AN EFFICIENT TAPER FOR POTENTIALLY OVERDIFFERENCED LONG-MEMORY TIME SERIES

By Clifford M. Hurvich and Willa W. Chen

Stern School of Business, New York University

First version received January 1998

Abstract. We propose a new complex-valued taper and derive the properties of a tapered Gaussian semiparametric estimator of the long-memory parameter $d \in (-0.5, 1.5)$. The estimator and its accompanying theory can be applied to generalized unit root testing. In the proposed method, the data are differenced once before the taper is applied. This guarantees that the tapered estimator is invariant with respect to deterministic linear trends in the original series. Any detrimental leakage effects due to the potential noninvertibility of the differenced series are strongly mitigated by the taper. The proposed estimator is shown to be more efficient than existing invariant tapered estimators. Invariance to kth order polynomial trends can be attained by differencing the data k times and then applying a stronger taper, which is given by the kth power of the proposed taper. We show that this new family of tapers enjoys strong efficiency gains over comparable existing tapers. Analysis of both simulated and actual data highlights potential advantages of the tapered estimator of d compared with the nontapered estimator.

Keywords. Periodogram; Gaussian semiparametric estimation; unit roots.

1. INTRODUCTION

Most theoretical results for long-memory time series assume that the memory parameter d lies in the interval $(-0.5, 0.5)$, so that the series is stationary and invertible. In applications, however, the estimated value of d based on the original data may exceed 0.5, indicating nonstationarity. Furthermore the potential presence of linear trend, excluded in much of the existing theory, could severely bias estimates of d. We will suppose that the observed series was generated by a process with $d \in (-0.5, 1.5)$, and may have an additive deterministic linear trend. This seems sufficient for many applications in areas such as finance, econometrics, hydrology, climatology, and network flow analysis.

Differencing is a very widely used technique for detrending and inducing stationarity. The ordinary first difference will convert the memory parameter to $d^* = d - 1$, and will completely remove a linear trend, without forcing the analyst to estimate the trend. Overdifferencing is a term to describe the situation where the differences are noninvertible, i.e. $d^* \leqslant -0.5$, but the differences are nevertheless used for modeling and parameter estimation.

0143-9782/00/02 155–180 JOURNAL OF TIME SERIES ANALYSIS Vol. 21, No. 2

Overdifferencing may arise as an unintended consequence of differencing, and causes many problems in parameter estimation. See Davis and Dunsmuir (1996) for theory on estimation of noninvertible MA(1) models, i.e. overdifferenced white noise. See Hurvich and Ray (1995) for a discussion of periodogram bias induced by the overdifferencing of long-memory time series.

Tapering (see, for example, Tukey, 1967) is a technique for reducing periodogram bias due to strong peaks and troughs in the spectral density. A taper is a nonrandom weight sequence with certain desired properties that is multiplied with the time series data prior to Fourier transformation or parameter estimation. It was suggested by Hurvich and Ray (1995) that the use of a taper can alleviate the detrimental effects of overdifferencing, most importantly the bias in estimates of d^* based on the periodogram of the differenced data. Deo and Hurvich (1998) showed that tapering can also be helpful for estimating the mean of a potentially overdifferenced long-memory time series, i.e. the linear trend in the original series. Assuming no deterministic trends and working with the original series, Velasco (1999a, 1999b) established asymptotic normality of the nontapered semiparametric estimators of d described in Robinson (1995a, 1995b) in the nonstationary case $d \in [0.5, 0.75)$, and explained why the theory breaks down when d exceeds 0.75. This serves to motivate the use of tapered estimators.

The discussion above indicates that the routine use of differencing followed by tapering may be helpful in many situations. The main difficulty with this strategy is that tapering may strongly inflate the variance of estimates of d^* and other parameters of the series. Velasco (1999a, 1999b) has obtained general consistency and asymptotic normality results for periodogram and log-periodogram semiparametric estimates of d, based on either levels or differences, with d in a potentially much wider range than considered here, and in the potential presence of additive polynomial trends of arbitrary degree. This can be achieved by using a class of tapers due to Kolmogorov (see Zhurbenko, 1979). Unfortunately, the efficiency loss incurred from using these tapers may be quite substantial. Even in the best case, the asymptotic variance of a Gaussian semiparametric estimator (GSE; Robinson, 1995b) with Kolmogorov tapering is $2.1/4m$, compared with $1/4m$ without tapering, where m is the number of periodogram ordinates used. This occurs for the Kolmogorov taper of order 2, which produces estimators that are invariant in the presence of linear trends in the levels.

In this paper we will introduce a new taper that can be safely used on differenced data having a constant mean and memory parameter $d^* \in (-1.5, 0.5)$. We will establish that the corresponding Gaussian semiparametric estimator of d^* based on the tapered differences is consistent and asymptotically normal, with an asymptotic variance of $1.5/4m$. This represents a substantial efficiency gain compared with tapered GSE estimators with the Kolmogorov and cosine tapers used in Velasco (1999a, 1999b). It also represents a small efficiency gain compared with the nontapered log-periodogram estimator proposed by Geweke and Porter-Hudak (1983) and justified theoretically for a trend-free Gaussian series with $d \in (-0.5, 0.5)$ by Robinson (1995a) and Hurvich *et al.* (1998).

The estimators of d developed here and in Velasco (1999a, 1999b) provide freedom from the necessity of having prior knowledge on the stationarity ($d < 0.5$) or nonstationarity ($d \geqslant 0.5$) of the nontrending component of the original series, and on the presence or absence of linear trend in the original series. There are several potential applications for such a consistent and asymptotically normal estimator of d. One is for the construction of asymptotically efficient estimators and asymptotically valid confidence intervals for linear trend, as discussed by Deo and Hurvich (1998). Another application is for a nonparametric mean function with additive long-memory errors having $d \in [0, 1.5)$. Here, d^* may be estimated from the tapered differences, and the estimate may then be used to set confidence intervals for the true regression function. The application we will consider in this paper is generalized unit root testing for economic time series. A sufficiently narrow confidence interval for d could be very helpful in this context. For example, a confidence interval of (0.2, 0.3) indicates that the series has long memory ($d > 0$) and is stationary ($d < 0.5$), while a confidence interval of (0.6, 0.7) indicates that the series is nonstationary ($d > 0.5$), but that the degree of nonstationarity is weaker than that of a random walk ($d < 1$), and an interval of (0.9, 1.1) indicates nonstationarity but does not rule out the possibility that $d = 1$. For more discussion of the use of estimates of d in the context of unit root testing, see Cheung (1993), Cheung and Lai (1993), Hassler (1993), Delgado and Robinson (1994), Hassler and Wolters (1995), Baillie *et al.* (1996) and Crato and de Lima (1997).

Throughout the paper, we suppose that we observe data $x_0, \ldots, x_n$ from a process having memory parameter $d \in (-0.5, 1.5)$. Equivalently, the differences $y_1, \ldots, y_n$ are generated from a weakly stationary process $\{y_t\} = \{x_t - x_{t-1}\}$ with memory parameter $d^* = d - 1 \in (-1.5, 0.5)$ and spectral density $f(\lambda) \sim \lambda^{-2d^*}$ as $\lambda \to 0^+$. The mean of y_t is a constant, μ, which need not equal zero, so that $\{x_t\}$ has a linear trend with slope μ

2. THE FOURIER TRANSFORM AND THE COSINE BELL TAPER

The discrete Fourier transform (DFT) of the data $\{y_t\}_{t=1}^n$ is given by

$$w_j = w(\lambda_j) = \frac{1}{(2\pi n)^{1/2}} \sum_{t=1}^{n} y_t \exp(i\lambda_j t) \qquad j = 1, \ldots, \tilde{n}$$

where $\tilde{n} = [(n-1)/2]$ and $\lambda_j = 2\pi j/n$ is the jth Fourier frequency. Note that if $\{y_t\}_{t=1}^n$ were replaced by $\{y_t + C\}_{t=1}^n$ for any constant C, the DFT values w_j would be unchanged, since $\sum_{t=1}^n \exp(i\lambda_j t) = 0$. Thus, the w_j values are free of any dependence on the mean μ. This property, which we will refer to as shift-invariance, is shared by the periodogram $I_j = |w_j|^2$, for $j = 1, \ldots, \tilde{n}$, and by any estimator obtained as a function of these periodogram values. Since μ has no effect on d^*, shift-invariance is a desirable attribute of estimators of d^*. It is also desirable that shift-invariance be attained without the need to directly estimate μ, as is the case for the w_j, in view of the potentially slow convergence rates of estimators of μ (see Samarov and Taqqu, 1988; Cheung and Diebold, 1994; Deo and Hurvich, 1998).

A widely used taper is the cosine bell (Tukey, 1967), given by

$$h_{t,\mathrm{CB}} = 0.5\left[1 - \cos\left\{\frac{2\pi(t-1/2)}{n}\right\}\right] \qquad t = 1, \ldots, n. \tag{1}$$

The tapered DFT and periodogram are defined by

$$w_{j,\mathrm{CB}} = \frac{1}{(2\pi\sum h_{t,\mathrm{CB}}^2)^{1/2}}\sum_{t=1}^{n} h_{t,\mathrm{CB}}\, y_t \exp(i\lambda_j t)$$

and $I_{j,\mathrm{CB}} = |w_{j,\mathrm{CB}}|^2$. It can be shown that $\sum h_{t,\mathrm{CB}}^2 = 3n/8$ and (see Bloomfield, 1976, pp. 80–84) that

$$w_{j,\mathrm{CB}} = (8/3)^{1/2}\{-0.25 w_{j-1}\exp(i\pi/n) + 0.5 w_j - 0.25 w_{j+1}\exp(-i\pi/n)\} \qquad j = 2, \ldots, \tilde{n}. \tag{2}$$

Since $w_{j,\mathrm{CB}}$ is a linear combination of the ordinary DFT values $\{w_j\}$ at nonzero Fourier frequencies, it follows that the $w_{j,\mathrm{CB}}$ are shift-invariant. A compromise between not tapering at all and the cosine bell is achieved by the split cosine bell taper (see Tukey, 1967; Bloomfield, 1976; Deo and Hurvich, 1998). Unfortunately, the resulting tapered Fourier transform is not shift-invariant, and therefore the split cosine bell is not suitable for use in estimating d^*.

3. A NEW TAPER

Analogously to Equation (1), we define a new taper by

$$h_t = 0.5\left[1 - \exp\left\{\frac{i2\pi(t-1/2)}{n}\right\}\right] \qquad t = 1, \ldots, n \tag{3}$$

so that $\sum |h_t|^2 = n/2$. Define the tapered Fourier transform by

$$w_j^{\mathrm{T}} = \frac{1}{(2\pi\sum|h_u|^2)^{1/2}}\sum_{t=1}^{n} h_t y_t \exp(i\lambda_j t). \tag{4}$$

The primary motivation for the Definition (3) of the new taper is provided by the relationship between the tapered and nontapered Fourier transforms, in analogy to Equation (2) for the cosine bell:

$$w_j^{\mathrm{T}} = \sqrt{2}\{0.5 w_j - 0.5 w_{j+1}\exp(-i\pi/n)\} \qquad j = 1, \ldots, \tilde{n}-1. \tag{5}$$

Note that only the two DFT ordinates w_j and w_{j+1} need be used to compute the tapered DFT w_j^{T}. Equation (5) suggests that the tapered periodogram $I_j^{\mathrm{T}} = |w_j^{\mathrm{T}}|^2$ may be viewed as an estimator of $f_{\tilde{j}} = f(\lambda_{\tilde{j}})$, where $\tilde{j} = j + 1/2$ and $\lambda_{\tilde{j}} = 2\pi\tilde{j}/n$. Equation (4) expresses w_j^{T} as a Fourier transform of a tapered data set, but the taper $\{h_t\}$ defined in Equation (3) has the unusual property that it is complex valued. This is something of a hindrance in both theoretical and practical considerations. However, complex $\{h_t\}$ are an inevitable consequence of our primary goal (Equation (5)) that w_j^{T} be a linear combination of two neighbouring DFT values, a choice that guarantees shift-invariance and improves efficiency relative to existing shift-invariant tapers. It is easily shown that $|h_t| = (h_{t,\mathrm{CB}})^{1/2} = \sin\{\pi(t - 1/2)/n\}$, and hence for large n the slope of the $|h_t|$ curve approaches ± 1 at the time end-points $t = 1$ and $t = n$.

4. THE TAPERED GAUSSIAN SEMIPARAMETRIC ESTIMATOR OF d_0^*

The Gaussian semiparametric estimator (GSE) was originally proposed by Kunsch (1987) for estimating the memory parameter $d_0 \in (-0.5, 0.5)$ of the untapered levels $\{x_t\}$, assuming no linear trend ($\mu = 0$). The GSE is implicitly defined, so the subscript in d_0 is needed to distinguish the true value from the variable d used in the objective function. Robinson (1995b) showed that the estimator, $\hat{d}_{\mathrm{GSE}}$, is consistent and asymptotically normal, with asymptotic variance $1/4m$. In the potential presence of polynomial trends and with d_0 not necessarily confined to $(-0.5, 0.5)$, Velasco (1999b) established consistency and asymptotic normality of a broad class of tapered GSE estimators in which the raw periodogram of $\{x_t\}$ is replaced by a tapered periodogram. However, Velasco (1999b) did not explicitly consider the noninvertible case, and in order to achieve invariance to polynomial trends of order $p - 1$ using the Kolmogorov tapers of order $p \geqslant 2$, he needed to omit all Fourier frequencies that are not a multiple of λ_p from the estimator. The asymptotic variance of the estimator exceeds $p/4m$, a strong inflation from the nontapered case. If the cosine bell taper is used and $\mu = 0$, then all Fourier frequencies may be used, but the asymptotic variance, at $35/18m$, is still much larger than in the nontapered case.

We will study the properties of a tapered GSE estimator of $d_0^* \in (-1.5, 0.5)$ based on the differences $\{y_t\}$, allowing for nonzero μ, using the new taper of Section 3. Unfortunately, this taper does not fit into the class considered by Velasco (1999b), since it is complex valued. The estimator is the minimizer of

$$Q(G, d^*) = \frac{1}{m}\sum_{j=1}^{m}\left\{\log g(\lambda_{\tilde{j}}) + \frac{I_j^{\mathrm{T}}}{g(\lambda_{\tilde{j}})}\right\}$$

where $g(\lambda) = G|\lambda|^{-2d^*}$. The objective function Q is to be minimized with respect to $G \in (0, \infty)$ and $d^* \in \Theta = [\Delta_1, \Delta_2]$, where $-1.5 < \Delta_1 < \Delta_2 < 0.5$. Equivalently, concentrating out the parameter G, the estimator is $\hat{d}^*_{\mathrm{GSET}} = \arg\min_{d^* \in \Theta} R(d^*)$, where

$$R(d^*) = \log \hat{G}(d^*) - 2d^* \frac{1}{m}\sum_{j=1}^{m} \log \lambda_{\tilde{j}} \qquad \hat{G}(d^*) = \frac{1}{m}\sum_{j=1}^{m} \lambda_{\tilde{j}}^{2d^*} I_j^{\mathrm{T}}.$$

5. THEORETICAL RESULTS

To derive the properties of $\hat{d}^*_{\mathrm{GSET}}$, we follow the path set out in Robinson (1995b) and Velasco (1999b), hereafter denoted by Rob and Vel, respectively. The results of Vel cannot be applied directly because our taper is complex valued, and because Vel did not give properties of tapered GSE estimators in the noninvertible case. We assume throughout that $y_t - \mu = \sum_{j=0}^{\infty}\alpha'_j\varepsilon_{t-j}$ where $\{\varepsilon_t\}$ is a martingale difference sequence with unit variance. Define

$$w^{\mathrm{T}}_{\varepsilon j} = \frac{1}{(2\pi\sum|h_u|^2)^{1/2}}\sum_{t=1}^{n} h_t\varepsilon_t \exp(i\lambda_j t)$$

and let $I^{\mathrm{T}}_{\varepsilon j} = |w^{\mathrm{T}}_{\varepsilon j}|^2$. The moving-average transfer function is denoted by $\alpha(\lambda)$, with $\alpha_j = \alpha(\lambda_j)\sum_{L=0}^{\infty}\alpha'_L \exp(iL\lambda_j)$. The spectral density of $\{y_t\}$ is $f(\lambda) = |\alpha(\lambda)|^2/2\pi$. For consistency and asymptotic normality of $\hat{d}^*_{\mathrm{GSET}}$, we require Assumptions A1–A4 and A1′–A4′ respectively, as given below. Except for A1, A1′ and A2, these assumptions are identical to those originally given and described in Rob.

A1. $f(\lambda) = G_0\lambda^{-2d_0^*}\{1 + E_\beta\lambda^\beta + \mathrm{o}(\lambda^\beta)\}$ as $\lambda \to 0^+$, for some $\beta \in (1, 2]$, $G_0 > 0$, $E_\beta < \infty$ and $d_0^* \in [\Delta_1, \Delta_2]$.

A2. In a neighborhood $(0, \delta)$ of the origin, $\alpha(\lambda)$ is differentiable and $\partial\alpha(\lambda)/\partial\lambda = \mathrm{O}(|\alpha(\lambda)|/\lambda)$ as $\lambda \to 0^+$.

A3. $\sum_{j=0}^{\infty}\alpha'^2_j < \infty$, $E(\varepsilon_t|F_{t-1}) = 0$, $E(\varepsilon_t^2|F_{t-1}) = 1$ almost surely (a.s.), $t = 0, \pm 1, \ldots$, in which F_t is the σ-field of events generated by ε_s, $s \leqslant t$, and there exists a random variable ε such that $E(\varepsilon^2) < \infty$ and, for all $\eta > 0$ and some $K > 0$, $P(|\varepsilon_t| > \eta) \leqslant KP(|\varepsilon| > \eta)$.

A4. As $n \to \infty$, $1/m + m/n \to 0$.

A1′. Assumption A1 holds.

A2′. Assumption A2 holds.

A3′. Assumption A3 holds and also $E(\varepsilon_t^3|F_{t-1}) = \mu_3$ a.s., $E(\varepsilon_t^4) = \mu_4$, $t = 0, \pm 1, \ldots$, for finite constants μ_3 and μ_4.

A4′. As $n \to \infty$, $1/m + m^{1+2\beta}(\log m)^2/n^{2\beta} \to 0$.

Our A1 is equivalent to Assumption 2 discussed in Velasco (1999a, pp. 328–9), and is stronger than A1 in Rob, which does not explicitly include the parameter β. Furthermore, we assume in A1 and A1′ that $\beta \in (1, 2]$, whereas A1′ of Rob assumes only that $\beta \in (0, 2]$. Our A2 is the same as A2′ of Rob, and implies that

$$\frac{d}{d\lambda} \log f(\lambda) = \mathrm{O}(\lambda^{-1})$$

as $\lambda \to 0^+$. The more restrictive assumptions made here, also used by Velasco (1999a, 1999b), allow for improved bounds on moments of tapered DFTs (see Lemma 4), and thereby greatly simplify part of the proof of asymptotic normality of $\hat{d}^*_{\mathrm{GSET}}$. Our assumptions are satisfied by the ARFIMA processes, which have $\beta = 2$.

Our first theorem establishes the consistency of $\hat{d}^*_{\mathrm{GSET}}$. Lemmas referred to in the proofs of the theorems are stated and proved in the Appendix.

THEOREM 1. *Under Assumptions A1–A4,* $\hat{d}^*_{\mathrm{GSET}} \xrightarrow{\mathrm{p}} d^*_0$ *as* $n \to \infty$.

PROOF. We follow the arguments and notation of the proof in Rob for the nontapered case. Define $\Theta_1 = \{d^*\colon \Delta \leqslant d^* \leqslant \Delta_2\}$, where $\Delta = \Delta_1$ when $d^*_0 < 1/2 + \Delta_1$ and $d^*_0 \geqslant \Delta > d^*_0 - 1/2$ otherwise. Note that $d^* - d^*_0 > -1/2$ for all $d^* \in \Theta_1$. When $d^*_0 \geqslant 1/2 + \Delta_1$, define $\Theta_2 = \{d^*\colon \Delta_1 \leqslant d^* < \Delta\}$, and otherwise take Θ_2 to be empty. Let $S(d^*) = R(d^*) - R(d^*_0)$. Then we can write $S(d^*) = U(d^*) - T(d^*)$, where

$$T(d^*) = \log\left\{\frac{\hat{G}(d^*_0)}{G_0}\right\} - \log\left\{\frac{\hat{G}(d^*)}{G(d^*)}\right\}$$
$$- \log\left[m^{-1}\{2(d^* - d^*_0) + 1\}\sum_{j=1}^{m}\left(\frac{\tilde{j}}{m}\right)^{2(d^* - d^*_0)}\right]$$
$$+ 2(d^* - d^*_0)\left\{m^{-1}\sum_{j=1}^{m}\log\tilde{j} - (\log m - 1)\right\} \qquad (6)$$

$$U(d^*) = 2(d^* - d^*_0) - \log\{2(d^* - d^*_0) + 1\}$$

$$G(d^*) = G_0 m^{-1}\sum_{j=1}^{m}\lambda_{\tilde{j}}^{2(d^* - d^*_0)}.$$

Combining Equations (3.2) and (3.3) of Rob, we have for $1/4 > \delta > 0$ that

$$P(|\hat{d}^*_{\mathrm{GSET}} - d^*_0| \geqslant \delta) \leqslant P\left\{\sup_{\Theta_1}|T(d^*)| \geqslant \inf_{\overline{N}_\delta \cap \Theta_1} U(d^*)\right\} + P\left\{\inf_{\Theta_2} S(d^*) \leqslant 0\right\} \qquad (7)$$

where $N_\delta = \{d^*: |d^* - d_0^*| < \delta\}$ and $\overline{N}_\delta = (-\infty, \infty) - N_\delta$. As in (3.4) of Rob, we have $\inf_{\overline{N}_\delta \cap \Theta_1} U(d^*) > \delta^2/2$. Thus, Lemma 1 implies that the first term on the right-hand side of Equation (7) tends to zero. The second term also tends to zero by Lemma 2, so the theorem is proved.

Our next theorem establishes the asymptotic distribution of $\hat{d}^*_{\mathrm{GSET}}$.

THEOREM 2. *Under Assumptions A1′–A4′,* $m^{1/2}(\hat{d}^*_{\mathrm{GSET}} - d_0^*) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \Phi/4)$, *where* $\Phi = 1.5$.

PROOF. By Theorem 1, with probability approaching 1 as $n \to \infty$, $\hat{d}^*_{\mathrm{GSET}}$ satisfies

$$0 = R'(\hat{d}^*_{\mathrm{GSET}}) = R'(d_0^*) + R''(\tilde{d}^*)(\hat{d}^*_{\mathrm{GSET}} - d_0^*)$$

where $|\tilde{d}^* - d_0^*| \leqslant |\hat{d}^*_{\mathrm{GSET}} - d_0^*|$. Using Lemmas 6 and 7 we can construct arguments along the lines given in Rob, pp. 1641–1644. In view of (4.6) and (4.7) of Rob, we need to verify that, for $0 < \delta < 1/4$,

$$6\sum_{r=1}^{m}\left(\frac{\tilde{r}}{m}\right)^{1-2\delta}\frac{1}{\tilde{r}^2}\left|\sum_{j=1}^{r}\left(\frac{I_j^{\mathrm{T}}}{g_{\tilde{j}}} - 1\right)\right| + \frac{3}{m}\left|\sum_{j=1}^{m}\left(\frac{I_j^{\mathrm{T}}}{g_{\tilde{j}}} - 1\right)\right| = \mathrm{o_p}\{(\log m)^{-6}\} \qquad (8)$$

where $g_{\tilde{j}} = G_0\lambda_{\tilde{j}}^{-2d_0^*}$ and $\tilde{r} = r + 1/2$. Using Lemmas 6 and 7, we obtain $\sum_{j=1}^{r}(I_j^{\mathrm{T}}/g_{\tilde{j}} - 1) = \mathrm{O_P}(r^{1/2} + r^{\beta+1}/n^{\beta})$, so that the left-hand side of (8) is $\mathrm{O_p}\{m^{-1/2} + (m/n)^{\beta}\} = \mathrm{O_p}(m^{-1/2})$ by A4′, and (8) is established. Following the lines of the proof in Rob, we conclude that

$$R''(\tilde{d}^*) \xrightarrow{\mathrm{p}} 4 \qquad (9)$$

and that

$$m^{1/2}R'(d_0^*) = 2m^{-1/2}\sum_{j=1}^{m} v_j\left(\frac{I_j^{\mathrm{T}}}{g_{\tilde{j}}} - 1\right)\{1 + \mathrm{o_p}(1)\} \qquad (10)$$

where $v_j = \log\tilde{j} - m^{-1}\sum_{j=1}^{m}\log\tilde{j}$. Using summation by parts, Lemmas 6 and 7, and Condition A4′, (10) is

$$\left\{2m^{-1/2}\sum_{j=1}^{m} v_j(2\pi I_{\varepsilon j}^{\mathrm{T}} - 1) + \mathrm{O_p}\left(m^{1/2-\beta/2}\log m + \frac{m^{\beta+1/2}}{n^{\beta}}\log m + m^{-1/2}\log^2 m\right)\right\}\{1 + \mathrm{o_p}(1)\}$$

$$= \left\{2m^{-1/2}\sum_{j=1}^{m} v_j(2\pi I_{\varepsilon j}^{\mathrm{T}} - 1) + \mathrm{o_p}(1)\right\}\{1 + \mathrm{o_p}(1)\}.$$

The theorem now follows from Lemma 8 and Equation (9).

6. SIMULATION RESULTS

Using the method of Davies and Harte (1987), we generated 500 realizations of a variety of zero-mean stationary and nonstationary ARFIMA(1, d, 0) models, with $n = 500$. The models are expressed as $(1-\rho B)(1-B)^d x_t = \varepsilon_t$, where the ε_t are independent standard normal and B is the backshift operator. We considered all combinations of $d = 0$, 0.2, 0.4, 0.6, 0.8, 1, 1.2 and $\rho = 0$, 0.5, 0.8.

For each realization of each process, we computed the nontapered and tapered Gaussian semiparametric estimators $\hat{d}^*_{\text{GSE}}$ and $\hat{d}^*_{\text{GSET}}$ of $d^* = d - 1$ from the differences $\{y_t\}_{t=1}^{500}$. For the empirical studies of this and the next section, we will use slightly different definitions of the estimators than given in Section 4, replacing $g(\lambda)$ by $\overline{g}(\lambda) = G|1-\exp(-i\lambda)|^{-2d^*} = G\{2\sin(\lambda/2)\}^{-2d^*}$. This modification is done for the sake of compatibility with ARFIMA models, and does not alter the asymptotic properties stated in Section 5. Thus, the estimators $\hat{d}^*_{\text{GSE}}$ and $\hat{d}^*_{\text{GSET}}$ used here are the minimizers of, respectively,

$$Q_1(G, d^*) = \frac{1}{m}\sum_{j=1}^{m}\left\{\log\overline{g}(\lambda_j) + \frac{I_j}{\overline{g}(\lambda_j)}\right\}$$

$$Q_2(G, d^*) = \frac{1}{m}\sum_{j=1}^{m}\left\{\log\overline{g}(\lambda_{\tilde{j}}) + \frac{I_j^{\mathrm{T}}}{\overline{g}(\lambda_{\tilde{j}})}\right\}$$

where the objective functions Q_1 and Q_2 are to be minimized with respect to $G \in (0, \infty)$ and $d^* \in \Theta = [-1.49, 0.49]$. The value of m used in the simulations was $m = [0.25 n^{4/5}]$, so that $m = 36$ for $n = 500$. The choice $m = \text{constant} \times n^{4/5}$ violates Condition A4′, but the choice may be optimal in terms of mean squared error, in view of the results of Henry and Robinson (1996). Nevertheless, to the best of our knowledge, no asymptotic theory has yet been derived for $\hat{d}^*_{\text{GSE}}$ in the case $d^* \leqslant -0.5$.

Before presenting the simulation results, we provide approximations for $\text{var}(\hat{d}^*_{\text{GSET}})$ and $\text{var}(\hat{d}^*_{\text{GSE}})$, which are more accurate than, but converge to, the asymptotic values. Define

$$\tilde{v}_j = \log\left(2\sin\frac{\lambda_{\tilde{j}}}{2}\right) - m^{-1}\sum_{j=1}^{m}\log\left(2\sin\frac{\lambda_{\tilde{j}}}{2}\right).$$

From the first equation in the proof of Theorem 2, we have $\sqrt{m}(\hat{d}^*_{\text{GSET}} - d_0^*) \approx -\sqrt{m}R'(d_0^*)/R''(\tilde{d}^*)$. From an analog to (4.10) of Rob, we have $R''(\tilde{d}^*) \approx 4m^{-1}\sum_{j=1}^{m}\tilde{v}_j^2$. From Equation (10) and the surrounding discussion, we have

$$\sqrt{m}R'(d_0^*) \approx \frac{2}{\sqrt{m}}\sum_{j=1}^{m}\tilde{v}_j(2\pi I_{\varepsilon j}^{\mathrm{T}} - 1)$$

the variance of which, from the proof of Lemma 8, may be approximated by $4(1.5)m^{-1}\sum\tilde{v}_j^2$. This leads to the approximation

$$\text{var}(\hat{d}^*_{\text{GSET}}) \approx 1.5 \bigg/ \left(4\sum_{j=1}^{m}\tilde{v}_j^2\right). \tag{11}$$

The analog to Equation (11) for the nontapered estimator is $\text{var}(\hat{d}^*_{\text{GSE}}) \approx 1/\{4\sum_{j=1}^{m}(X_j - \overline{X})^2\}$, where $X_j = \log\{2\sin(\lambda_j/2)\}$ and $\overline{X} = m^{-1}\sum_{j=1}^{m}\log\{2\sin(\lambda_j/2)\}$.

Table I presents the mean and variance of $\hat{d}_{\text{GSE}} = \hat{d}^*_{\text{GSE}} + 1$ and $\hat{d}_{\text{GSET}} = \hat{d}^*_{\text{GSET}} + 1$, based on the 500 realizations of the ARFIMA processes. The empirical variances in the table may be compared with the following theoretical values. For $\text{var}(\hat{d}_{\text{GSE}})$, we have $1/4m = 0.00694$, and the variance approximation given above is 0.00959. For $\text{var}(\hat{d}_{\text{GSET}})$, we have $1.5/4m = 0.01042$, and the variance approximation from Equation (11) is 0.01685. It should be noted that the ratio of the variance approximations for $\hat{d}_{\text{GSET}}$ and $\hat{d}_{\text{GSE}}$ is 1.757, which markedly exceeds the asymptotic value of 1.5.

When d is small, $\hat{d}_{\text{GSET}}$ is often substantially less biased than $\hat{d}_{\text{GSE}}$. The most extreme example of this occurs when $d = 0$ and $\rho = 0$, so that $\{y_t\}$ is a

TABLE I

PROPERTIES OF NONTAPERED AND TAPERED GAUSSIAN SEMIPARAMETRIC ESTIMATORS OF d FOR ARFIMA (1, d, 0) MODELS, $n = 500$

d	ρ	mean($\hat{d}_{\text{GSE}}$)	var($\hat{d}_{\text{GSE}}$)	Mean($\hat{d}_{\text{GSET}}$)	var($\hat{d}_{\text{GSET}}$)
0	0	0.2742	0.0403	−0.0013	0.0186
0	0.5	0.2116	0.0255	0.0574	0.0188
0	0.8	0.3534	0.0154	0.3116	0.0198
0.2	0	0.3192	0.0215	0.1994	0.0173
0.2	0.5	0.3098	0.0133	0.2580	0.0175
0.2	0.8	0.5008	0.0108	0.5112	0.0190
0.4	0	0.4389	0.0130	0.3964	0.0174
0.4	0.5	0.4665	0.0107	0.4551	0.0176
0.4	0.8	0.6878	0.0106	0.7091	0.0190
0.6	0	0.6048	0.0114	0.5949	0.0173
0.6	0.5	0.6548	0.0111	0.6533	0.0176
0.6	0.8	0.8833	0.0113	0.9079	0.0191
0.8	0	0.7929	0.0107	0.7945	0.0173
0.8	0.5	0.8453	0.0106	0.8535	0.0174
0.8	0.8	1.0758	0.0110	1.1079	0.0190
1	0	0.9942	0.0101	0.9895	0.0187
1	0.5	1.0459	0.0104	1.0488	0.0187
1	0.8	1.2764	0.0113	1.2999	0.0171
1.2	0	1.1923	0.0101	1.1981	0.0169
1.2	0.5	1.2444	0.0102	1.2553	0.0164
1.2	0.8	1.4414	0.0048	1.4453	0.0056

Based on 500 realizations of the ARFIMA model, AR(1) parameter $= \rho$. Nontapered and tapered estimators $\hat{d}_{\text{GSE}}$ and $\hat{d}_{\text{GSET}}$. Estimators are constructed from differenced data, with $m = 36$. Approximate and asymptotic variances for $\hat{d}_{\text{GSE}}$: 0.00959 and 0.00694. Approximate and asymptotic variances for $\hat{d}_{\text{GSET}}$: 0.01685 and 0.01042.

noninvertible MA(1) process. For a given value of ρ, the bias of $\hat{d}_{\text{GSE}}$ decreases as d increases, while the bias of $\hat{d}_{\text{GSET}}$ remains nearly constant. The relatively poor bias properties of $\hat{d}_{\text{GSE}}$ may be attributed to leakage suffered by the nontapered periodogram. For a given value of d, the bias of both estimators typically increases as ρ increases, due to contamination of the periodogram from the short-memory component of the spectral density.

The variance of $\hat{d}_{\text{GSE}}$ is typically substantially less than that of $\hat{d}_{\text{GSET}}$, as would be expected from asymptotic theory, and from the variance approximations. Surprisingly, however, $\text{var}(\hat{d}_{\text{GSET}}) < \text{var}(\hat{d}_{\text{GSE}})$ for $(d, \rho) = (0, 0)$, $(0, 0.5)$ and $(0.2, 0)$. In the most extreme case where $d = 0$, $\rho = 0$, the ratio is $\text{var}(\hat{d}_{\text{GSE}})/\text{var}(\hat{d}_{\text{GSET}}) = 2.17$. The variance inflation in $\hat{d}_{\text{GSE}}$ here may be attributed to correlation in the nontapered periodogram induced by leakage. Except for the cases described above as well as $(d, \rho) = (1.2, 0.8)$, $\text{var}(\hat{d}_{\text{GSE}})$ remains reasonably constant as d and ρ are changed. Meanwhile, $\text{var}(\hat{d}_{\text{GSET}})$ remains reasonably constant for all cases except $(d, \rho) = (1.2, 0.8)$, where the variance of both estimators is inexplicably small.

Next, we discuss the quality of the theoretical variance expressions, ignoring the unusual cases described above. The asymptotic variances $1/4m$ and $1.5/4m$, for $\text{var}(\hat{d}_{\text{GSE}})$ and $\text{var}(\hat{d}_{\text{GSET}})$ respectively, both strongly understate the actual observed variances, while the variance approximations (Equation (11) and the formula following it) are reasonably accurate. For example, when $d = 1$ and $\rho = 0$, so that $\{y_t\}$ is white noise, the ratio of the observed to asymptotic variances is 1.46 for $\hat{d}_{\text{GSE}}$ and 1.79 for $\hat{d}_{\text{GSET}}$, while the ratio of the observed to approximate variances is 1.05 and 1.11 for the two estimators, respectively.

Next, we consider several one-tailed hypothesis tests regarding d that may be useful, particularly in an econometric context. The tests, together with their null hypotheses H_0 and alternative hypotheses H_1, are as follows.

T_{LM}, H_0: $d = 0$, H_1: $d > 0$.

T_{DS}, H_0: $d < 0.5$, H_1: $d \geqslant 0.5$.

T_{TS}, H_0: $d \geqslant 0.5$, H_1: $d < 0.5$.

T_{MR}, H_0: $d \geqslant 1$, H_1: $d < 1$.

The subscripts used in the names of the tests describe the alternative hypothesis. Thus T_{LM} is a test for long memory ($d > 0$) with a null hypothesis of short memory, T_{DS} is a test for difference stationarity with a null hypothesis of trend stationarity, T_{TS} is a test for trend stationarity with a null hypothesis of difference stationarity, and T_{MR} is a test for mean reversion with a null hypothesis of no mean reversion. For an explanation as to why a series with $d < 1$ is referred to as mean-reverting, see, for example, Cheung and Lai (1993).

We performed the hypothesis tests described above based on $\hat{d}_{\text{GSET}}$, assumed to be normally distributed with a mean of d and a variance given by Equation (11). All of the tests are one-tailed, at a nominal significance level of 0.05. The results on test performance given here cannot be directly compared with those of Hassler (1993) and Crato and de Lima (1997), due to the variation in choices on differencing, tapering, type of estimator, and number of frequencies used.

Table II gives, for each test and model, the proportion of rejections of the null hypothesis out of the 500 replications. For $\rho = 0$ and $\rho = 0.5$, all tests are reasonably powerful and hold their sizes reasonably well, although the power is unsurprisingly low when d is near to its value under the null hypothesis. In all cases, the power increases monotonically as d moves away from its null value. When $\rho = 0.8$, the substantial bias in $\hat{d}_{\mathrm{GSET}}$ induced by the autoregressive parameters causes size and power distortions.

For testing a null hypothesis of $d = 1$, a score test based on first differences as described in Lobato and Robinson (1998) should be equivalent to the version of our (Wald) test based on $\hat{d}_{\mathrm{GSE}}$. Although the score test could be carried out without the computational expense of directly estimating d, the score test procedure does not yield confidence intervals for d.

TABLE II

Proportion of Rejections of Null Hypotheses in Tests for d in 500 Replications of ARFIMA (1, d, 0) Series, $n = 500$

d	ρ	Long memory	Difference stationarity	Trend stationarity	Mean reversion
0	0	0.044	0.000	0.982	1.000
0.2	0	0.468	0.000	0.736	1.000
0.4	0	0.916	0.004	0.192	1.000
0.6	0	0.998	0.186	0.016	0.928
0.8	0	1.000	0.734	0.000	0.470
1	0	1.000	0.968	0.000	0.064
1.2	0	1.000	1.000	0.000	0.002
0	0.5	0.130	0.000	0.964	1.000
0.2	0.5	0.640	0.000	0.578	1.000
0.4	0.5	0.956	0.022	0.090	0.996
0.6	0.5	0.998	0.334	0.002	0.842
0.8	0.5	1.000	0.858	0.000	0.288
1	0.5	1.000	0.986	0.000	0.040
1.2	0.5	1.000	1.000	0.000	0.000
0	0.8	0.746	0.000	0.412	1.000
0.2	0.8	0.976	0.058	0.054	0.982
0.4	0.8	1.000	0.478	0.002	0.698
0.6	0.8	1.000	0.920	0.000	0.174
0.8	0.8	1.000	0.998	0.000	0.014
1	0.8	1.000	1.000	0.000	0.000
1.2	0.8	1.000	1.000	0.000	0.000

AR(1) parameter $= \rho$. Tests are one-tailed, at level of significance 0.05. Tests based on tapered estimators $\hat{d}_{\mathrm{GSET}}$ with $m = 36$, and variance approximation of Equation (11). Alternative hypotheses are long memory ($d > 0$), difference stationarity ($d \geqslant 0.5$), trend stationarity ($d < 0.5$) and mean reversion ($d < 1$).

7. APPLICATIONS

Here, we report results on $\hat{d}_{\mathrm{GSE}}$ and $\hat{d}_{\mathrm{GSET}}$ for several scientific and economic data sets. Two of the data sets were previously analyzed in Deo and Hurvich (1998): seasonally adjusted monthly temperatures (in degrees Celsius) for the northern hemisphere for the years 1854–1989 as given by Beran (1994, pp. 257–61), and the natural logarithms of the daily levels of the S&P 500 composite stock index from 2 July 1962 to 29 December 1995, not adjusted for dividends, taken from the Center for Research in Security Prices Database. The remaining data sets are monthly economic series from January 1957 to December 1997, taken from the International Monetary Fund's International Financial Statistics CD-ROM. These series are the consumer price index (CPI)-based inflation rates for the USA, the UK and France, as well as the US log real wages (manufacturing) and log industrial production. All of these series are seasonally adjusted. The CPI-based inflation rate is defined as the first difference of the log of the CPI. Real wages are defined as the ratio of hourly earnings to CPI. Time series plots of the data (not shown here) indicated that the global temperatures, S&P 500 and industrial production series had roughly linear trends.

The estimates were calculated as described in Section 6. For each data set, the value of m was selected based on an examination of a log–log plot of the tapered periodogram of the differences. Frequencies corresponding to periods of 1 year or less were avoided when this seemed to be warranted, i.e. for the global temperatures and real wages. For the S&P 500 data, m was taken to be $n^{4/5}$, and no discernible seasonal peaks were found in the periodogram. Table III presents the values of n, m, the estimates of d, and approximate standard errors from Equation (11) and the formula that follows it.

It is seen from Table III that all series studied exhibited statistically significant long-memory effects. For the inflation series, there were substantial differences between the nontapered and tapered estimates. For the S&P 500 index, the estimated value of d did not differ significantly from unity,

TABLE III

NONTAPERED AND TAPERED GAUSSIAN SEMIPARAMETRIC ESTIMATES OF d FOR SEVERAL DATA SETS

Data description	n	m	$\hat{d}_{\mathrm{GSE}}$	$\hat{d}_{\mathrm{GSET}}$
Global temperatures	1631	130	0.54 (0.047)	0.45 (0.060)
S&P 500 Stock Index	8431	1383	0.99 (0.014)	0.99 (0.018)
Inflation, USA	490	40	0.70 (0.093)	0.57 (0.123)
Inflation, UK	490	40	0.44 (0.093)	0.33 (0.123)
Inflation, France	490	40	0.43 (0.093)	0.67 (0.123)
Real wages, USA	491	35	1.30 (0.092)	1.43 (0.121)
Industrial production, USA	491	100	1.27 (0.058)	1.34 (0.075)

$\hat{d}_{\mathrm{GSE}}$ and $\hat{d}_{\mathrm{GSET}}$ are the nontapered and tapered estimates, respectively. Approximate standard errors are given in parentheses.

indicating long-term market efficiency. The inflation series and the global temperature series had estimated values of d that were significantly greater than zero and less than one, indicating mean-reverting long memory, but the values were typically not significantly different from 0.5, so that definitive conclusions on the stationarity of these series cannot be made. Industrial production and real wages both had estimated values of d that were significantly greater than unity, indicating a fractional unit root.

8. INVARIANCE TO POLYNOMIAL TRENDS

There are situations where it is desirable that an estimator of d be invariant to quadratic or higher-order polynomial trends in the levels $\{x_t\}$. For example, Velasco (1999a) found evidence of a quadratic trend in logs of nominal US production worker wages in manufacturing. For each positive integer p, there is a Kolmogorov taper which is of *order p* in the sense defined in Velasco (1999a, p. 339). Any taper of order p, if applied to $\{x_t\}$, will yield a tapered periodogram that is invariant to polynominal trends of order $p-1$, provided that the tapered periodogram is evaluated on the grid $\lambda_p, \lambda_{2p}, \ldots$. Thus, for nominal wages, Velasco (1999a) found it appropriate to use a Kolmogorov taper of order 3 or higher. As a price paid for protection against higher-order trends using Kolmogorov tapers of order p, the corresponding tapered GSE estimator has an asymptotic variance that increases with p and exceeds $p/4m$.

Here, we propose to use the taper $\{h_t^{p-1}\}$, i.e. the $(p-1)$th power of $\{h_t\}$ given by Equation (3), to attain invariance to polynominal trends of order $p-1$, and compare these tapers with the corresponding Kolmogorov tapers. Generalizing the case $p=2$ considered previously, we apply the taper $\{h_t^{p-1}\}$ to the $(p-1)$th difference $\{y_t\}_{t=1}^n$ of the original series $\{x_t\}$. This differencing will remove any $(p-1)$th-order polynominal trends in $\{x_t\}$, but the resulting $\{y_t\}$ may be strongly noninvertible, a problem that is remedied by tapering. We will assume throughout this section that $\{y_t\}$ is weakly stationary with mean μ and spectral density $f(\lambda) \sim G_0\lambda^{-2d_0^*}$ as $\lambda \to 0^+$, where $d_0^* \in (-p+1/2, 1/2)$.

We note that $\{h_t^{p-1}\}$ with $p=1$ corresponds to no tapering, and that the modulus of $\{h_t^{p-1}\}$ with $p=3$ yields the cosine bell taper. This helps to clarify the role of the coefficients 0.25, 0.5 and 0.25 in Equation (2) for the cosine bell: they are proportional to the binomial coefficients

$$\binom{2}{0}, \binom{2}{1} \text{ and } \binom{2}{2}.$$

Although $\{h_t^{p-1}\}$ is not of order p as defined in Velasco (1999a), it does have a Fourier transform that decays rapidly. Specifically, define $D_p^{\mathrm{T}}(\lambda) = \sum_{t=1}^n h_t^{p-1}\exp(i\lambda t)$. It follows from Equation (3), the binomial theorem, and some elementary calculations that

$$D_p^{\mathrm{T}}(\lambda) = \left(\frac{1}{2}\right)^{p-1} \exp\left\{\frac{i(n+1)\lambda}{2}\right\} \sum_{k=0}^{p-1} \binom{p-1}{k} (-1)^k \frac{\sin n\lambda/2}{\sin(\lambda+\lambda_k)/2}. \qquad (12)$$

It follows that $D_p^{\mathrm{T}}(\lambda_j) = 0$ for $j = 1, \ldots, n-p$, so that the tapered Fourier transform of $\{y_t\}$ is invariant to shifts in $\{y_t\}$, and therefore to $(p-1)$th-order polynomial trends in $\{x_t\}$. This invariance is achieved without the need to restrict attention to a coarse grid of Fourier frequencies, as is necessary for the Kolmogorov tapers applied to $\{x_t\}$. To describe the decay properties of $D_p^{\mathrm{T}}(\lambda)$, we have the following lemma.

LEMMA 0. *There exists a finite constant K depending on p but not on n such that*

$$|D_p^{\mathrm{T}}(\lambda)| \leqslant K \min(n, n^{1-p}|\lambda|^{-p}) \qquad \lambda \in [-\pi, \pi].$$

Let $\hat{d}^*_{\mathrm{GSET},p}$ denote the tapered GSE estimator of d_0^* based on $\{y_t\}$ as defined in Section 4, but using the taper $\{h_t^{p-1}\}$ and replacing $\lambda_{\tilde{j}}$ by $2\pi\tilde{j}_p/n$, where $\tilde{j}_p = j + (p-1)/2$. Using Lemma 0 together with ideas from the proofs of Theorems 1 and 2, it can be shown that $\hat{d}^*_{\mathrm{GSET},p}$ is asymptotically normal with mean d_0^* and variance $\Phi_p/4m$, where

$$\Phi_p = \lim_{n\to\infty} \frac{n\sum|h_t^{p-1}|^4}{(\sum|h_t^{p-1}|^2)^2} = \frac{\pi\Gamma^2(2p-1)\Gamma^2\{(4p-3)/2\}}{\Gamma^4\{(2p-1)/2\}\Gamma(4p-3)}. \qquad (13)$$

For example, we have $\Phi_1 = 1$, $\Phi_2 = 3/2$, $\Phi_3 = 35/18$, $\Phi_4 = 2.31$. The first expression in (13) for Φ_p is a generalization for complex-valued tapers of the expression given on p. 101 of Velasco (1999b). We can use the right-hand expression in (13) together with Stirling's formula to show that, as $p \to \infty$, $\Phi_p \sim (p\pi/2)^{1/2}$. Thus, the asymptotic efficiency of $\hat{d}^*_{\mathrm{GSET},p}$ relative to the corresponding tapered GSE estimator based on the Kolmogorov taper of order p, which exceeds p/Φ_p, can be made arbitrarily large by taking p sufficiently large.

The proof of Lemma 0 (given in the Appendix) reveals a curious duality in the roles played by differencing for the taper $\{h_t^{p-1}\}$. From Equation (12), it is seen that, apart from a multiplicative factor of modulus bounded by unity, $D_p^{\mathrm{T}}(\lambda)$ is the $(p-1)$th difference of the function $1/\sin(\lambda/2)$ for λ at a spacing of one Fourier frequency. The fact that a $(p-1)$th-order difference renders constant any $(p-1)$th-order polynomial and annihilates any lower-order polynomial is a crucial element in our proof that a Taylor expansion of Equation (12) is of order $\mathrm{O}(n^{1-p}|\lambda|^{-p})$. Curiously, the theme of differencing in the proof of Lemma 0 seems to mirror the context in which the taper $\{h_t^{p-1}\}$ is to be used: for the $(p-1)$th difference of a time series that may have originally possessed a polynomial trend of order up to $p-1$.

APPENDIX

Lemmas 1–5 below require A1–A4. Lemmas 6–8 require A1′–A4′. Lemma 6 is similar to Equation (A23) of Vel, p. 116, and may be compared with (4.8) of Rob. Lemma 7 is similar to (4.9) of Rob.

LEMMA 1. $\sup_{\Theta_1}|T(d^*)| \xrightarrow{\text{p}} 0$, *where* $T(d^*)$ *is defined in Equation (6).*

LEMMA 2. $P\{\inf_{\Theta_2} S(d^*) \leqslant 0\} \to 0$, *where* $S(d^*)$ *is defined in the proof of Theorem 1.*

LEMMA 3. *There exists a finite constant* C *such that, for all sufficiently large* n, $E|I_j^{\text{T}}/g_{\tilde{j}}| \leqslant C$, $j = 1, \ldots, m$, *where* $g_{\tilde{j}} = G_0\lambda_{\tilde{j}}^{-2d_0^*}$.

LEMMA 4. $E(I_j^{\text{T}} - f_{\tilde{j}})/f_{\tilde{j}} = \text{O}(j^{-\beta})$ *and* $E\{w_j^{\text{T}}\overline{w}_{\varepsilon j}^{\text{T}} - \alpha_{\tilde{j}}/(2\pi)\}/\sqrt{f_{\tilde{j}}} = \text{O}(j^{-\beta})$, *uniformly for* $j = 1, 2, \ldots, m$.

LEMMA 5

$$\sum_{r=1}^{m-1}\left(\frac{\tilde{r}}{m}\right)^{2(\Delta-d_0^*)+1}\frac{1}{\tilde{r}^2}\left|\sum_{j=1}^{r}(2\pi I_{\varepsilon j}^{\text{T}} - 1)\right| \xrightarrow{\text{p}} 0 \text{ where } \tilde{r} = r + 1/2.$$

LEMMA 6. *For* $1 \leqslant r \leqslant m$, $\sum_{j=1}^{r}(I_j^{\text{T}}/g_{\tilde{j}} - 2\pi I_{\varepsilon j}^{\text{T}}) = \text{O}_{\text{p}}(r^{1-\beta/2} + \log r + r^{\beta+1}/n^{\beta})$ *as* $n \to \infty$.

LEMMA 7. *For* $1 \leqslant r \leqslant m$, $\sum_{j=1}^{r}(2\pi I_{\varepsilon j}^{\text{T}} - 1) = \text{O}_{\text{p}}(r^{1/2})$ *as* $n \to \infty$.

LEMMA 8

$$\frac{1}{\sqrt{m}}\sum_{j=1}^{m} \upsilon_j(2\pi I_{\varepsilon j}^{\text{T}} - 1) \xrightarrow{\text{d}} \text{N}(0, \Phi)$$

where $\Phi = 1.5$.

PROOF OF LEMMA 1. For any nonnegative random variable Y, if $\varepsilon \leqslant 1$, we have

$$P(|\log Y| \geqslant \varepsilon) \leqslant 2P(|Y - 1| \geqslant \varepsilon/2). \tag{A1}$$

Thus, $\sup_{\Theta_1}|T(d^*)| \xrightarrow{\text{p}} 0$ if

$$\sup_{\Theta_1}\left|\frac{\hat{G}(d^*) - G(d^*)}{G(d^*)}\right| \xrightarrow{\text{p}} 0 \tag{A2}$$

$$\sup_{\Theta_1}\left|\frac{2(d^* - d_0^*) + 1}{m}\sum_{j=1}^{m}\left(\frac{\tilde{j}}{m}\right)^{2(d^*-d_0^*)} - 1\right| \to 0 \tag{A3}$$

and

$$\left|m^{-1}\sum_{j=1}^{m}\log\tilde{j} - (\log m - 1)\right| \to 0. \tag{A4}$$

Lemma 2 of Rob establishes (A4). Since $d^* - d_0^* > -1/2$ for all $d^* \in \Theta_1$, Lemma 1 of Rob can be applied to yield (A3). Combining (3.9)–(3.13) of Rob, we find that (A2) holds provided that

$$\sum_{r=1}^{m-1}\left(\frac{\tilde{r}}{m}\right)^{2(\Delta-d_0^*)+1}\frac{1}{\tilde{r}^2}\left|\sum_{j=1}^{r}\left(\frac{I_j^{\mathrm{T}}}{g_{\tilde{j}}}-1\right)\right|\xrightarrow{\mathrm{p}}0 \tag{A5}$$

and

$$\frac{1}{m}\left|\sum_{j=1}^{m}\left(\frac{I_j^{\mathrm{T}}}{g_{\tilde{j}}}-1\right)\right|\xrightarrow{\mathrm{p}}0. \tag{A6}$$

It remains to demonstrate (A5) and (A6). We will prove (A5) since the proof of (A6) is similar. We have

$$\frac{I_j^{\mathrm{T}}}{g_{\tilde{j}}}-1=\left(1-\frac{g_{\tilde{j}}}{f_{\tilde{j}}}\right)\frac{I_j^{\mathrm{T}}}{g_{\tilde{j}}}+\frac{1}{f_{\tilde{j}}}(I_j^{\mathrm{T}}-|\alpha_{\tilde{j}}|^2I_{\varepsilon j}^{\mathrm{T}})+(2\pi I_{\varepsilon j}^{\mathrm{T}}-1).$$

Using Lemma 3, and arguing as immediately below (3.16) of Rob, we conclude that

$$E\sum_{r=1}^{m-1}\left(\frac{\tilde{r}}{m}\right)^{2(\Delta-d_0^*)+1}\frac{1}{\tilde{r}^2}\left|\sum_{j=1}^{r}\left(1-\frac{g_{\tilde{j}}}{f_{\tilde{j}}}\right)\frac{I_j^{\mathrm{T}}}{g_{\tilde{j}}}\right|\to 0.$$

Lemma 4 together with the tapered analog of (3.17) of Rob yields

$$E|I_j^{\mathrm{T}}-|\alpha_{\tilde{j}}|^2I_{\varepsilon j}^{\mathrm{T}}|=\mathrm{O}(j^{-\beta/2}f_{\tilde{j}}).$$

Arguing as in Rob, p. 1637, we find that

$$E\sum_{r=1}^{m-1}\left(\frac{\tilde{r}}{m}\right)^{2(\Delta-d_0^*)+1}\frac{1}{\tilde{r}^2}\left|\sum_{j=1}^{r}\frac{1}{f_{\tilde{j}}}(I_j^{\mathrm{T}}-|\alpha_{\tilde{j}}|^2I_{\varepsilon j}^{\mathrm{T}})\right|\to 0.$$

It remains to show that

$$\sum_{r=1}^{m-1}\left(\frac{\tilde{r}}{m}\right)^{2(\Delta-d_0^*)+1}\frac{1}{\tilde{r}^2}\left|\sum_{j=1}^{r}(2\pi I_{\varepsilon j}^{\mathrm{T}}-1)\right|\xrightarrow{\mathrm{p}}0. \tag{A7}$$

The validity of (A7) is established by Lemma 5. This completes the proof of Lemma 1.

PROOF OF LEMMA 2. Lemma 2 follows from Lemmas 3 and 4, using a proof along the lines given in Rob, pp. 1638–40, and in the proof of Theorem 5 of Vel.

PROOF OF LEMMA 3. Lemma 3 follows from an argument similar to that used in the proof of Robinson (1995a), Theorem 2.

PROOF OF LEMMA 4. The first part of Lemma 4 follows from the proof of Theorem 6, part (a) of Velasco (1999a, pp. 341, 354–360). The original result was given for a tapered Fourier transform of the levels $\{x_t\}$, assumed to be invertible, using a taper of order $p \geqslant 2$ as defined in Velasco (1999a, p. 339). Velasco's original proof goes through with a few minor changes for the w_j^{T}, which are based on the differences $\{y_t\}$, even in the noninvertible case $-1.5 < d_0^* \leqslant -0.5$. Define

$$K^{\mathrm{T}}(\lambda)=\frac{1}{2\pi\sum|h_u|^2}|D^{\mathrm{T}}(\lambda)|^2$$

where $D^{\mathrm{T}}(\lambda)=\sum_{t=1}^{n}h_t\exp(i\lambda t)$. It follows from Lemma 0 with $p=2$ that $K^{\mathrm{T}}(\lambda) \leqslant \text{constant}\times\min\{n,\ n^{-3}|\lambda|^{-4}\}$ for $\lambda\in[-\pi,\pi]$. (An identical bound holds for the tapers of order $p=2$ considered in Velasco (1999a).) It can be shown that

$$D^{\mathrm{T}}(\lambda) = \frac{1}{2} D_n(\lambda) - \frac{1}{2} D_n(\lambda + \lambda_1) \exp\left(-i \frac{\lambda_1}{2}\right)$$

where

$$D_n(\lambda) = \exp\left\{ i(n+1) \frac{\lambda}{2} \right\} \frac{\sin(n\lambda/2)}{\sin(\lambda/2)}$$

is the Dirichlet kernel. It follows after some algebraic manipulations that $K^{\mathrm{T}}(\cdot)$ is symmetric about $\lambda = -\lambda_{1/2}$, i.e. that $K^{\mathrm{T}}(-\lambda_{1/2} + \lambda) = K^{\mathrm{T}}(-\lambda_{1/2} - \lambda)$. Since $\int_{-\pi}^{\pi} K^{\mathrm{T}}(\lambda)\, d\lambda = 1$, we have

$$E(I_j^{\mathrm{T}} - f_{\tilde{j}}) = \int_{-\pi}^{\pi} \{ f(\lambda) - f_{j+1/2} \} K^{\mathrm{T}}(\lambda_j - \lambda)\, d\lambda.$$

The interval of integration $[-\pi, \pi]$ is partitioned into the subintervals $[-\pi, -\varepsilon]$, $[-\varepsilon, -\lambda_{j/2}]$, $[-\lambda_{j/2}, \lambda_{(j+1)/2}]$, $[\lambda_{(j+1)/2}, \lambda_{(3j+1)/2}]$, $[\lambda_{(3j+1)/2}, \varepsilon]$ and $[\varepsilon, \pi]$, where ε is a fixed number in $(0, \pi)$ such that $f(\lambda) \leqslant \text{constant} \times |\lambda|^{-2d_0^*}$ for $|\lambda| \in (0, \varepsilon)$. Bounds in these subintervals for the absolute value of the above integral can be obtained as in the proof of Velasco (1999a), with a slight modification for the subinterval $[\lambda_{(j+1)/2}, \lambda_{(3j+1)/2}]$, which we now present. Using Assumption A1, we have

$$\begin{aligned} &\left| \int_{\lambda_{(j+1)/2}}^{\lambda_{(3j+1)/2}} \{ f(\lambda) - f_{j+1/2} \} K^{\mathrm{T}}(\lambda_j - \lambda)\, d\lambda \right| \\ &\quad = \left| \int_{-\lambda_{j/2}}^{\lambda_{j/2}} \{ f(\lambda_{j+1/2} + \lambda) - f(\lambda_{j+1/2}) \} K^{\mathrm{T}}(-\lambda_{1/2} - \lambda)\, d\lambda \right| \\ &\quad = \left| \int_{-\lambda_{j/2}}^{\lambda_{j/2}} \{ \lambda f'(\lambda_{j+1/2}) + \mathrm{O}(\lambda_{j+1/2}^{-\beta-2d_0^*} |\lambda|^{\beta}) \} K^{\mathrm{T}}(-\lambda_{1/2} - \lambda)\, d\lambda \right| \\ &\quad \leqslant |f'(\lambda_{\tilde{j}})| \left| \int_{-\lambda_{j/2}}^{\lambda_{j/2}} \lambda K^{\mathrm{T}}(-\lambda_{1/2} - \lambda)\, d\lambda \right| + \lambda_{\tilde{j}}^{-\beta-2d_0^*} \mathrm{O}\left\{ \int_{-\lambda_{j/2}}^{\lambda_{j/2}} |\lambda|^{\beta} K^{\mathrm{T}}(-\lambda_{1/2} - \lambda)\, d\lambda \right\}. \end{aligned}$$

The first term on the right is zero, since the integral is

$$\int_0^{\lambda_j/2} \lambda \{ K^{\mathrm{T}}(-\lambda_{1/2} - \lambda) - K^{\mathrm{T}}(-\lambda_{1/2} + \lambda) \}\, d\lambda = 0.$$

Using the bound given earlier for K^{T}, the second term on the right can be shown to be $\mathrm{O}(f_{\tilde{j}} j^{-\beta})$. Using this together with the bounds for the other subintervals, we obtain the desired result, $E(I_j^{\mathrm{T}} - f_{\tilde{j}})/f_{\tilde{j}} = \mathrm{O}(j^{-\beta})$.

The proof of the second part of Lemma 4 is similar. First, it can be shown that

$$E\left(w_j^{\mathrm{T}} \overline{w}_{\varepsilon j}^{\mathrm{T}} - \frac{\alpha_{\tilde{j}}}{2\pi} \right) = \int_{-\pi}^{\pi} \left\{ \frac{\alpha(\lambda)}{2\pi} - \frac{\alpha(\lambda_{\tilde{j}})}{2\pi} \right\} K^{\mathrm{T}}(\lambda_j - \lambda)\, d\lambda.$$

The bound analogous to the one shown above can be obtained from Assumptions A1 and A2, which imply that for $\lambda \in [-\lambda_{j/2}, \lambda_{j/2}]$,

$$\alpha(\lambda_{j+1/2} - \lambda) - \alpha(\lambda_{j+1/2}) = -\lambda \alpha'(\lambda_{j+1/2}) + \mathrm{O}(\lambda_{j+1/2}^{-\beta-d_0^*} |\lambda|^{\beta}).$$

This leads to the conclusion that

$$\left| \int_{\lambda_{(j+1)/2}}^{\lambda_{(3j+1)/2}} \left\{ \frac{\alpha(\lambda)}{2\pi} - \frac{\alpha(\lambda_{j+1/2})}{2\pi} \right\} K^{\mathrm{T}}(\lambda_j - \lambda)\, d\lambda \right| = \mathrm{O}(\sqrt{f_{\tilde{j}}}\, j^{-\beta}).$$

Bounds for the other subintervals can be obtained in a straightforward manner.

PROOF OF LEMMA 5. We have

$$2\pi I^{\mathrm{T}}_{\varepsilon j} - 1 = \left(\frac{1}{\sum |h_u|^2}\sum_{t=1}^{n}|h_t|^2\varepsilon_t^2 - 1\right) + \frac{1}{\sum |h_u|^2}\sum\sum_{s\neq t} h_s\overline{h}_t \exp\{i(s-t)\lambda_j\}\varepsilon_s\varepsilon_t.$$

Thus,

$$\sum_{r=1}^{m-1}\left(\frac{\tilde{r}}{m}\right)^{2(\Delta-d_0^*)+1}\frac{1}{\tilde{r}^2}\left|\sum_{j=1}^{r}(2\pi I^{\mathrm{T}}_{\varepsilon j}-1)\right| \leqslant \left|\frac{1}{\sum |h_u|^2}\sum_{t=1}^{n}|h_t|^2\varepsilon_t^2 - 1\right|\sum_{r=1}^{m}\left(\frac{\tilde{r}}{m}\right)^{2(\Delta-d_0^*)+1}\frac{1}{\tilde{r}}$$

$$+\sum_{r=1}^{m-1}\left(\frac{\tilde{r}}{m}\right)^{2(\Delta-d_0^*)+1}\frac{1}{\tilde{r}^2\sum |h_u|^2}\left|\sum\sum_{s\neq t} h_s\overline{h}_t\varepsilon_s\varepsilon_t\sum_{j=1}^{r}\exp\{i(s-t)\lambda_j\}\right|. \quad \text{(A8)}$$

By Theorem 2.23, p. 44, of Hall and Heyde (1980),

$$\frac{1}{\sum |h_u|^2}\sum_{t=1}^{n}|h_t|^2\varepsilon_t^2 - 1 \xrightarrow{\mathrm{p}} 0$$

so the first term on the right of (A8) is $o_p(1)$. Now,

$$E\left|\sum\sum_{s\neq t} h_s\overline{h}_t\varepsilon_s\varepsilon_t\sum_{j=1}^{r}\exp\{i(s-t)\lambda_j\}\right|^2$$

$$\leqslant 2\sum\sum_{s\neq t}|h_s|^2|h_t|^2\left|\sum_{j=1}^{r}\exp\{i(s-t)\lambda_j\}\right|^2$$

$$\leqslant 2\sum_{s=1}^{n}\sum_{t=1}^{n}\left|\sum_{j=1}^{r}\exp\{i(s-t)\lambda_j\}\right|^2$$

$$= 2\sum_{j=1}^{r}\sum_{k=1}^{r}\sum_{s=1}^{n}\sum_{t=1}^{n}\exp\{i(s-t)\lambda_j\}\exp\{-i(s-t)\lambda_k\}$$

$$= 2\sum_{j=1}^{r}\sum_{k=1}^{r}\sum_{s=1}^{n}\exp(is\lambda_{j-k})\sum_{t=1}^{n}\exp(-it\lambda_{j-k}) = 2rn^2$$

so the expectation of the second term on the right of (A8) is

$$\mathrm{O}\left\{\sum_{r=1}^{m}\left(\frac{r}{m}\right)^{2(\Delta-d_0^*)+1}\frac{1}{r^2 n}\sqrt{(rn^2)}\right\} = \mathrm{O}\{m^{2(d_0^*-\Delta)-1} + (\log m)m^{-1/2}\} = \mathrm{o}(1).$$

PROOF OF LEMMA 6. We have

$$E\left|\sum_{j=1}^{r}\left(\frac{I^{\mathrm{T}}_j}{g_{\tilde{j}}} - 2\pi I^{\mathrm{T}}_{\varepsilon j}\right)\right| \leqslant \sum_{j=1}^{r}E\left|\frac{I^{\mathrm{T}}_j}{f_{\tilde{j}}} - 2\pi I^{\mathrm{T}}_{\varepsilon j}\right| + \sum_{j=1}^{r}E\left|\frac{I^{\mathrm{T}}_j}{g_{\tilde{j}}} - \frac{I^{\mathrm{T}}_j}{f_{\tilde{j}}}\right|.$$

Furthermore,

$$E\left|\frac{I_j^{\mathrm{T}}}{f_{\tilde{j}}}-2\pi I_{\varepsilon j}^{\mathrm{T}}\right|=\frac{1}{f_{\tilde{j}}}E|I_j^{\mathrm{T}}-2\pi f_{\tilde{j}}I_{\varepsilon j}^{\mathrm{T}}|=\frac{1}{f_{\tilde{j}}}E|I_j^{\mathrm{T}}-|\alpha_{\tilde{j}}|^2 I_{\varepsilon j}^{\mathrm{T}}|=\mathrm{O}(j^{-\beta/2})$$

by the tapered analog to (3.17) of Rob, and Lemma 4. Thus,

$$\sum_{j=1}^{r} E|I_j^{\mathrm{T}}/f_{\tilde{j}}-2\pi I_{\varepsilon j}^{\mathrm{T}}|=\mathrm{O}(r^{1-\beta/2}+\log r)$$

where the $\log r$ term appears if $\beta = 2$. Using Lemma 3, we obtain

$$E\sum_{j=1}^{r}\left|\frac{I_j^{\mathrm{T}}}{g_{\tilde{j}}}-\frac{I_j^{\mathrm{T}}}{f_{\tilde{j}}}\right|=E\sum_{j=1}^{r}\frac{I_j^{\mathrm{T}}}{f_{\tilde{j}}}\left|\frac{f_{\tilde{j}}}{g_{\tilde{j}}}-1\right|$$

$$=\mathrm{O}\left(\sum_{j=1}^{r}\left|\frac{f_{\tilde{j}}}{g_{\tilde{j}}}-1\right|\right)=\mathrm{O}\left(\sum_{j=1}^{r}\lambda_{\tilde{j}}^{\beta}\right)=\mathrm{O}(n^{-\beta}r^{\beta+1})$$

since, by Assumption A1′, $f_{\tilde{j}}/g_{\tilde{j}}-1=\mathrm{O}(\lambda_{\tilde{j}}^{\beta})$.

Proof of Lemma 7. Define $d_s=(\sum|h_u|^2)^{-1}\sum_{j=1}^{r}\exp(is\lambda_j)$. Then

$$\sum_{j=1}^{r}(2\pi I_{\varepsilon j}^{\mathrm{T}}-1)=\sum_{j=1}^{r}\left\{\left(\sum|h_u|^2\right)^{-1}\sum_{t=1}^{n}h_t\varepsilon_t\exp(i\lambda_j t)\sum_{s=1}^{n}\overline{h}_s\varepsilon_s\exp(-i\lambda_j s)-1\right\}$$

$$=r\left(\sum|h_u|^2\right)^{-1}\sum_{t=1}^{n}|h_t|^2(\varepsilon_t^2-1)+\sum_{t=2}^{n}\sum_{s=1}^{t-1}\varepsilon_t\varepsilon_s 2\,\mathrm{Re}(h_t\overline{h}_s d_{t-s}). \tag{A9}$$

Clearly, $|d_s|\leqslant r/\sum|h_u|^2$. Furthermore, $|d_s|=|d_{n-s}|$ for all s, and for $1\leqslant s\leqslant n/2$ we have

$$|d_s|=\frac{2}{n}\left|\sum_{j=0}^{r}\exp(ij\lambda_s)-1\right|\leqslant\frac{2}{n}(|\sin(\lambda_s/2)|^{-1}+1)\leqslant\frac{3}{s}.$$

The first term on the right-hand side of (A9) has mean zero and a variance of order $\mathrm{O}(r^2/n)$, since $\sum|h_u|^2=\mathrm{O}(n)$ and $|h_t|\leqslant 1$. Note that $|\mathrm{Re}(h_t\overline{h}_s d_{t-s})|\leqslant|h_t\overline{h}_s d_{t-s}|\leqslant|d_{t-s}|$. The second term on the right-hand side of (A9) has mean zero and a variance of order

$$\mathrm{O}\left(\sum_{t=2}^{n}\sum_{s=1}^{t-1}|d_{t-s}|^2\right)=\mathrm{O}\left(n\sum_{s=1}^{n}|d_s|^2\right)=\mathrm{O}\left\{n\frac{n}{r}\left(\frac{r}{n}\right)^2+n\sum_{[n/r]}^{\infty}s^{-2}\right\}=\mathrm{O}(r).$$

Proof of Lemma 8. Define

$$c_r=\frac{2}{\sum|h_u|^2}\frac{1}{\sqrt{m}}\sum_{j=1}^{m}v_j\exp(i\lambda_j r).$$

Since $\sum_{j=1}^{m}v_j=0$, we have $c_0=0$, and

$$\frac{2}{\sqrt{m}}\sum_{j=1}^{m}v_j(2\pi I_{\varepsilon j}^{\mathrm{T}}-1)=\frac{2}{\sqrt{m}}2\pi\sum_{j=1}^{m}v_jI_{\varepsilon j}^{\mathrm{T}}$$

$$=\frac{1}{\sum|h_u|^2}\frac{2}{\sqrt{m}}\sum_{j=1}^{m}v_j\sum_{t=1}^{n}h_t\varepsilon_t\exp(i\lambda_jt)\sum_{s=1}^{n}\overline{h}_s\varepsilon_s\exp(-i\lambda_js)$$

$$=\sum_{t=1}^{n}\sum_{s=1}^{n}h_t\varepsilon_t\overline{h}_s\varepsilon_sc_{t-s}$$

$$=c_0\sum_{t=1}^{n}|h_t|^2\varepsilon_t^2+\sum_{t=1}^{n}\sum_{s<t}\varepsilon_t\varepsilon_s2\,\mathrm{Re}(h_t\overline{h}_sc_{t-s})=2\sum_{t=1}^{n}z_t$$

where $z_t=\varepsilon_t\sum_{s<t}\varepsilon_sC(s,\,t)$ with $z_1=0$, and

$$C(s,\,t)=\mathrm{Re}(h_t\overline{h}_sc_{t-s})=\frac{1}{2}(h_t\overline{h}_sc_{t-s}+\overline{h}_th_s\overline{c}_{t-s}).$$

Since $\{\varepsilon_t\}$ is a martingale difference sequence, so is $\{z_t\}$. We need to show that $\sum_{t=1}^{n}z_t\xrightarrow{d}\mathrm{N}(0,\,\Phi)$. This will follow from both

$$\sum_{t=1}^{n}E(z_t^2|F_{t-1})-\Phi\xrightarrow{p}0 \tag{A10}$$

$$\sum_{t=1}^{n}E\{z_t^2\chi(|z_t|>\delta)\}\to0\qquad\text{for all }\delta>0 \tag{A11}$$

where $\chi(\cdot)$ denotes an indicator function (cf. (4.12) and (4.13) of Rob). First we prove (A10). We have

$$\sum_{t=1}^{n}E(z_t^2|F_{t-1})-\Phi=\left\{\sum_{t=1}^{n}\sum_{s=1}^{t-1}C^2(s,\,t)-\Phi\right\}+\sum_{t=1}^{n}\sum_{s=1}^{t-1}(\varepsilon_s^2-1)C^2(s,\,t)$$
$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{r\neq s}^{t-1}\varepsilon_s\varepsilon_rC(s,\,t)C(r,\,t).$$

Denote the terms on the right-hand side of the equation above by T_1, T_2 and T_3, respectively. At the end of this proof, we will show that $T_1\to0$. As for T_2, we have $E(T_2)=0$, and we can write

$$T_2=\sum_{s=1}^{n-1}(\varepsilon_s^2-1)\sum_{t=s+1}^{n}C^2(s,\,t)=\sum_{t=1}^{n-1}(\varepsilon_t^2-1)\sum_{s=t+1}^{n}C^2(t,\,s)=\sum_{t=1}^{n-1}(\varepsilon_t^2-1)\sum_{u=1}^{n-t}C^2(t,\,u+t).$$

Since, for $v<t$, $E(\varepsilon_t^2-1)(\varepsilon_v^2-1)=0$, we conclude that

$$\mathrm{var}(T_2)=\mathrm{O}\left[\sum_{t=1}^{n-1}\left\{\sum_{s=1}^{n-t}C^2(t,\,s+t)\right\}^2\right].$$

Since $C(t,\,s+t)=\mathrm{Re}(h_{s+t}\overline{h}_tc_s)$ and since $|h_t|\leqslant1$ for all t, we have $|C^2(t,\,s+t)|\leqslant|c_s|^2$. Thus,

$$\mathrm{var}(T_2) = \mathrm{O}\left\{\sum_{t=1}^{n-1}\left(\sum_{s=1}^{n-t}|c_s|^2\right)^2\right\}.$$

As in (4.20) of Rob, $|c_s| = \mathrm{O}(m^{1/2}\log m\, n^{-1})$, so the argument on p. 1646 of Rob with $|c_s|$ in place of the c_s of Rob implies that $\mathrm{var}(T_2) \to 0$, so $T_2 = \mathrm{o_p}(1)$.

As for T_3, we have $E(T_3) = 0$. Furthermore, $\mathrm{var}(T_3) = \mathrm{o}(1)$ by an argument that proceeds as follows. First, an expression analogous to that below Equation (4.22) of Rob for $\mathrm{var}(T_3)$ is obtained, with $C(\cdot, \cdot)$ replacing his c_r. Next, the bound $|C(t, s+t)| \leqslant |c_s|$ allows us to bound $\mathrm{var}(T_3)$ using the bounds of (4.23) and the equation immediately below of Rob (pp. 1646–47), which continue to hold when his c_s^2 is replaced by our $|c_r|^2$ as can be shown using the analogs of his bounds (4.20)–(4.22).

The proof of (A11) follows the argument of Rob, p. 1647. We check the sufficient condition $\sum_{t=1}^{n} E(z_t^4) \to 0$. The left-hand side of this is bounded by constant $\times\, n(\sum_{t=1}^{n}|c_t|^2)^2$, by an argument similar to the one given above, combined with the arguments of Rob, p. 1647. This expression, in turn, is $\mathrm{O}\{(\log m)^4/n\}$ by an analog to (4.22) of Rob, p. 1646.

We now prove that $T_1 \to 0$. We use ideas from the proof of Lemmas 6 and 7 of Vel (pp. 121–126), although our argument is simpler due to the special properties of the new taper. For arbitrary complex sequences $\{\alpha_t\}$, $\{\beta_s\}$, $\{\gamma_{t-s}\}$ it can be shown that

$$\sum_{t=1}^{n}\sum_{s=1}^{t-1}\alpha_t\beta_s\gamma_{t-s} = \sum_{t=1}^{n-1}\beta_t\sum_{s=1}^{n-t}\alpha_{s+t}\gamma_s. \tag{A12}$$

It can also be shown that

$$\sum_{t=1}^{n-1}|h_t|^2\sum_{s=1}^{n-t}|h_{s+t}|^2\cos(s\lambda_j) = \frac{1}{2}\left|\sum_{t=1}^{n}|h_t|^2\exp(i\lambda_j t)\right|^2 - \frac{1}{2}\sum_{t=1}^{n}|h_t|^4 \tag{A13}$$

$$\mathrm{Re}\sum_{t=1}^{n-1}\overline{h}_t^2\sum_{s=1}^{n-t}h_{t+s}^2\exp(-i\lambda_j s) = \frac{1}{2}\left|\sum_{t=1}^{n}\overline{h}_t^2\exp(i\lambda_j t)\right|^2 - \frac{1}{2}\sum_{t=1}^{n}|h_t|^4. \tag{A14}$$

Using (A12)–(A14), and noting that $C^2(s, t)$ is real, we obtain

$$\begin{aligned}
4\sum_{t=1}^{n}\sum_{s=1}^{t-1}C^2(s, t) &= \sum_{t=1}^{n}\sum_{s=1}^{t-1}(h_t^2\overline{h}_s^2c_{t-s}^2 + \overline{h}_t^2h_s^2\overline{c}_{t-s}^2 + 2|h_t|^2|h_s|^2|c_{t-s}|^2)\\
&= 2\,\mathrm{Re}\sum_{t=1}^{n-1}\overline{h}_t^2\sum_{s=1}^{n-t}h_{t+s}^2\overline{c}_s^2 + 2\sum_{t=1}^{n-1}|h_t|^2\sum_{s=1}^{n-t}|h_{s+t}|^2|c_s|^2\\
&= \frac{4}{m(\sum|h_u|^2)^2}\sum_{j=1}^{m}\sum_{k=1}^{m}v_jv_k\\
&\quad\times\left\{2\,\mathrm{Re}\sum_{t=1}^{n-1}\overline{h}_t^2\sum_{s=1}^{n-t}h_{t+s}^2\exp(-i\lambda_{j+k}s) + 2\sum_{t=1}^{n-1}|h_t|^2\sum_{s=1}^{n-t}|h_{s+t}|^2\cos(s\lambda_{j-k})\right\}\\
&= \frac{4}{m(\sum|h_u|^2)^2}\sum_{j=1}^{m}\sum_{k=1}^{m}v_jv_k\\
&\quad\times\left\{\left|\sum_{t=1}^{n}\overline{h}_t^2\exp(i\lambda_{j+k}t)\right|^2 + \left|\sum_{t=1}^{n}|h_t|^2\exp(i\lambda_{j-k}t)\right|^2 - 2\sum_{t=1}^{n}|h_t|^4\right\}.
\end{aligned}$$

For the new taper $\{h_t\}$ given by Equation (3), and for j, k in the range $1, \ldots, m$, it can be shown that

$$\left|\sum_{t=1}^{n} \overline{h}_t^2 \exp(i\lambda_{j+k} t)\right|^2 + \left|\sum_{t=1}^{n} |h_t|^2 \exp(i\lambda_{j-k} t)\right|^2$$
$$= \frac{n^2}{16}\chi(j+k=2) + \frac{n^2}{4}\chi(j-k=0) + \frac{n^2}{16}\chi(|j-k|=1).$$

Since

$$\sum_{j=1}^{m}\sum_{k=1}^{m} v_j v_k = \left(\sum_{j=1}^{m} v_j\right)^2 = 0$$

and since $\sum |h_u|^2 = n/2$, we obtain

$$4\sum_{t=1}^{n}\sum_{s=1}^{t-1} C^2(s,\ t) = \frac{4}{m(n/2)^2}\left(\frac{n^2}{16}v_1^2 + \frac{n^2}{4}\sum_{j=1}^{m} v_j^2 + \frac{n^2}{8}\sum_{j=1}^{m-1} v_j v_{j+1}\right)$$
$$= \mathrm{O}\left(\frac{\log^2 m}{m}\right) + \frac{4}{m}\sum_{j=1}^{m} v_j^2(1+0.5) + \mathrm{O}\left(\frac{1}{m}\sum_{j=1}^{m-1} |v_j|\,|v_{j+1} - v_j|\right).$$

From Rob, p. 1645,

$$\frac{1}{m}\sum_{j=1}^{m} v_j^2 = 1 + \mathrm{O}\left(\frac{\log^2 m}{m}\right).$$

Furthermore,

$$|v_{j+1} - v_j| = |\log\{(\tilde{j}+1)/\tilde{j}\}| = |\log(1 + 1/\tilde{j})| = \mathrm{O}(1/j).$$

Thus,

$$\sum_{t=1}^{n}\sum_{s=1}^{t-1} C^2(s,\ t) = 1.5 + \mathrm{O}\left(\frac{\log^2 m}{m}\right) = \Phi + \mathrm{o}(1).$$

PROOF OF LEMMA 0. We suppose $p > 1$, since it is well known that Lemma 0 holds when $p = 1$. Since $|h_t| \leqslant 1$, it follows that $|h_t^{p-1}| \leqslant 1$ and hence $|D_p^{\mathrm{T}}(\lambda)| \leqslant n$. We now show that $|D_p^{\mathrm{T}}(\lambda)| \leqslant K n^{1-p} |\lambda|^{-p}$ for $1/n \leqslant |\lambda| \leqslant \pi$. From Equation (12), we have

$$|D_p^{\mathrm{T}}(\lambda)| \leqslant \left|\sum_{k=0}^{p-1}\binom{p-1}{k}(-1)^k \frac{1}{\sin(\lambda + \lambda_k)/2}\right|.$$

If $\tilde{g}(\lambda) = 1/\sin(\lambda)$ and $\tilde{g}^{(L)}(\lambda) = \partial^L \tilde{g}(\lambda)/\partial\lambda^L$, then by Taylor's theorem with remainder,

$$\tilde{g}\left(\frac{\lambda + \lambda_k}{2}\right) = \frac{1}{\sin(\lambda/2)} + \sum_{L=1}^{p-1}\frac{\tilde{g}^{(L)}(\lambda/2)}{L!}\left(\frac{\pi k}{n}\right)^L + \frac{\tilde{g}^{(p)}(\xi_k)}{p!}\left(\frac{\pi k}{n}\right)^p$$

where $\lambda/2 \leqslant \xi_k \leqslant (\lambda + \lambda_k)/2$. It follows that

$$\sum_{k=0}^{p-1}\binom{p-1}{k}(-1)^k\frac{1}{\sin(\lambda+\lambda_k)/2}$$

$$=\sum_{L=1}^{p-1}\frac{\tilde{g}^{(L)}(\lambda/2)}{L!}\left(\frac{\pi}{n}\right)^L\sum_{k=0}^{p-1}\binom{p-1}{k}(-1)^k k^L+\sum_{k=0}^{p-1}\binom{p-1}{k}(-1)^k\frac{\tilde{g}^{(p)}(\xi_k)}{p!}\left(\frac{\pi k}{n}\right)^p. \tag{A15}$$

We will show below that

$$|\tilde{g}^{(L)}(\lambda)|\leqslant\frac{C_L}{\sin^{L+1}(\lambda)}\qquad L\geqslant 1,\ \lambda\in(0,\pi] \tag{A16}$$

where C_L is a constant depending on L, and that

$$\sum_{k=0}^{p-1}\binom{p-1}{k}(-1)^k k^L=\begin{cases}0 & 1\leqslant L\leqslant p-2\\ (-1)^{p-1}(p-1)! & L=p-1.\end{cases} \tag{A17}$$

Thus, the first term on the right-hand side of Equation (A15) is

$$\tilde{g}^{(p-1)}(\lambda/2)(\pi/n)^{p-1}(-1)^{p-1}=\mathrm{O}\{n^{1-p}\sin^{-p}(\lambda/2)\}=\mathrm{O}(n^{1-p}|\lambda|^{-p})$$

and the second term is

$$\mathrm{O}\left\{n^{-p}\sum_{k=0}^{p-1}|\tilde{g}^{(p)}(\xi_k)|k^p\right\}=\mathrm{O}(n^{-p}|\lambda|^{-(p+1)})=\mathrm{O}(n^{1-p}|\lambda|^{-p})$$

since $|\lambda|>1/n$.

It remains to prove Equations (A16) and (A17). It can be shown by induction that

$$\tilde{g}^{(L)}(\lambda)=(-1)^L L!\sin^{-(L+1)}(\lambda)\cos^L(\lambda)+\tilde{h}(\lambda)$$

where $\tilde{h}(\lambda)$ is a linear combination of products of the form $\sin^{-a}(\lambda)\cos^b(\lambda)$ with $0<a\leqslant L$, $0\leqslant b\leqslant L$, so Equation (A16) holds. To prove Equation (A17), for a given L, define $h(x)=x^L$, a polynomial in the real variable x. Define the differencing operator Δ by $\Delta h(x)=h(x)-h(x-1)$. Then $\Delta h(x)$ is a polynomial in x of order $L-1$ with leading term Lx^{L-1}. It follows inductively that $\Delta^L h(x)\equiv L!$ and $\Delta^m h(x)\equiv 0$ if $m>L$. For any integer $m\geqslant 1$,

$$\begin{aligned}\Delta^m h(x)|_{x=m}&=\sum_{k=0}^{m}\binom{m}{k}(-1)^k h(m-k)\\&=\sum_{k=0}^{m}\binom{m}{k}(-1)^{m-k}h(k)\\&=(-1)^m\sum_{k=0}^{m}\binom{m}{k}(-1)^k k^L\\&=\begin{cases}0 & 1\leqslant L\leqslant m-1\\ m! & L=m.\end{cases}\end{aligned} \tag{A18}$$

Equation (A17) follows on setting $m=p-1$ in Equation (A18).

ACKNOWLEDGEMENT

The authors thank the Associate Editor and Referee for suggestions that substantially improved the paper. The authors also thank Rohit Deo for many helpful discussions and suggestions.

REFERENCES

BAILLIE, R. T., CHUNG, C.-F. and TIESLAU, M. A. (1996) Analysing inflation by the fractionally integrated ARFIMA-GARCH model. *J. Appl. Economet.* 11, 23–40.

BERAN, J. (1994) *Statistics for Long-memory Processes*. New York: Chapman & Hall.

BLOOMFIELD, P. (1976) *Fourier Analysis of Time Series: An Introduction*. New York: Wiley.

CHEUNG, Y.-W. (1993) Long memory in foreign exchange rates. *J. Bus. Econ. Stat.* 11, 93–101.

—— and DIEBOLD, F. X. (1994) On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean. *J. Economet.* 62, 301–16.

—— and LAI, K. S. (1993) A fractional cointegration analysis of purchasing power parity. *J. Bus. Econ. Stat.* 11, 103–12.

CRATO, N. and DE LIMA, P. J. F. (1997) On the power of underdifferencing and overdifferencing tests against nearly nonstationary alternatives. *Commun. Stat. Simul. Comput.* 26, 1431–46.

DAVIES, R. B. and HARTE, D. S. (1987) Tests for Hurst effect. *Biometrika* 74, 95–102.

DAVIS, R. A. and DUNSMUIR, W. T. M. (1996) Maximum likelihood estimation of MA(1) processes with a root on or near the unit circle. *Economet. Theory* 12, 1–29.

DELGADO, M. and ROBINSON, P. M. (1994) New methods for the analysis of long-memory time-series: application to Spanish inflation. *J. Forecasting* 13, 97–107.

DEO, R. S. and HURVICH, C. M. (1998) Linear trend with fractionally integrated errors. *J. Time Ser. Anal.* 19, 379–397.

GEWEKE, J. and PORTER-HUDAK, S. (1983) The estimation and application of long memory time series models. *J. Time Ser. Anal.* 4, 221–38.

HALL, P. and HEYDE, C. C. (1980) *Martingale Limit Theory and its Applications*. Boston, MA: Academic Press.

HASSLER, U. (1993) Unit root tests: the autoregressive approach in comparison with the periodogram regression. *Stat. Pap.* 34, 67–82.

—— and WOLTERS, J. (1995) Long memory in inflation rates: international evidence. *J. Bus. Econ. Stat.* 13, 37–45.

HENRY, M. and ROBINSON, P. M. (1996) Bandwidth choice in Gaussian semiparametric estimation of long range dependence. In *Athens Conference on Applied Probability and Time Series* (eds P. M. Robinson and M. Rosenblatt). New York: Springer, pp. 220–32.

HURVICH, C. M., DEO, R. S. and BRODSKY, J. (1998) The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *J. Time Ser. Anal.* 19, 19–46.

HURVICH, C. M. and RAY, B. K. (1995) Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *J. Time Ser. Anal.* 16, 17–41.

KUNSCH, H. R. (1987) Statistical aspects of self-similar processes. In *Proceedings of the First World Congress of the Bernoulli Society*, Vol. 1 (eds Yu. Prohorov and V. V. Sazanov). Utrecht: VNU Science Press, pp. 67–74.

LOBATO, I. and ROBINSON, P. M. (1998) A nonparametric test for I(0). *Rev. Econ. Stud.* 65, 475–95.

ROBINSON, P. M. (1995a) Log-periodogram regression of time series with long range dependence. *Ann. Stat.* 23, 1048–72.

—— (1995b) Gaussian semiparametric estimation of long range dependence. *Ann. Stat.* 23, 1630–61.

SAMAROV, A. and TAQQU, M. S. (1988) On the efficiency of the sample mean in long-memory noise. *J. Time Ser. Anal.* 9, 191–200.

TUKEY, J. W. (1967) An introduction to the calculations of numerical spectrum analysis. In *Advanced Seminar on Spectral Analysis of Time Series* (ed. B. Harris). New York: Wiley, pp. 25–46.

Velasco, C. (1999a) Non-stationary log-periodogram regression. *J. Economet.* 91, 325–371.
—— (1999b) Gaussian semiparametric estimation of non-stationary time series. *J. Time Ser. Anal.* 20, 87–127.
Zhurbenko, I. G. (1979) On the efficiency of estimates of a spectral density. *Scand. J. Stat.* 6, 49–56.