpose: to derive the asymptotic properties of the estimator when applied to a pure long-memory process. HDB later showed that this trimming is unnecessary.

3. This includes all cases for which $h(\cdot)$ is monotonic and bounded because we can simply subtract $h(1)$ from $h(\cdot)$ and add $h(1)$ to k to have the same DGP.

4. One could also formulate an analogous adaptive procedure that coincides with the final term of Assumption 3*. This would lead to more trimming but may be useful if asymptotic normality is a concern.

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

5. We chose this criterion because, from unreported results, it performs relatively well with very high rates of convergence in larger ($T = 2{,}000$) samples without being too computationally intensive.

6. Note that some of the bandwidth/trimming parameter combinations under scrutiny do not satisfy Assumption 3* (e.g., the MSE-optimal bandwidth rate of $T^{4/5}$) because the primary goal of this simulation exercise is to compare estimators in terms of bias and RMSE performance.

7. 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 rescaling the variance of η_t to be comparable to that of v_t, the standard estimator is RMSE-dominated in all cases considered. Still, using this rescaling, the reductions in RMSE from using the robust estimators decrease as d increases but remain quite substantial.

8. We thank Shinsuke Ikeda for providing this data set. For details on its construction see Ikeda (2013).

REFERENCES

Anderson, T.G., T. Bollerslev, T. Diebold, & P. Labys (2001) The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96, 42–55.

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

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

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

Dahlhaus, R. (1989) Efficient parameter estimation for self similar processes. *Annals of Statistic*s 17, 1749–1766.

Deo, R.S. & C.M. Hurvich (2001) On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models. *Econometric Theory* 17, 686–710.

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

Dolado, J., J. Gonzalo & L. Mayoral (2005) What Is What? A Simple Test of Long-Memory versus Structural Breaks in the Time Domain. Manuscript, Departments of Economics, Universidad Carlos III de Madrid and Universitat Pompeu Fabra.

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

Frederiksen, P., F.S. Nielsen, & M.Ø. Nielsen (2012) Local polynomial Whittle estimation of perturbed fractional processes. *Journal of Econometrics* 167, 426–447.

Garcia, R. & P. Perron (1996) An analysis of the real interest rate under regime shifts. *Review of Economics and Statistics* 78, 111–125.

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

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

Granger, C.W.J. & N. Hyung (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.

Haldrup, N. & M.Ø. Nielsen (2007) Estimation of fractional integration in the presence of data noise. *Computational Statistics and Data Analysis* 51, 3100–3114.

Hosking, J.R.M. (1981) Fractional differencing. *Biometrika* 68, 165–176.

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

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

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

Hurvich, C.M., R. Deo & J. Brodsky (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.M., G. Lang & P. Soulier (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. (2013) Two Scale Realized Kernels: A Univariate Case. Manuscript, Faculty of Economics, National Graduate Institute of Policy Studies.

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

Künsch, H.R. (1987) Statistical aspects of self-similar processes. In Y. Prohorov & V. Sazarov (eds.), *Proceedings of the First World Congress of the Bernoulli Society*, vol. 1, pp. 67–74. VNU Science Press.

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

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

Ohanissian, A., J.R. Russell, & R.S. Tsay (2004) True or Spurious Long Memory in Volatility: Does It Matter for Pricing Options?, Manuscript, Booth School of Business, University of Chicago.

Ohanissian, A., J.R. Russell, & R.S. Tsay (2008) True or spurious long memory? A new test. *Journal of Business & 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. & Z. Qu (2010) Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business & Economic Statistics* 28, 275–290.

Qu, Z. (2011) A test against spurious long memory. *Journal of Business & Economic Statistics* 29, 423–438.

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

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

Shimotsu, K. (2006) Simple (but effective) Tests of Long Memory versus Structural Breaks, Working paper 1101, Department of Economics, Queen's University.

Skovgaard, I.M. (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 & Economic Statistics* 23, 321–335.

Sun, Y. & P.C.B. Phillips (2003) Nonlinear log-periodogram regression for perturbed fractional processes. *Journal of Econometrics* 115, 355–389.

Taylor, S.J. (2000) Consequences for Option Pricing of a Long Memory in Volatility. 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.

MATHEMATICAL APPENDIX

Proof of Theorem 1. We start with the case of the DGP of A1(a).

(i) Because $\{v_t\}$ and $\{u_t\}$ are independent, $\mathbb{E}I_x\left(\lambda_j\right) = \mathbb{E}I_v\left(\lambda_j\right) + \mathbb{E}I_u\left(\lambda_j\right)$, where $I_v(\cdot)$ and $I_u(\cdot)$ denote the periodograms of $\{v_t\}_{t=1}^T$ and $\{u_t\}_{t=1}^T$. By Theorem 2(a)

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

of Robinson (1995), $\mathbb{E}[I_v(\lambda_j)/(f^*(0)\lambda_j^{-2d})] = 1 + O[(\log j)/j + (j/T)^2]$, and by A2 and Proposition 3 of Perron and Qu (2010),

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

(ii) Again using the independence of $\{v_t\}$ and $\{u_t\}$, $\mathbb{E}w_x(\lambda_j)^2 = \mathbb{E}w_v(\lambda_j)^2 + \mathbb{E}w_u(\lambda_j)^2$, where $w_v(\cdot)$ and $w_u(\cdot)$ denote the discrete Fourier transforms of $\{v_t\}_{t=1}^T$ and $\{u_t\}_{t=1}^T$. By Theorem 2(b) of Robinson (1995), $\mathbb{E}[w_v(\lambda_j)^2/(f^*(0)\lambda_j^{-2d})] = O((\log j)/j)$.The arguments used to find the order of magnitude of (A.6) in the proof of Lemma A.3 also follow through for $n = 2$. Hence, $\mathbb{E}[w_u(\lambda_j)^2] = O(T/j^2)$ is a special case of this result so that $\mathbb{E}[w_u(\lambda_j)^2/(f^*(0)\lambda_j^{-2d})] = O\left(T^{1-2d}/j^{2-2d}\right)$.

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

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

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

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

and applying Theorem 2(d) of Robinson (1995). We now consider DGP A1(b). First,

$$\mathbb{E}I_x(\lambda_j) = \mathbb{E}I_v(\lambda_j) + \frac{1}{2\pi T}\sum_{i=1}^{B}\sum_{k=1}^{B}c_i c_k \sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{I}\left(T_{i-1} < t \le T_i\right) \times \mathbb{I}\left(T_{k-1} < s \le T_k\right)\left[\cos(\lambda_j t)\cos(\lambda_j s) + \sin(\lambda_j t)\sin(\lambda_j s)\right]. \quad \textbf{(A.1)}$$

Note that by using the fact that $\cos(2\pi jx) = \partial(\sin(2\pi jx)/(2\pi j))/\partial x$,

$$\begin{aligned}
&(j/T)\sum_{t=1}^{T}\mathbb{I}(T_{i-1} < t \le T_i)\cos(\lambda_j t)\\
&\quad = (j/T)\sum_{t=T_{i-1}+1}^{T_i}\cos(\lambda_j t)\\
&\quad = (j/T)\sum_{t=T_{i-1}+1}^{T_i}\left\{(\sin(2\pi j(t+1)/T) - \sin(2\pi jt/T))/(2\pi j/T) + o\left(T^{-1}\right)\right\}\\
&\quad = \sin(2\pi j(T_i+1)/T)/2\pi - \sin(2\pi j(T_{i-1}+1)/T)/2\pi + o(j/T) = O(1).
\end{aligned}$$

Analogous results hold for the other terms that compose (A.1). Hence, (A.1) is $O(T/j^2)$. Thus, applying the results of Theorem 2(a) of Robinson (1995),

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

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

The proofs of parts (ii)–(iv) are similar. We now consider the DGP of A1(c). First note that Qu (2011) and Künsch (1986) have shown that $|\sum_{t=1}^{T} h(t/T)\cos(\lambda_j t)|$ and $|\sum_{t=1}^{T} h(t/T)\sin(\lambda_j t)|$ are $O(T/j)$ when h is Lipschitz continuous or monotonic and bounded (respectively). Together with Robinson (1995, Thm. 2), this implies the results because, e.g.,

$$\mathbb{E}\left[I_x\left(\lambda_j\right)\right] = \mathbb{E}\left[I_v\left(\lambda_j\right)\right] + \frac{1}{2\pi T}\left|\sum_{t=1}^{T} h(t/T)\cos\left(\lambda_j t\right)\right|^2 + \frac{1}{2\pi T}\left|\sum_{t=1}^{T} h(t/T)\sin\left(\lambda_j t\right)\right|^2,$$

$$\mathbb{E}\left[w_x\left(\lambda_j\right)^2\right] = \mathbb{E}\left[w_v\left(\lambda_j\right)^2\right] + \frac{1}{2\pi T}\left(\sum_{t=1}^{T} h(t/T)\exp\left(i\lambda_j t\right)\right)^2,$$

$$\left|\sum_{t=1}^{T} h(t/T)\exp\left(i\lambda_j t\right)\right| \leq \left(\left|\sum_{t=1}^{T} h(t/T)\cos\left(\lambda_j t\right)\right| + \left|\sum_{t=1}^{T} h(t/T)\sin\left(\lambda_j t\right)\right|\right).$$

Using Theorem 1(ii) in Iacone (2010), the proof under A1(d) is nearly identical. ■

The proofs of Theorems 2 and 3 follow those of Theorems 1 and 2 in HDB. First, we state and prove some lemmas, using the notation of HDB. Let $a_j = Y_j - \bar{Y}$ and $S_{YY} = \sum_{k=l}^{m} a_k^2$, so that

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

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

LEMMA A.1. *Under A3, $a_j = O(\log m)$ for all $l \leq j \leq m$ and $S_{YY} = m + o(m)$.*

Proof. The first statement follows directly from a slight modification of arguments used by Hurvich and Beltrao (1994, pp. 299–301). For the second, following the same arguments, for all $l \leq j \leq m$, $a_j = \log j - (m-l+1)^{-1}\left[\log m! - \log(l-1)!\right] + o(1)$. By Stirling's formula and A3, $(m-l+1)^{-1}\left[\log m! - \log(l-1)!\right] = (m/(m-l+1))\log m - 1 + o(1)$. Thus,

$$\begin{aligned}\frac{1}{m}\sum_{j=l}^{m} a_j^2 &= \frac{1}{m}\sum_{j=l}^{m}\left(\log j - \frac{m}{m-l+1}\log m + 1 + o(1)\right)^2 \\ &= \frac{1}{m}\sum_{j=l}^{m}\left(\log j - \frac{m}{m-l+1}\log m + 1\right)^2 \\ &\quad + o\left(\frac{1}{m}\sum_{j=l}^{m}\left(\log j - \frac{m}{m-l+1}\log m + 1\right)^2\right) + o(1).\end{aligned} \tag{A.3}$$

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

The first term of (A.3) is equal to

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

by A3. Hence, (A.3) is equal to $1+o(1)$. ■

LEMMA A.2. *Under A2 and A3,*

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

Proof. Using results in the proof of Lemma 1 of HDB (pp. 37–38) and A2,

$$
\sum_{j=l}^{m} a_j \log f_j^* = (1/2)(f^{*\prime\prime}(0)/f^*(0))\sum_{j=l}^{m} a_j \lambda_j^2 + R,
$$

where $R = O\left(T^{-3}m^4\log m\right)$. From Hurvich and Beltrao (1994, pp. 299–301), $a_j = \log j - (m-l+1)^{-1}\sum_{k=l}^{m}\log k + O\left(m^2/T^2\right)$, uniformly in j. Also note that

$$
\sum_{j=l}^{m} j^2 \log j = (1/6)m(m+1)(2m+1)\log m - \left(m^3/9\right) + o\left(m^3\right)
$$

by HDB (p. 38) and A3, $\sum_{j=l}^{m} j^2 = (1/6)m(m+1)(2m+1) - (1/6)l(l-1)(2l-1)$, and

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

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

by Stirling's formula and A3. Using these results, we have

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

by A3. Thus Lemma A.1 and A3 provide

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

■

Note that the normalized periodogram can be expressed as follows:

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

where $A_j \equiv (2\pi T)^{-1/2}\sum_{t=1}^{T} x_t \cos\left(\lambda_j t\right)$ and $B_j \equiv (2\pi T)^{-1/2}\sum_{t=1}^{T} x_t \sin\left(\lambda_j t\right)$. Define the frequency-dependent vectors $\gamma = (A_j/f_j^{1/2}, B_j/f_j^{1/2}, A_k/f_k^{1/2}, B_k/f_k^{1/2})'$ and $\widetilde{\gamma} = (\gamma_1, \gamma_2)'$. The next lemma details how well γ and $\widetilde{\gamma}$ can be approximated by multivariate Gaussian random variables under A1. It will be used later to derive Edgeworth approximations.

LEMMA A.3. *Under A1 and A4, 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 $n > 2$: $\kappa\left(\gamma_1^{n_1}, \gamma_2^{n_2}, \gamma_3^{n_3}, \gamma_4^{n_4}\right) = O\left(T^{n/2-nd}/(j^{(1-d)(n_1+n_3)}k^{(1-d)(n_2+n_4)})\right)$, where $\kappa(X_1, \ldots, X_4)$ denotes the joint cumulant of random variables $X_1, \ldots, X_4$ and $n_1, \ldots, n_4$ are nonnegative integers that sum to n. Similarly, for $n > 2$, the nth cumulants of $\widetilde{\gamma}$ are $O(T^{n/2-nd}/j^{n-nd})$.*

Proof. We provide the proof under A1(a); the others are similar. Note that

$$A_j = (2\pi T)^{-1/2}\sum_{t=1}^{T} v_t \cos\left(\lambda_j t\right) + (2\pi T)^{-1/2}\sum_{t=1}^{T} u_t \cos\left(\lambda_j t\right) \equiv A_j^v + A_j^u$$

and similarly for A_k, B_j, and B_k. By definition of κ, we have

$$\kappa\left(X_1, \ldots, X_4\right) = \sum_{\mathcal{P}}(|\mathcal{P}|-1)!(-1)^{|\mathcal{P}|-1}\prod_{\mathcal{B}\in\mathcal{P}}\mathbb{E}\left(\prod_{i\in\mathcal{B}} X_i\right), \tag{A.4}$$

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

where $\mathcal{P}$ runs through all partitions of $\{1,\ldots,4\}$, $\mathcal{B}$ runs through all blocks of the partition $\mathcal{P}$, and $|\cdot|$ denotes the number of elements in a set. Using the independence of $\{v_t\}$ and $\{u_t\}$ and the properties of cumulants, we have for any nth joint cumulant of the entries of γ:

$$\begin{aligned}
&\kappa\left(\gamma_1^{n_1},\gamma_2^{n_2},\gamma_3^{n_3},\gamma_4^{n_4}\right)\\
&\quad=\kappa\left(\left[A_j^v f_j^{-1/2}\right]^{n_1},\left[A_k^v f_k^{-1/2}\right]^{n_2},\left[B_j^v f_j^{-1/2}\right]^{n_3},\left[B_k^v f_k^{-1/2}\right]^{n_4}\right)\\
&\qquad+\kappa\left(\left[A_j^u f_j^{-1/2}\right]^{n_1},\left[A_k^u f_k^{-1/2}\right]^{n_2},\left[B_j^u f_j^{-1/2}\right]^{n_3},\left[B_k^u f_k^{-1/2}\right]^{n_4}\right)\\
&\quad=f_j^{-(n_1+n_3)/2}f_k^{-(n_2+n_4)/2}\kappa\left(\left[A_j^u\right]^{n_1},\left[A_k^u\right]^{n_2},\left[B_j^u\right]^{n_3},\left[B_k^u\right]^{n_4}\right). \qquad \textbf{(A.5)}
\end{aligned}$$

The second equality follows from A4. Upon inspection of (A.4), it becomes apparent that what is of concern are sums of products over $\mathcal{B}\in\mathcal{P}$ of terms of the form

$$\mathbb{E}\left[(A_j^u)^{r_1(\mathcal{B})}(A_k^u)^{r_2(\mathcal{B})}(B_j^u)^{r_3(\mathcal{B})}(B_k^u)^{r_4(\mathcal{B})}\right], \qquad \textbf{(A.6)}$$

where $r_i(\mathcal{B})\equiv n_i\mathbb{I}(i\in\mathcal{B})$. Let $\sum_{i=1}^4 r_i(\mathcal{B})\equiv R(\mathcal{B})\in[0,n]$ for each $\mathcal{B}\in\mathcal{P}$. Now consider as given block $\mathcal{B}$ of a fixed partition $\mathcal{P}$. Suppressing dependence of quantities upon $\mathcal{B}$, and scaling (A.6) by $j^{r_1+r_3}k^{r_2+r_4}/T^{R/2}$, we obtain

$$\begin{aligned}
&\frac{j^{r_1+r_3}k^{r_2+r_4}}{(2\pi)^{R/2}T^R}\sum_{t_1=1}^{T}\cdots\sum_{t_R=1}^{T}\mathbb{E}\left[\prod_{i=1}^{R}u_{t_i}\right]\prod_{i=1}^{r_1}\cos(\lambda_j t_i)\prod_{i=r_1+1}^{r_1+r_2}\cos(\lambda_k t_i)\\
&\qquad\prod_{i=r_1+r_2+1}^{R-r_4}\sin(\lambda_j t_i)\prod_{i=R-r_4+1}^{R}\sin(\lambda_k t_i). \qquad \textbf{(A.7)}
\end{aligned}$$

The expectation inside of the summations can be decomposed as follows. Without loss of generality, suppose $t_1\le t_2\le\cdots\le t_R$. Then, with $u_{t_0}\equiv 0$,

$$\mathbb{E}\left[\prod_{i=1}^{R}u_{t_i}\right]=\mathbb{E}\left[\prod_{i=1}^{R}\sum_{k=1}^{i}(u_{t_k}-u_{t_{k-1}})\right]=\sum_{k_R=1}^{R}\sum_{k_{R-1}=1}^{R-1}\cdots\sum_{k_1=1}^{1}\mathbb{E}\left[\prod_{j=1}^{R}\left(u_{t_{k_j}}-u_{t_{k_j-1}}\right)\right]. \qquad \textbf{(A.8)}$$

Because $\{u_t\}$ is a mean zero process with independent differences, this is a finite sum of terms of the form $\prod_{k=1}^{R}\mathbb{E}[(u_{t_k}-u_{t_{k-1}})^{\alpha_k}]$, where the α_k's are nonnegative integers that sum to R. Also, $(u_{t_k}-u_{t_{k-1}})\overset{d}{\sim}u_{t_k-t_{k-1}}$ so that $\mathbb{E}[(u_{t_k}-u_{t_{k-1}})^{\alpha_k}]=\mathbb{E}[u_{t_k-t_{k-1}}^{\alpha_k}]$. Now, letting $\widetilde{N}_t=\sum_{i=1}^{t}\pi_i$, for any nonnegative integer α, we have

$$\begin{aligned}
\mathbb{E}[u_t^\alpha]&=\mathbb{E}\left[\left(\sum_{j=1}^{\widetilde{N}_t}\eta_j\right)^\alpha\right]=\mathbb{E}\left[\sum_{j_1=1}^{\widetilde{N}_t}\cdots\sum_{j_\alpha=1}^{\widetilde{N}_t}\mathbb{E}\left[\eta_{j_1}\ldots\eta_{j_\alpha}|\widetilde{N}_t\right]\right]\\
&=C\mathbb{E}\left[\widetilde{N}_t^\alpha\right]=C\sum_{i_1=1}^{t}\cdots\sum_{i_\alpha=1}^{t}\mathbb{E}[\pi_{i_1}\ldots\pi_{i_\alpha}], \qquad \textbf{(A.9)}
\end{aligned}$$

where C is a finite constant because all of the moments of η_i exist and η_i and $\widetilde{N}_t$ are independent for all i and t. A1(a) provides that the term $\mathbb{E}[\pi_{i_1}\ldots\pi_{i_\alpha}]$ is a polynomial in

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

(t/T) of maximum order α so that (A.9) is also. In turn, (A.8) is a finite sum of a finite product of polynomials in the increments $(t_k/T - t_{k-1}/T)$, each of degree less than or equal to R. Thus, we can deduce that the term (A.8) is a function of $t_1/T, \ldots, t_R/T$ that is bounded for all $t_i \in \{1, \ldots, T\}$. Denoting this function as $g(t_1/T, \ldots, t_R/T)$, we can further deduce that, for any finite constant c, $g(c, \ldots, c, t_i/T, c, \ldots, c)$ is a finite-order polynomial in t_i/T. We can thus analyze the terms composing (A.7) as follows. Using the fact that $x^\alpha \cos(2\pi jx) = \partial(x^\alpha \sin(2\pi jx)/(2\pi j))/\partial x - \alpha x^{\alpha-1} \sin(2\pi jx)/(2\pi j)$,

$$
\begin{aligned}
&(j/T)(2\pi)^{-1/2} \sum_{t=1}^{T} (t/T)^\alpha \cos(2\pi jt/T) \\
&\quad = (j/T)(2\pi)^{-1/2} \sum_{t=1}^{T} \Big\{ \left[((t+1)/T)^\alpha \sin(2\pi j(t+1)/T) - (t/T)^\alpha \sin(2\pi jt/T) \right] / \\
&\qquad\qquad (2\pi j/T) + o\left(T^{-1}\right) \Big\} \\
&\qquad - (j/T)(2\pi)^{-1/2} \sum_{t=1}^{T} \alpha (t/T)^{\alpha-1} \sin(2\pi jt/T)/(2\pi j) \\
&\quad = (2\pi)^{-3/2} \left\{ ((T+1)/T)^\alpha \sin(2\pi j(T+1)/T) - T^{-\alpha} \sin(2\pi j/T) \right\} + o(j/T) \\
&\qquad - \alpha(2\pi)^{-3/2} T^{-1} \sum_{t=1}^{T} (t/T)^{\alpha-1} \sin(2\pi jt/T) = O(1)
\end{aligned}
$$

for $\alpha > 0$. Analogous results hold when $\alpha = 0$ and for the other terms in (A.7). Hence, (A.7) is $O(1)$ and (A.6) is $O(T^{R/2}/j^{r_1+r_3}k^{r_2+r_4})$. From earlier reasoning derived from (A.4),

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

Finally, from (A.5) we then have

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

■

We now present a lemma that applies to all DGPs in A1, using ε_j defined by (A.2). In what follows, mention of uniformity refers to results holding uniformly for $l \le k < j \le m$.

LEMMA A.4. *Under A1–A4, we have uniformly*

$$
\mathrm{Cov}(\varepsilon_j, \varepsilon_k) = O\left(\frac{\log^2 j}{k^2} + \frac{T^{1-2d} \log j}{j^{1-d} k^{2-d}} + \frac{T^{2-4d}}{j^{2-2d} k^{2-2d}} + \frac{T^{2-4d} \log j}{k^{5-4d}} + \frac{T^{3-6d}}{k^{6-6d}} \right).
$$

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

Proof. The proof involves the use of Lemma A.3 via an Edgeworth expansion. Define $\chi_j = \log(I_j/f_j) - \mathbb{E}[\log(I_j/f_j)]$ and $\psi = \Sigma^{-1}$, where $\Sigma = \mathrm{Cov}(\gamma)$. 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 density because, under A3, 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 and, for four-dimensional g, $\phi_\psi(g) = (2\pi)^{-2}|\psi|^{1/2}\exp\left(-g'\psi g/2\right)$ (the multivariate zero-mean normal density with covariance matrix Σ, the second-order expansion provides (see Skovgaard, 1986, p. 172)

$$f_\gamma(g) \approx \phi_\psi(g)\left\{1 + \sum_{1\le m,n,r\le 4} \kappa(\gamma_m,\gamma_n,\gamma_r) \times \left[(1/6)\sum_{1\le\alpha,\beta,\delta\le 4}\psi_{m\alpha}\psi_{n\beta}\psi_{r\delta}g_\alpha g_\beta g_\delta - (/2)\psi_{mn}\sum_{\alpha=1}^{4}\psi_{r\alpha}g_\alpha\right]\right\}.$$

From this Edgeworth expansion, we obtain

$$\mathbb{E}\left[\chi_j\chi_k\right] \approx (2\pi)^{-2}|\psi|^{1/2}\int \chi_j\chi_k\exp\left(-g'\psi g/2\right)dg \tag{A.10}$$

plus terms of the form

$$(1/6)(2\pi)^{-2}|\psi|^{1/2}\kappa(\gamma_m,\gamma_n,\gamma_r)\psi_{m\alpha}\psi_{n\beta}\psi_{r\delta}\int \chi_j\chi_k g_\alpha g_\beta g_\delta\exp\left(-g'\psi g/2\right)dg, \tag{A.11}$$

$$-(1/2)(2\pi)^{-2}|\psi|^{1/2}\kappa(\gamma_m,\gamma_n,\gamma_r)\psi_{mn}\psi_{r\alpha}\int \chi_j\chi_k g_\alpha\exp\left(-g'\psi g/2\right)dg \tag{A.12}$$

uniformly. Applying Theorem 1, by Lemmas 2 and 3 of HDB (with minor modification), (A.10) is $O((\log^2 j)/k^2 + T^{1-2d}\log j/(j^{1-d}k^{2-d}) + T^{2-4d}/(j^{2-2d}k^{2-2d}))$ uniformly. Turning now to (A.11) and (A.12), partition ψ as

$$\psi = \begin{bmatrix}\psi_{11}^P & \psi_{12}^P\\ \psi_{21}^P & \psi_{22}^P\end{bmatrix} \quad \text{and let} \quad \widetilde{\psi} = \begin{bmatrix}\psi_{11}^P & 0\\ 0 & \psi_{22}^P\end{bmatrix}$$

and $\overline{\psi} = \psi - \widetilde{\psi}$, where ψ_{ij}^P are 2×2 submatrices. Adapting Theorem 1 and Robinson (1995),

$$\mathbb{E}\left[\frac{A_j^2}{f_j}\right], \mathbb{E}\left[\frac{B_j^2}{f_j}\right] = \frac{1}{2} + O\left(\frac{\log j}{j}\right) + O\left(\frac{T^{1-2d}}{j^{2-2d}}\right),$$

$$\mathbb{E}\left[\frac{w_x(\lambda_j)^2}{f_j}\right] = \mathbb{E}\left[\frac{A_j^2 - 2iA_jB_j - B_j^2}{f_j}\right] = O\left(\frac{\log j}{j}\right) + O\left(\frac{T^{1-2d}}{j^{2-2d}}\right),$$

$$\mathbb{E}\left[\frac{w_x(\lambda_j)w_x(\lambda_k)}{f_j^{1/2}f_k^{1/2}}\right] = \mathbb{E}\left[\frac{A_jA_k - iA_jB_k - iA_kB_j - B_jB_k}{f_j^{1/2}f_k^{1/2}}\right] = O\left(\frac{\log j}{k}\right) + O\left(\frac{T^{1-2d}}{j^{1-d}k^{1-d}}\right),$$

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

$$\mathbb{E}\left[\frac{w_x(\lambda_j)w_x(\lambda_k)^*}{f_j^{1/2}f_k^{1/2}}\right] = \mathbb{E}\left[\frac{A_jA_k + iA_jB_k - iA_kB_j + B_jB_k}{f_j^{1/2}f_k^{1/2}}\right]$$
$$= O\left(\frac{\log j}{k}\right) + O\left(\frac{T^{1-2d}}{j^{1-d}k^{1-d}}\right),$$

so that we may conclude $\psi = 2I_4 + R(T)$, where the entries of $R(T)$ are of order $O(k^{-1}\log k) + O(T^{1-2d}/k^{2-2d})$, $O(k^{-1}\log j) + O(T^{1-2d}/(j^{1-d}k^{1-d}))$, or smaller. Similarly, the entries of $\overline{\psi}$ are of order $O(k^{-1}\log j) + O(T^{1-2d}/(j^{1-d}k^{1-d}))$. As an aside, let $n_1, \ldots, n_4$ denote nonnegative integers that sum to some odd number. Then the quantity

$$\int \chi_j\chi_k \exp\left(-g'\widetilde{\psi}g/2\right) g_1^{n_1}g_2^{n_2}g_3^{n_3}g_4^{n_4}dg \tag{A.13}$$

is proportional to $\mathbb{E}_{\widetilde{\psi}}^{\sim}[\chi_j\chi_k\gamma_1^{n_1}\gamma_2^{n_2}\gamma_3^{n_3}\gamma_4^{n_4}]$, where $\mathbb{E}_{\widetilde{\psi}}^{\sim}[\cdot]$ denotes the expectation operator assuming that γ is multivariate normal with mean zero and covariance matrix $\widetilde{\psi}^{-1}$. Under this assumption, the vectors $(\gamma_1, \gamma_2)'$ and $(\gamma_3, \gamma_4)'$ are independent. Hence,

$$\mathbb{E}_{\widetilde{\psi}}^{\sim}\left[\chi_j\chi_k\gamma_1^{n_1}\gamma_2^{n_2}\gamma_3^{n_3}\gamma_4^{n_4}\right] = \mathbb{E}_{\widetilde{\psi}}^{\sim}\left[\chi_j\gamma_1^{n_1}\gamma_2^{n_2}\right]\mathbb{E}_{\widetilde{\psi}}^{\sim}\left[\chi_k\gamma_3^{n_3}\gamma_4^{n_4}\right] = 0$$

because either $\mathbb{E}_{\widetilde{\psi}}^{\sim}[\chi_j\gamma_1^{n_1}\gamma_2^{n_2}] = 0$ or $\mathbb{E}_{\widetilde{\psi}}^{\sim}[\chi_k\gamma_3^{n_3}\gamma_4^{n_4}] = 0$. Hence, (A.13) is equal to zero. The same holds if $n_i = 0$ for $i = 1, \ldots, 4$. Now note that

$$\int \chi_j\chi_k g_\alpha g_\beta g_\delta \exp\left(-g'\psi g/2\right) dg$$
$$= \int \chi_j\chi_k \exp\left(-g'\widetilde{\psi}g/2\right) g_\alpha g_\beta g_\delta dg$$
$$\int \chi_j\chi_k g_\alpha g_\beta g_\delta \exp\left(-g'\widetilde{\psi}g/2\right)\left\{\exp\left(-g'\widetilde{\psi}g/2\right) - 1\right\} dg. \tag{A.14}$$

The first term of (A.14) takes the form of (A.13) and is thus zero. Following similar arguments to those on page 40 of HDB, the second term on the right-hand side of (A.14) is

$$\int \chi_j\chi_k g_\alpha g_\beta g_\delta \exp\left(-g'\widetilde{\psi}g/2\right)\exp\left(-g'\widetilde{\psi}g/2\right) dg$$
$$+ O(\Psi^2 \int \|g\|^4 |\chi_j\chi_k| |g_\alpha g_\beta g_\delta| \exp\left(-g'(\widetilde{\psi} - 4\Psi I_4)g/2\right) dg), \tag{A.15}$$

where Ψ is the largest absolute entry of $\overline{\psi}$. The first term of (A.15) is composed of a linear combination of terms of the form $\mathbb{E}_{\widetilde{\psi}}^{\sim}[\chi_j\chi_k\gamma_1^{n_1}\gamma_2^{n_2}\gamma_3^{n_3}\gamma_4^{n_4}]$, where $\sum_{i=1}^4 n_i = 5$. Thus, this term is zero because (A.13) is. Using the results of Theorem 1, Ψ is $O((\log j)/k + T^{1-2d}/(j^{1-d}k^{1-d}))$ uniformly, and $\widetilde{\psi} - 4\Psi I_4 = 2I_4 + o(1)$. Hence, the second term of (A.15) is

$$O\left(\left(\log^2 j\right)/k^2 + T^{1-2d}\log j/\left(j^{1-d}k^{2-d}\right) + T^{2-4d}\Big/\left(j^{2-2d}k^{2-2d}\right)\right)$$

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

uniformly so that (A.14) is also. Lemma A.3 tells us that $\kappa(\gamma_m, \gamma_n, \gamma_r)$ is $O(T^{3/2-3d}/k^{3-3d}) = o(1)$ so that (A.11) is $o((\log^2 j)/k^2 + T^{1-2d}\log j/(j^{1-d}k^{2-d}) + T^{2-4d}/(j^{2-2d}k^{2-2d}))$ uniformly. Very similar arguments show that (A.12) is too. Thus, (A.10)–(A.12) are $O(k^{-2}\log^2 j + T^{1-2d}\log j/(j^{1-d}k^{2-d}) + T^{2-4d}/(j^{2-2d}k^{2-2d}))$ uniformly. The higher order terms for approximating $\mathbb{E}[\chi_j\chi_k]$ via a third-order Edgeworth expansion take the following forms:

$$(1/24)(2\pi)^{-2}|\psi|^{1/2}\kappa(\gamma_m,\gamma_n,\gamma_r,\gamma_q)\psi_{m\alpha}\psi_{n\beta}\psi_{r\xi}\psi_{q\delta} \times \int \chi_j\chi_k g_\alpha g_\beta g_\xi g_\delta \exp\left(-g'\psi g/2\right) dg, \tag{A.16}$$

$$-(1/4)(2\pi)^{-2}|\psi|^{1/2}\kappa(\gamma_m,\gamma_n,\gamma_r,\gamma_q)\psi_{m\alpha}\psi_{n\beta}\psi_{rq}\int \chi_j\chi_k g_\alpha g_\beta \exp\left(-g'\psi g/2\right) dg, \tag{A.17}$$

$$(1/8)(2\pi)^{-2}|\psi|^{1/2}\kappa(\gamma_m,\gamma_n,\gamma_r,\gamma_q)\psi_{mn}\psi_{rq}\int \chi_j\chi_k \exp\left(-g'\psi g/2\right) dg, \tag{A.18}$$

plus terms that take the form

$$C\kappa(\gamma_m,\gamma_n,\gamma_r)\kappa(\gamma_q,\gamma_s,\gamma_u)\int \chi_j\chi_k g_1^{n_1}g_2^{n_2}g_3^{n_3}g_4^{n_4}\exp\left(-g'\psi g/2\right) dg$$

for some finite constant C and nonnegative integers $n_1,\ldots,n_4$. These latter terms are $O(T^{3-6d}/k^{6-6d})$ uniformly because, from Lemma A.3, $\kappa(\gamma_m,\gamma_n,\gamma_r) = O(T^{3/2-3d}/k^{3-3d})$ and

$$(2\pi)^{-2}|\psi|^{1/2}\int \chi_j\chi_k g_1^{n_1}g_2^{n_2}g_3^{n_3}g_4^{n_4}\exp\left(-g'\psi g/2\right) dg < \infty \tag{A.19}$$

uniformly for any nonnegative integers $n_1,\ldots,n_4$ by the properties of the multivariate Gaussian distribution. Using the same technique used for (A.14) and (A.15), (A.16) is equal to

$$(1/24)(2\pi)^{-2}|\psi|^{1/2}\kappa(\gamma_m,\gamma_n,\gamma_r,\gamma_q)\psi_{m\alpha}\psi_{n\beta}\psi_{r\xi}\psi_{q\delta} \times \int \chi_j\chi_k g_\alpha g_\beta g_\xi g_\delta \exp\left(-g'\widetilde{\psi}g/2\right) dg \tag{A.20}$$

$$+ O(\kappa(\gamma_m,\gamma_n,\gamma_r,\gamma_q))\int \chi_j\chi_k g_\alpha g_\beta g_\xi g_\delta \exp\left(-g'\widetilde{\psi}g/2\right)\left(-g'\overline{\psi}g/2\right) dg \tag{A.21}$$

$$+ O(\kappa(\gamma_m,\gamma_n,\gamma_r,\gamma_q)\Psi^2\int \|g\|^4|\chi_j\chi_k||g_\alpha g_\beta g_\xi g_\delta|\exp\left(-g'(\widetilde{\psi}-4\Psi I_4)g/2\right) dg), \tag{A.22}$$

where Ψ is the largest absolute entry of $\overline{\psi}$. Now, the integral inside of (A.20) can only be nonzero when two, and only two, elements of the set $\{\alpha,\beta,\xi,\delta\}$ are in $\{1,2\}$. To see this, note that if every element of $\{\alpha,\beta,\xi,\delta\}$ is in $\{1,2\}$ or $\{3,4\}$, this integral is proportional to $\mathbb{E}_{\widetilde{\psi}}[\chi_j]\mathbb{E}_{\widetilde{\psi}}[\chi_k\gamma_\alpha\gamma_\beta\gamma_\xi\gamma_\delta] = 0$ or $\mathbb{E}_{\widetilde{\psi}}[\chi_j\gamma_\alpha\gamma_\beta\gamma_\xi\gamma_\delta]\mathbb{E}_{\widetilde{\psi}}[\chi_k] = 0$. Similarly, if only one element of $\{\alpha,\beta,\xi,\delta\}$ is in $\{1,2\}$ or $\{3,4\}$, this integral is proportional to, e.g., $\mathbb{E}_{\widetilde{\psi}}[\chi_j\gamma_\alpha]\mathbb{E}_{\widetilde{\psi}}[\chi_k\gamma_\beta\gamma_\xi\gamma_\delta] = 0$. To determine the asymptotic order of (A.20) when two, and only two, elements of the set $\{\alpha,\beta,\xi,\delta\}$ are in $\{1,2\}$, we must consider the possible orders that $\kappa(\gamma_m,\gamma_n,\gamma_r,\gamma_q)$ and $\psi_{m\alpha}\psi_{n\beta}\psi_{r\xi}\psi_{q\delta}$ can simultaneously take given the results of Lemma A.3 and that

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

$\psi = 2I_4 + R(T)$. First, if all four terms of $\psi_{m\alpha}\psi_{n\beta}\psi_{r\xi}\psi_{q\delta}$ are on the main diagonal of ψ, $\kappa(\gamma_m,\gamma_n,\gamma_r,\gamma_q) = \kappa(\gamma_\alpha,\gamma_\beta,\gamma_\xi,\gamma_\delta) = O(T^{2-4d}/(j^{2-2d}k^{2-2d}))$ and $\psi_{m\alpha}\psi_{n\beta}\psi_{r\xi}\psi_{q\delta} = O(1)$ so that (A.20) is $O(T^{2-4d}/(j^{2-2d}k^{2-2d}))$ uniformly. Second, if three of four terms of $\psi_{m\alpha}\psi_{n\beta}\psi_{r\xi}\psi_{q\delta}$ are on the main diagonal, the maximum order of $\kappa(\gamma_m,\gamma_n,\gamma_r,\gamma_q)$ is $O(T^{2-4d}/(j^{1-d}k^{3-3d}))$ and that of $\psi_{m\alpha}\psi_{n\beta}\psi_{r\xi}\psi_{q\delta}$ is $O(k^{-1}\log k + T^{1-2d}/k^{2-2d})$ so that the maximum order of (A.20) is $O(k^{-1}\log k + T^{1-2d}/k^{2-2d})O(T^{2-4d}/(j^{1-d}k^{3-3d}))$. Third, if two of the terms are on the main diagonal, the maximum order of (A.20) is $O(k^{-1}\log k + T^{1-2d}/k^{2-2d})^2 O(T^{2-4d}/k^{4-4d})$. Finally, when one or zero terms are on the main diagonal, the maximum order of (A.20) is $o(k^{-1}\log k + T^{1-2d}/k^{2-2d})^2 O(T^{2-4d}/k^{4-4d})$. Turning to (A.21), $g'\overline{\psi}g$ is a finite sum of terms of the form $\psi_{mn}g_\alpha g_\beta$, where $\psi_{mn} = O(k^{-1}\log j + T^{1-2d}/(j^{1-d}k^{1-d}))$ by Lemma A.3. Hence, the maximum order of (A.21) is $O(T^{2-4d}/k^{4-4d})$ $O(k^{-1}\log j + T^{1-2d}/(j^{1-d}k^{1-d}))$. Similar to prior cases, (A.22) is $o(k^{-2}\log^2 j + T^{1-2d}\log j/(j^{1-d}k^{2-d}) + T^{2-4d}/(j^{2-2d}k^{2-2d}))$ uniformly. Using similar decompositions, (A.17) is $O(T^{2-4d}/k^{4-4d})O(k^{-1}\log j + T^{1-2d}/(j^{1-d}k^{1-d})) + o(k^{-2}\log^2 j + T^{1-2d}\log j/(j^{1-d}k^{2-d}) + T^{2-4d}/(j^{2-2d}k^{2-2d}))$ and (A.18) is $o(k^{-2}\log^2 j + j^{d-1}k^{d-2}T^{1-2d}\log j + j^{2d-2}k^{2d-2}T^{2-4d})$, uniformly.

To summarize, the addition of terms (A.10)–(A.12) and (A.16)–(A.18), provides the third-order Edgeworth approximation to $\mathbb{E}[\chi_j\chi_k]$. The remainder to this approximation has a maximum order of $O(T^{3-6d}/k^{6-6d})$. To see this, note that the terms in the Edgeworth expansion following the third-order approximation will either take the form of $C\kappa(\gamma_m,\gamma_n,\gamma_r,\gamma_q,\gamma_s)$ times an integral of the form given by (A.19) with $n_1,\ldots,n_4$ summing to an odd number or a product of cumulants with multiplied order of $O(T^{3-6d}/k^{6-6d})$ (by Lemma A.3) times a finite constant and an integral of the form given by (A.19). To see that the former is $O(T^{3-6d}/k^{6-6d})$, note that $\kappa(\gamma_m,\gamma_n,\gamma_r,\gamma_q,\gamma_s) = O(T^{5/2-5d}/k^{5-5d})$ by Lemma A.3 and (A.19) with $n_1,\ldots,n_4$ summing to an odd number is $O((\log^2 j)/k^2 + T^{1-2d}\log j/(j^{1-d}k^{2-d}) + T^{2-4d}/(j^{2-2d}k^{2-2d}))$ by nearly identical arguments to those showing the same for (A.14). Therefore by A3, uniformly

$$
\begin{aligned}
\mathbb{E}[\chi_j\chi_k] &= O\left(\left(\log^2 j\right)/k^2 + T^{1-2d}\log j/\left(j^{1-d}k^{2-d}\right) + T^{2-4d}/\left(j^{2-2d}k^{2-2d}\right)\right)\\
&\quad + O\left(k^{-1}\log k + k^{2d-2}T^{1-2d}\right)O\left(j^{d-1}k^{3d-3}T^{2-4d}\right)\\
&\quad + O\left(k^{-1}\log k + k^{2d-2}T^{1-2d}\right)^2 O\left(k^{4d-4}T^{2-4d}\right)\\
&\quad + O\left(k^{-1}\log j + T^{1-2d}/(j^{1-d}k^{1-d})\right)O\left(T^{2-4d}/k^{4-4d}\right)\\
&\quad + O\left(T^{3-6d}/k^{6-6d}\right)\\
&= O\left(k^{-2}\log^2 j + j^{d-1}k^{d-2}T^{1-2d}\log j + j^{2d-2}k^{2d-2}T^{2-4d}\right.\\
&\qquad \left. + k^{4d-5}T^{2-4d}\log j + k^{6d-6}T^{3-6d}\right).
\end{aligned}
$$

■

LEMMA A.5. *Under A1–A4,* $\mathbb{E}[\varepsilon_j] = O(j^{-1}\log j) + O\left(T^{1-2d}/j^{2-2d}\right)$ *uniformly.*

Proof. To begin, note that the same arguments used in Lemma A.4 allow us to conclude that $\widetilde{\Sigma} \equiv \mathrm{Cov}(\widetilde{\gamma}) = (1/2)I_2 + O\left(j^{-1}\log j\right) + O\left(j^{2d-2}T^{1-2d}\right)$. Let $\hat{\psi} = \widetilde{\Sigma}^{-1}$. Then

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

we can use Lemma A.3 and a second-order Edgeworth expansion to obtain

$$\mathbb{E}[\varepsilon_j] = (2\pi)^{-1}|\hat{\psi}|^{1/2}\int \varepsilon_j \exp(-g'\hat{\psi}g/2)dg + O\left(j^{3d-3}T^{3/2-3d}\right)(2\pi)^{-1}|\hat{\psi}|^{1/2}$$
$$\times \int \varepsilon_j \exp(-g'\hat{\psi}g/2)\left(\sum_{1\le m,n,r\le 2} g_m g_n g_r + \sum_{1\le m\le 2} g_m\right)dg$$
$$+ O\left(j^{4d-4}T^{2-4d}\right). \tag{A.23}$$

Using (A.19), the last two terms of (A.23) are $O(T^{3/2-3d}/j^{3-3d})$ uniformly. Turning to the first term, note that $\hat{\psi} = 2I_2 + R_j(T)$, where $R_j(T)$ is some some matrix with entries that are $O(j^{-1}\log j) + O(T^{1-2d}/j^{2-2d})$ uniformly. Hence, the integral inside the first term is equal to

$$\int \varepsilon_j \exp(-g'g)\exp\left(-g'R_j(T)g/2\right)dg. \tag{A.24}$$

By the mean value theorem, $|e^u - 1| \le |u|e^{|u|}$ so that

$$\left|\exp\left(-g'R_j(T)g/2\right) - 1\right| \le \left|g'R_j(T)g/2\right|\exp\left(\left|g'R_j(T)g/2\right|\right).$$

Hence, because $|g'R_j(T)g/2| \le 3Q_j(T)\|g\|^2/2$, with $Q_j(T)$ the largest absolute entry of $R_j(T)$,

$$\exp\left(-g'R_j(T)g/2\right) = 1 + O\left(Q_j(T)\|g\|^2\exp\left((3/2)Q_j(T)\|g\|^2\right)\right),$$

and (A.24) is equal to

$$\int \varepsilon_j \exp(-g'g)dg + O(Q_j(T)\int \varepsilon_j\|g\|^2\exp(-g'[2-3Q_j(T)]I_2g/2)dg).$$

Because $Q_j(T) = O(j^{-1}\log j) + O(T^{1-2d}/j^{2-2d})$, $[2-3Q_j(T)]I_2 = 2I_2 + o(1)$ uniformly so that the second term of the preceding expression is $O(Q_j(T)) = O(j^{-1}\log j) + O(T^{1-2d}/j^{2-2d})$ uniformly. Putting these results together, (A.23) is equal to

$$\left((1/2)|\hat{\psi}|^{1/2}\right)(1/\pi)\int \varepsilon_j \exp(-g'g)dg + O\left(j^{-1}\log j\right) + O\left(j^{2d-2}T^{1-2d}\right) \tag{A.25}$$

uniformly. The first term of this expression is $|\hat{\psi}|^{1/2}/2$ times the expectation of $\log(\gamma_1^2 + \gamma_2^2) + C$ under the assumption that γ_1 and γ_2 are independent standard normal random variables. Under this assumption, $-\log(\gamma_1^2 + \gamma_2^2)$ has a Gumbel distribution with parameters zero and one. The mean of this distribution is known to be C so that the first term of (A.25) is identically zero because $|\hat{\psi}| = 4 + o(1)$ uniformly. This then implies the lemma's claim. ■

LEMMA A.6. *Under A1–A4,* $\mathrm{Var}(\varepsilon_j) = \pi^2/6 + O(j^{-1}\log j) + O(T^{1-2d}/j^{2-2d})$ *uniformly.*

Proof. This is very similar to the proof of Lemma A.5. ■

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

LEMMA A.7. *Under A1–A4,*

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

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

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

■

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

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

$$\mathrm{Var}(\hat{d}) = (4S_{YY}^2)^{-1}\sum_{j=l}^{m} a_j^2 \mathrm{Var}\left(\varepsilon_j\right) + (2S_{YY}^2)^{-1}\sum_{k=l}^{m}\sum_{j=k+1}^{m} a_j a_k \mathrm{Cov}\left(\varepsilon_j, \varepsilon_k\right). \tag{A.26}$$

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

$$\begin{aligned}&\sum_{j=l}^{m} a_j^2 \mathrm{Var}\left(\varepsilon_j\right) + 2\sum_{k=l}^{m}\sum_{j=k+1}^{m} a_j a_k \mathrm{Cov}\left(\varepsilon_j, \varepsilon_k\right)\\ &\quad= \sum_{j=l}^{m} a_j^2\left\{(\pi^2/6) + O\left(j^{-1}\log j\right) + O\left(T^{1-2d}/j^{2-2d}\right)\right\}\\ &\qquad + O\left(\log^2 m \sum_{k=l}^{m}\sum_{j=k+1}^{m}\left[k^{-2}\log^2 j + T^{1-2d}\log j/(j^{1-d}k^{2-d})\right.\right.\\ &\qquad\qquad + T^{2-4d}/\left(j^{2-2d}k^{2-2d}\right)\\ &\qquad\qquad \left.\left. + T^{2-4d}(\log j)/k^{5-4d} + T^{3-6d}/k^{6-6d}\right]\right)\\ &\quad= \pi^2 m/6 + o(m) + O\left(l^{4d-2}T^{2-4d}\log^4 m\right) + O\left(l^{4d-4}T^{2-4d}m\log^4 m\right)\\ &\qquad + O\left(l^{6d-5}T^{3-6d}m\log^3 m\right).\end{aligned}$$

Using Lemma A.1 together with (A.26) shows part (ii). Part (iii) is a direct consequence of parts (i) and (ii). ■

We introduce an additional lemma to prove Theorem 3.

LEMMA A.8. *Under A3, the sequence* $\{a_j\}_{j=l}^{m}$ *satisfies* $\max_{l\le j\le m}|a_j| = o(m)$, $m^{-1}\sum_{k=l}^{m} a_k^2 \to 1$, $\sum_{k=l}^{m}|a_k|^p = O(m)$ *for all* $p \ge 1$.

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

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

Proof of Theorem 3. Let $\varepsilon_j^v = \log(I_v(\lambda_j)/f_j) + C$. By Lemma A.5 and Lemma 6 of HDB, $\mathbb{E}[\varepsilon_j] = \mathbb{E}[\varepsilon_j^v] + O\left(j^{-1}\log j\right) + O\left(T^{1-2d}/j^{2-2d}\right)$, uniformly. Hence,

$$\begin{aligned}\mathbb{E}\left[m^{-1/2}\sum_{j=l}^{m} a_j(\varepsilon_j - \varepsilon_j^v)\right] &= m^{-1/2}\sum_{j=l}^{m} a_j O\left(j^{-1}\log j + T^{1-2d}/j^{2-2d}\right)\\ &= O\left(m^{-1/2}\log^3 m\right) + O\left(m^{-1/2}l^{2d-1}T^{1-2d}\log^2 m\right)\\ &= o(1) \qquad \textbf{(A.27)}\end{aligned}$$

under A3*. A very similar proof to that of lemma A.6 shows that $\mathrm{Cov}(\varepsilon_j, \varepsilon_j^v) = \pi^2/6 + O(j^{-1}\log j) + O(T^{1-2d}/j^{2-2d})$ uniformly so that by Lemma A.6 and Lemma 7 of HDB,

$$\mathrm{Cov}(\varepsilon_j, \varepsilon_j^v), \mathrm{Var}(\varepsilon_j) = \mathrm{Var}(\varepsilon_j^v) + O\left(j^{-1}\log j\right) + O\left(T^{1-2d}/j^{2-2d}\right) \qquad \textbf{(A.28)}$$

uniformly. Likewise, a very similar proof to the Proof of Lemma A.4 shows that

$$\mathrm{Cov}(\varepsilon_j, \varepsilon_k^v), \mathrm{Cov}(\varepsilon_j^v, \varepsilon_k) = O\left(\frac{\log^2 j}{k^2} + \frac{T^{1-2d}\log j}{j^{1-d}k^{2-d}} + \frac{T^{2-4d}}{j^{2-2d}k^{2-2d}} + \frac{T^{2-4d}\log j}{k^{5-4d}} + \frac{T^{3-6d}}{k^{6-6d}}\right)$$

uniformly so that by Lemma A.4 and Lemmas 2 and 3 of HDB,

$$\begin{aligned}&\mathrm{Cov}(\varepsilon_j, \varepsilon_k), \mathrm{Cov}(\varepsilon_j, \varepsilon_k^v), \mathrm{Cov}(\varepsilon_j^v, \varepsilon_k)\\ &\quad = Cov(\varepsilon_j^v, \varepsilon_k^v) + O\left(\frac{\log^2 j}{k^2} + \frac{T^{1-2d}\log j}{j^{1-d}k^{2-d}} + \frac{T^{2-4d}}{j^{2-2d}k^{2-2d}}\right.\\ &\qquad\qquad \left. + \frac{T^{2-4d}\log j}{k^{5-4d}} + \frac{T^{3-6d}}{k^{6-6d}}\right) \qquad \textbf{(A.29)}\end{aligned}$$

uniformly. Hence, (A.28) and (A.29) provide

$$\begin{aligned}&\mathrm{Var}\left[m^{-1/2}\sum_{j=l}^{m} a_j(\varepsilon_j - \varepsilon_j^v)\right]\\ &\quad = m^{-1}\sum_{j=l}^{m} a_j^2\left\{\mathrm{Var}(\varepsilon_j) + \mathrm{Var}\left(\varepsilon_j^v\right) - 2\,\mathrm{Cov}\left(\varepsilon_j, \varepsilon_j^v\right)\right\}\\ &\qquad + 2m^{-1}\sum_{k=l}^{m}\sum_{j=k+1}^{m} a_j a_k\left\{\mathrm{Cov}(\varepsilon_j, \varepsilon_k) - \mathrm{Cov}(\varepsilon_j, \varepsilon_k^v) - \mathrm{Cov}(\varepsilon_j^v, \varepsilon_k) + \mathrm{Cov}(\varepsilon_j^v, \varepsilon_k^v)\right\}\\ &\quad = m^{-1}\sum_{j=l}^{m} a_j^2 O\left(j^{-1}\log j + T^{1-2d}/j^{2-2d}\right)\\ &\qquad + 2m^{-1}\sum_{k=l}^{m}\sum_{j=k+1}^{m} a_j a_k O\left(\frac{\log^2 j}{k^2} + \frac{T^{1-2d}\log j}{j^{1-d}k^{2-d}} + \frac{T^{2-4d}}{j^{2-2d}k^{2-2d}}\right.\\ &\qquad\qquad \left. + \frac{T^{2-4d}\log j}{k^{5-4d}} + \frac{T^{3-6d}}{k^{6-6d}}\right)\end{aligned}$$

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press

$$= O\left(\frac{T^{1-2d}\log^3 m}{ml^{1-2d}}\right) + O\left(\frac{\log^5 m}{l}\right) + O\left(\frac{T^{1-2d}\log^5 m}{m^{1-\max\{d,0\}}l^{1-d-\min\{d,0\}}}\right)$$

$$+ O\left(\frac{T^{2-4d}\log^4 m}{ml^{2-4d}}\right) + O\left(\frac{T^{2-4d}\log^4 m}{l^{4-4d}}\right) + O\left(\frac{T^{3-6d}\log m}{l^{5-6d}}\right) = o(1) \tag{A.30}$$

by A3*. Equations (A.27) and (A.30) imply

$$m^{-1/2}\sum_{j=l}^{m} a_j\varepsilon_j = m^{-1/2}\sum_{j=l}^{m} a_j\varepsilon_j^v + o_p(1). \tag{A.31}$$

Now note that

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

By Lemma A.2, $-m^{1/2}(2S_{YY})^{-1}\sum_{j=l}^{m} a_j \log f_j^* = o(1)$, because $m = o(T^{4/5})$ so that (A.31) and Lemma A.1 imply

$$m^{1/2}(\hat{d}-d) = -m(2S_{YY})^{-1}m^{-1/2}\sum_{j=l}^{m} a_j\varepsilon_j^v + o_p(1). \tag{A.32}$$

Let $U_j^v = \varepsilon_j^v + \log\left[f^*\left(\lambda_j\right)/f^*(0)\right] - 2d\log\left[\left|1-\exp\left(-i\lambda_j\right)\right|/\lambda_j\right]$. Then,

$$m^{-1/2}\sum_{j=l}^{m} a_j\varepsilon_j^v = m^{-1/2}\sum_{j=l}^{m} a_j U_j^v - m^{-1/2}\sum_{j=l}^{m} a_j \log\left\{f^*\left(\lambda_j\right)/f^*(0)\right\}$$

$$+ 2dm^{-1/2}\sum_{j=l}^{m} a_j \log\left\{\left|1-\exp\left(-i\lambda_j\right)\right|/\lambda_j\right\} \equiv N_1 + N_2 + N_3.$$

We also have $N_2 = -m^{-1/2}\sum_{j=l}^{m} a_j \log f^*\left(\lambda_j\right) + m^{-1/2}\log f^*(0)\sum_{j=l}^{m} a_j$, which is $o(1)$ because 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), $\log\left\{\left|1-\exp\left(-i\lambda_j\right)\right|/\lambda_j\right\} = O\left(m^2/T^2\right)$, uniformly. Thus,

$$N_3^2 \le m^{-1}4d^2\sum_{j=1}^{m} a_j^2\sum_{j=1}^{m}\left[\log\left\{\left|1-\exp\left(-i\lambda_j\right)\right|/\lambda_j\right\}\right]^2$$

$$= O(1/m)O(m)O\left(m^5/T^4\right) = o(1)$$

by Lemma 1 of Hurvich and Beltrao (1994) because $m = o\left(T^{4/5}\right)$. Finally, because Lemma A.8 holds, the results of Robinson (1995, pp. 1067–1070) provide that the moments of N_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, then $N_1 \xrightarrow{d} N\left(0,\pi^2/6\right)$. In summary, given (A.32),

$$m^{1/2}(\hat{d}-d) = -m(2S_{YY})^{-1}\left(N_1+N_2+N_3\right) + o_p(1)$$

$$= -m^{-1}2[m+o(m)]N_1 + o_p(1) \xrightarrow{d} N\left(0,\pi^2/24\right). \qquad \blacksquare$$

https://doi.org/10.1017/S0266466613000042 Published online by Cambridge University Press