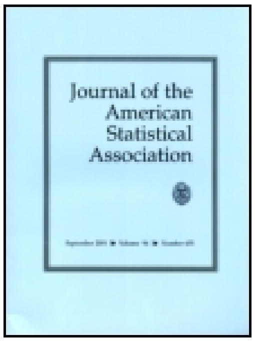

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Estimation of Long Memory in the Presence of a Smooth Nonparametric Trend

Clifford Hurvich, Gabriel Lang & Philippe Soulier

Estimation of Long Memory in the Presence of a Smooth Nonparametric Trend

Clifford HURVICH, Gabriel LANG, and Philippe SOULIER

We consider semiparametric estimation of the long-memory parameter of a stationary process in the presence of an additive nonparametric mean function. We use a semiparametric Whittle-type estimator, applied to the tapered, differenced series. Because the mean function is not necessarily a polynomial of finite order, no amount of differencing will completely remove the mean. We establish a central limit theorem for the estimator of the memory parameter, assuming that a slowly increasing number of low frequencies are trimmed from the estimator's objective function. We find in simulations that tapering and trimming, applied either separately or together, are essential for the good performance of the estimator in practice. In our simulation study, we also compare the proposed estimator of the long-memory parameter with a direct estimator obtained from the raw data without differencing or tapering, and finally we study the question of feasible inference for the regression function. We find that the proposed estimator of the long-memory parameter is potentially far less biased than the direct estimator, and consequently that the proposed estimator may lead to more accurate inference on the regression function.

KEY WORDS: Long-range dependence; Nonparametric regression; Periodogram; Tapering.

1. INTRODUCTION

The semiparametric estimation of long memory for weakly stationary univariate series has been studied extensively (see, e.g., Robinson 1994, 1995a,b; Hurvich, Deo, and Brodsky 1998; Moulines and Soulier 1999). Generalizations to the case where additive polynomial trends may be present were considered by Velasco (1999a,b), Hurvich and Chen (2000), and Hurvich, Moulines, and Soulier (2002). All four of these articles use tapering schemes, and the latter two use differencing before tapering.

The idea of differencing to detrend the data followed by tapering to handle difficulties induced by possible noninvertibility was suggested by Hart (1989) in a nonparametric context. In the presence of polynomial trends, adequate differencing will completely annihilate the trends, but if the trend is an arbitrary smooth function, then differencing serves as only an approximate detrending device. Hart (1989) focused on estimation of the autocovariances of the noise process (stochastic component), which was assumed to have short memory, in the presence of a smooth additive nonparametric signal (trend). Here we explore the use of differencing and tapering for estimation of the memory parameter of a long-memory noise process in the presence of a smooth additive nonparametric signal.

There is an existing literature on long memory in the presence of nonpolynomial trends. Künsch (1986) discussed the difficulty of distinguishing certain monotonic trends from long memory. Hall and Hart (1990), Csörgo and Mielnikzuk (1995a,b), and Deo (1997) discussed the properties of kernel estimators of the mean function in the presence of a long-memory noise. Robinson (1997) discussed this same topic, and also provided a method for estimating the memory parameter in the presence of a nonparametric signal. Our theoretical results here focus only on estimation of the memory parameter of the noise, not on estimation of the signal. Nevertheless, our estimator of the memory parameter of the noise, which does not require any preliminary estimator of the signal, might be useful for estimating standard errors for the signal or in constructing optimal estimators of the signal.

Robinson (1997) showed that the memory parameter of the noise may be estimated consistently from the raw data, even in the presence of an unknown nonparametric signal. He established the $\log^{1/2}(n)$-consistency of the estimator, a rate that was sufficient for the purpose at hand, under conditions on the bandwidth that become extremely stringent as the short-memory case is approached. Our procedure, which applies the Gaussian semiparametric estimator (see Robinson 1995b) to the tapered, differenced data, yields $n^{2/5-\delta}$-consistency and asymptotic normality for any sufficiently small positive value of δ, assuming that the process is stationary with a spectral pole at zero frequency.

Section 2 presents the main theoretical results on consistency and asymptotic normality of the proposed estimator. It is assumed in these theorems that an increasing number of low frequencies are trimmed from the objective function of the estimator. Stronger tapers lead to a decrease in the amount of trimming required. In Section 3 we investigate the possibility of using trimming alone, without any differencing or tapering. We find that this direct approach, a trimmed version of the estimator of Robinson (1997), can attain $n^{2/5-\delta}$-consistency, but that an excessive amount of trimming is required, so this approach would not be expected to work well in practice even with thousands of observations. Simulation results, presented in Section 4, examine the performance of the proposed estimator and explore the effects of tapering and trimming. Furthermore, a finite-sample correction to the asymptotic variance eventually agrees well with the variances of the estimators as found in our simulations. In our simulation study, we also compare the proposed estimator of the long-memory parameter with the direct estimator considered by Robinson (1997), and finally we study the question of feasible inference for the regression function. We find that the proposed estimator of the long-memory parameter is potentially far less biased than the direct estimator, and consequently the proposed estimator may lead to more accurate inference on the regression function. Section 5 presents an application of our proposed methodology to a series of monthly

Clifford Hurvich is Professor of Statistics, New York University, New York, NY 10012 (E-mail: *churvich@stern.nyu.edu*). Gabriel Lang is Senior Researcher, UMR 518 MIA, INRA, INAP–G, ENGREF, 75732 Paris Cedex 15, France. Philippe Soulier is Professor of Mathematics, University of Paris X, 92000 Nanterre Cedex, France. The authors thank the editor, the associate editor, and the anonymous referees for their close attention to this article, and their helpful suggestions.

Journal of the American Statistical Association
September 2005, Vol. 100, No. 471, Theory and Methods
DOI 10.1198/016214504000002096

global temperatures. The article concludes with a mathematical Appendix, containing proofs of theoretical results.

2. ASSUMPTIONS AND MAIN RESULTS

We consider the model

$$X_t = r(t/n) + \epsilon_t, \qquad t = 0, \dots, n, \tag{1}$$

where r is a sufficiently smooth function and ϵ is a linear process with long-range dependence. Denote

$$Y_t = X_t - X_{t-1},$$
$$\Delta r(t/n) = r(t/n) - r((t-1)/n), \qquad \text{and}$$
$$\eta_t = \epsilon_t - \epsilon_{t-1}.$$

This yields

$$Y_t = \Delta r(t/n) + \eta_t, \qquad t = 1, \dots, n.$$

We assume that the process η is linear with respect to a mean-0 unit variance white noise $Z = (Z_t)_{t\in\mathbb{Z}}$,

$$\eta_t = \sum_{j\in\mathbb{Z}} a_j Z_{t-j}, \qquad \sum_{j\in\mathbb{Z}} a_j^2 < \infty. \tag{2}$$

We further assume that the spectral density of η, denoted by f, can be expressed as

$$f(x) = |x|^{-2d_0} f^*(x), \tag{3}$$

where d_0 represents the memory parameter of the differenced series and $f^*(x)$ satisfies some smoothness condition in the neighborhood of zero frequency (see the assumptions in the statement of Thm. 1). In the sequel we will assume that the original series is stationary with long memory; thus, $d_0 \in I \subset (-1, -1/2)$.

Let h be a complex-valued function defined on $[0, 1]$. For any positive integer n, define $H_n = \sum_{t=1}^{n} |h(t/n)|^2$. The tapered discrete Fourier transform (DFT) and the tapered periodogram ordinates of a process ξ are defined as

$$d_{\xi,k} = \frac{1}{\sqrt{2\pi H_n}} \sum_{t=1}^{n} h(t/n)\xi_t \mathrm{e}^{ix_k t} \qquad \text{and}$$
$$I_{\xi,k} = |d_{\xi,k}|^2, \tag{4}$$

where $x_k := 2\pi k/n$ ($k = 1, \dots, [(n-1)/2]$) are the Fourier frequencies.

We use the following assumptions on the mean function r and on the taper h. Throughout the article, p is a fixed integer. For our theoretical results, we assume that $p \geq 1$, whereas in the simulation section we also consider $p = 0$, that is, no tapering.

(A1) r is a $p+1$ times continuously differentiable function on $[0, 1]$.

(A2) $h(x) = \sum_{u=0}^{v} b_u \mathrm{e}^{2i\pi xu}$ for a nonnegative integer $v \geq p$ and real coefficients b_u, $0 \leq u \leq v$, such that

$$\sum_{u=0}^{v} b_u u^k = 0 \quad \text{for } k = 0, \dots, p-1.$$

A function h that satisfies (A2) has the following important properties:

a. h has at least $p-1$ vanishing derivatives at 0 and 1,

$$h^{(j)}(0) = h^{(j)}(1) = 0, \qquad j = 0, \dots, p-1. \tag{5}$$

b. There exists a constant C such that for all $x \in [-\pi, \pi]$,

$$\left|\sum_{t=1}^{n} h(t/n)\mathrm{e}^{itx}\right| \leq C \frac{n}{(1+n|x|)^{p+1}}. \tag{6}$$

An example of a function h that satisfies (A2) is $h(x) = (1 - e^{2i\pi x})^p$. This yields the family of tapers introduced by Hurvich and Chen (2000). For this taper, which has $v = p$, (5) clearly holds, and (6) was proved by Hurvich and Chen (2000). In fact, it is easily seen that (5) implies (6). Chen (2001) has constructed certain tapers that satisfy (A2). These new tapers may have improved efficiency properties compared with the Hurvich–Chen tapers.

The following bound for the tapered DFT of the differenced mean function $d_{\Delta r,k}$ is crucial for the derivation of the properties of our estimator. It should be noted that the lemma does not require the specific form for the taper as given in assumption (A2).

Lemma 1. Assume (A1) and let h be a complex-valued p times continuously differentiable function that satisfies (5). Then there exists a constant C such that, for $1 \leq k \leq n/2$,

$$|d_{\Delta r,k}| \leq C\big(k^{-p} n^{-1/2} + n^{-3/2}\big). \tag{7}$$

We now introduce the assumptions on the structure of the process η:

(A3) (Z_l) is a fourth-order homoscedastic martingale difference sequence; that is, almost surely,

$$E[Z_k|\mathcal{F}_{k-1}] = 0, \qquad \mathbb{E}[Z_k^2|\mathcal{F}_{k-1}] = 1, \qquad \text{and}$$
$$\mathbb{E}[Z_k^4|\mathcal{F}_{k-1}] = \mu_4,$$

where $\mathcal{F}_k = \sigma(Z_l, l \leq k)$.

(A4) $a(x) := \sum_{j\in\mathbb{Z}} a_j \mathrm{e}^{ijx}$ can be expressed as $a(x) = x^{-d_0} \times a^*(x)$ $(x > 0)$, where $d_0 \in (-1, -1/2)$ and a^* is twice continuously differentiable in a neighborhood $[-\vartheta, \vartheta]$ of 0 and absolutely integrable on $[-\pi, \pi]$.

The local Whittle contrast is defined as

$$W_m(C, d) = \sum_{k=\ell}^{m} \{\log(C x_k^{-2d}) + C^{-1} x_k^{2d} I_{Y,k}\}, \tag{8}$$

where $m < n/2$ is a bandwidth parameter and $\ell < m$ is a lower trimming number. Concentrating C out of W_m yields the profile likelihood

$$\hat{J}_{\ell,m}(d) = \log\left(\frac{1}{m-\ell+1} \sum_{k=\ell}^{m} k^{2d} I_{Y,k}\right) - 2d\gamma_{\ell,m}, \tag{9}$$

where $\gamma_{\ell,m} = \frac{1}{m-\ell+1} \sum_{k=\ell}^{m} \log(k)$. We define the local Whittle (or Gaussian semiparametric in the terminology of Robinson 1995b) estimate of d_0 as

$$\hat{d}_n = \arg\min_{d\in[-1,-1/2]} \hat{J}_{\ell,m}(d).$$

Theorem 1. Assume that (A1)–(A4) hold. If ℓ and m are nondecreasing sequences of integers such that $\ell/m \to 0$ and

$$m/n + n/m\ell^{1+2p} = O(n^{-\zeta}) \tag{10}$$

for some $\zeta > 0$, then $\hat{d}_n - d_0 = O_P(n^{-\eta})$ for some $\eta > 0$.

Remark 1. A suitable choice of the sequences ℓ and m is $\ell = [n^{\delta/2p}]$ and $m = [n^{1-\delta}]$ for some arbitrarily small $\delta > 0$.

Theorem 2. Assume (A1)–(A4). Let ℓ and m be nondecreasing sequences of integers such that (10) holds and

$$\lim_{n\to\infty} \{n \log(m)/(m^{1/2}\ell^{1+2p}) + m^5/n^4 + \ell/m\} = 0. \tag{11}$$

Then $m^{1/2}(\hat{d}_n - d_0)$ converges weakly to the Gaussian distribution with mean-0 and variance

$$\frac{\sum_{z=-v}^{v} (\sum_{u=0}^{v-|z|} b_u b_{u+|z|})^2}{4(\sum_{u=0}^{v} b_u^2)^2}.$$

Remark 2. A suitable choice of the sequences ℓ and m is $\ell = [n^{3/(5(1+2p))+\delta/(4p)}]$ and $m = [n^{4/5-\delta}]$ for some arbitrarily small $\delta \in (0, 4/5)$. Hence the number of trimmed lower frequencies is not too high.

Remark 3. For the Hurvich–Chen taper of order p, the asymptotic variance is $(4p)!(p!)^4/\{4(2p)!^4\}$.

3. DIRECT ESTIMATION OF THE MEMORY PARAMETER WITH TRIMMING

Robinson (1997) showed that the memory parameter of the original sequence $\{X_t\}$ can be estimated consistently by the standard Gaussian semiparametric or local Whittle estimator. The rate of convergence that Robinson obtained is $\log^{1/2}(n)$. In this section we prove that, in theory, trimming can improve this rate of convergence. In our context, the memory parameter of $\{X_t\}$ is $d^* = d_0 + 1 \in (0, 1/2)$.

In this section only we consider the nontapered periodogram of the nondifferenced data,

$$I_{X,k} = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} X_t e^{itx_k} \right|^2.$$

We then define the objective function and the estimator of $d^* = d_0 + 1$ by

$$\tilde{J}_{\ell,m}(d) = \log\left(\frac{1}{m-\ell+1} \sum_{k=\ell}^{m} k^{2d} I_{X,k}\right) - 2d\gamma_{\ell,m}$$

and

$$\tilde{d}_n = \arg\min_{d\in[0,1/2]} \tilde{J}_{\ell,m}(d).$$

Proposition 1. Assume that (A3) and (A4) hold and that r is twice continuously differentiable. Set $m = [n^{4/5-\epsilon}]$ and $\ell = [n^{3/5+\epsilon}]$. Then $\tilde{d}_n - d^* = O_P(n^{-2/5+\epsilon})$.

A proof of this result is sketched in Section A.4. Thus trimming alone without tapering or differencing allows for the efficient estimation of d^*, in terms of the rate of convergence. In small and moderate-sized samples, however (even with thousands of observations), the tremendous amount of trimming required would render this procedure useless from a practical standpoint.

4. SIMULATIONS

4.1 Properties of the Proposed Estimator of the Memory Parameter

We investigated the properties of the estimator $\hat{d}_n$ based on first-differences of $\{X_t\}$, where $\{X_t\}$ is generated by model (1). We took the noise process $\{\epsilon_t\}$ to be a Gaussian ARFIMA(1, $d_0 + 1$, 0) process with memory parameter $d_0 + 1 = .4$, and AR(1) parameter (lag-1 autocorrelation of short-memory component) equal to .2. We considered two values of the mean function, $r(x) = 10x^4$ and $r(x) \equiv 0$. We considered both fixed and adaptively selected bandwidths m for estimation of d_0. The fixed bandwidth was $m = n^{.7}$. This choice is somewhat arbitrary, but our intent was to select an exponent that is relatively large for the sake of efficiency, but not greater than or equal to .8, to rule out asymptotic bias. We describe the adaptive bandwidth choice later. We applied pth-order Hurvich–Chen tapers $h(x) = (1 - e^{i2\pi x})^p$ with $p = 0, 1, 2$ to the first differences, which have a memory parameter of $d_0 = -.6$. Because the differences are noninvertible, the standard theory of Robinson (1995b) for the nontapered case ($p = 0$) does not apply, but we present this nontapered case for the sake of comparison. We considered three sample sizes, $n = 100, 1{,}000, 5{,}000$. For each of these sample sizes and each choice of the mean function r, we simulated 500 realizations of the process, using the method of Davies and Harte (1987).

For computation of $\hat{d}_n$, we used a slightly modified version the profile likelihood (9),

$$\hat{J}^*_{\ell,m}(d) = \log\left(\frac{1}{m-\ell+1} \sum_{k=\ell}^{m} (k + p/2)^{2d} I_{Y,k}\right) - \frac{2d}{m-\ell+1} \sum_{k=\ell}^{m} \log(k + p/2).$$

We use $k + p/2$ here in place of k in (9), because it is advisable to consider $I_{Y,k}$ as an estimator of $f(2\pi(k+p/2)/n)$ rather than as an estimator of $f(2\pi k/n)$. More discussion on this slight frequency shifting due to the taper has been given by Hurvich and Chen (2000) and Hurvich, Moulines, and Soulier (2002).

We minimized the profile likelihood $\hat{J}^*_{\ell,m}$ for d on a grid of mesh size .01, in the range $[-.99, -.51]$. This range is allowed by our theoretical results.

We considered three choices for the degree of trimming: $\ell = 1$, $\ell = n^{.15}$, and $\ell = n^{.25}$, denoted by Tr0, Tr1, and Tr2. The choice $\ell = 1$ corresponds to no trimming and is included for the sake of comparison, although our theoretical results do not apply in this case. The choices $\ell = n^{.15}$ and $\ell = n^{.25}$ satisfy the condition given in Remark 2 for $p = 2$ and $p = 1$. For $n = 5{,}000$, these values of ℓ reduce to 8 and 3.

The adaptive choice of m is a modified version of a plug-in method described by Henry (2001), based on an approximation of Delgado and Robinson (1996) to the optimal bandwidth for the Gaussian semiparametric estimator,

$$m^{opt} = (3/4\pi)^{4/5} |\tau^* + (d_0 + 1)/12|^{-2/5} n^{4/5},$$

where $\tau^* = f^{*\prime\prime}(0)/2f^*(0)$, and we have used $d_0 + 1$ instead of d_0 because we are working here with differenced data. We first estimate $K^* = 2\tau^*$ as the coefficient $\hat{K}^*$ of $\{x_k^2/2\}_{k=\ell}^{M}$

in a regression of $\{\log I_{Y,k}\}_{k=\ell}^{M}$ on $\{\log|2\sin(x_{\tilde{k}}/2)|\}_{k=\ell}^{M}$ and $\{x_{\tilde{k}}^2/2\}_{k=\ell}^{M}$, where $M=.2n^{6/7}$ and $x_{\tilde{k}}=2\pi(k+p/2)/n$. This estimator, along with this particular arbitrary choice of M, was used by Hurvich and Deo (1999) and has the advantage of not requiring an auxiliary estimator of d_0, as was required in the method of Henry (2001), who used a regression of the periodogram rather than the log periodogram. We now proceed iteratively, precisely as was done by Henry (2001), starting with an initial bandwidth $\hat{m}(0)=n^{4/5}$ and then computing for $k=0,1,2$ both $\hat{d}^{(k)}=\arg\min \hat{J}^*_{\ell,\hat{m}^{(k)}}(d)$ and

$$\hat{m}^{(k+1)}=(3/4\pi)^{4/5}\left|\hat{K}^*/2+\left(\hat{d}^{(k)}+1\right)/12\right|^{-2/5}n^{4/5},$$

and using $\hat{m}^{(2)}$ and $\hat{d}^{(2)}$ as the adaptively selected bandwidth and corresponding estimate of d_0.

We have not undertaken an adaptive choice of ℓ, because the plug-in methods do not give any guidance on this question. A possible adaptive choice of ℓ was mentioned in the context of a cross-validatory method proposed by Hurvich and Beltrao (1994), although it would be difficult to generalize this methodology to allow for tapering.

Tables 1 and 2 give the bias and root mean squared error (RMSE) for $\hat{d}_n$, and Figures 1 and 2 give boxplots for $\hat{d}_n$ for the cases of the quartic mean function $r(x)=10x^4$ and the constant mean function $r(x)\equiv 0$. The figures pertain only to the fixed bandwidth case, whereas the tables give results for both the fixed and the adaptive bandwidths. For the quartic mean function, the failure to use tapering or trimming ($p=0$, $\ell=1$) leads to substantially biased estimators, whereas the bias is reduced considerably by tapering and trimming, applied separately or together. In general, the stronger the taper or the greater the degree of trimming, the smaller the bias. The reduced bias for $p=2$ compared with $p=1$ in the case of fixed bandwidth $m=n^{.7}$ is consistent with Lemma 1, because the lemma says that the magnitude of the DFT of the mean function becomes smaller when p is increased, and because the non-0 mean function makes a substantial contribution to the bias of $\hat{d}_n$. For the case of the constant mean function and fixed bandwidth, the bias in all of the estimators is attributable to the short-memory component of the process, and the bias generally decreases as n increases. The use of an adaptive bandwidth usually had a reasonably small effect on the bias except when $n=100$, but almost always gave a substantially greater RMSE compared with the corresponding case with fixed bandwidth. For the quartic mean function, there was typically some degree of tapering or trimming that gave substantially reduced RMSE compared with $p=0$, $\ell=1$. For the constant mean function, the choice at $p=0$, $\ell=1$ was typically best or nearly best, as would be expected because tapering and trimming both have the effect of inflating variance, whereas the potential for bias reduction is not so great here as in the case of the quartic mean function.

Table 1. Bias (RMSE) for $\hat{d}_n$, Quartic Mean Function, $r(x)=10x^4$

n	ℓ	p = 0	1	2
$m=n^{.7}$				
100	1	$.090_{(.090)}$	$.052_{(.092)}$	$.030_{(.112)}$
	$n^{.15}$	$.090_{(.090)}$	$.052_{(.092)}$	$.030_{(.112)}$
	$n^{.25}$	$.057_{(.096)}$	$.032_{(.112)}$	$.022_{(.128)}$
1,000	1	$.088_{(.088)}$	$.028_{(.058)}$	$.013_{(.064)}$
	$n^{.15}$	$.049_{(.063)}$	$.014_{(.060)}$	$.010_{(.068)}$
	$n^{.25}$	$.030_{(.060)}$	$.016_{(.067)}$	$.013_{(.074)}$
5,000	1	$.074_{(.076)}$	$.017_{(.036)}$	$.011_{(.039)}$
	$n^{.15}$	$.023_{(.038)}$	$.011_{(.036)}$	$.011_{(.040)}$
	$n^{.25}$	$.016_{(.037)}$	$.013_{(.040)}$	$.012_{(.044)}$
Adaptive m				
100	1	$.089_{(.090)}$	$-.031_{(.171)}$	$-.082_{(.216)}$
	$n^{.15}$	$.089_{(.090)}$	$-.031_{(.171)}$	$-.082_{(.216)}$
	$n^{.25}$	$-.078_{(.224)}$	$-.101_{(.240)}$	$-.108_{(.248)}$
1,000	1	$.089_{(.089)}$	$.017_{(.087)}$	$-.021_{(.118)}$
	$n^{.15}$	$.047_{(.077)}$	$-.016_{(.107)}$	$-.031_{(.132)}$
	$n^{.25}$	$.003_{(.101)}$	$-.025_{(.130)}$	$-.031_{(.141)}$
5,000	1	$.086_{(.087)}$	$.014_{(.053)}$	$-.006_{(.065)}$
	$n^{.15}$	$.026_{(.056)}$	$-.002_{(.058)}$	$-.006_{(.066)}$
	$n^{.25}$	$.016_{(.056)}$	$.003_{(.062)}$	$-.001_{(.070)}$

Table 2. Bias (RMSE) for $\hat{d}_n$, Constant Mean Function, $r(x)\equiv 0$

n	ℓ	p = 0	1	2
$m=n^{.7}$				
100	1	$.054_{(.086)}$	$.036_{(.099)}$	$.028_{(.111)}$
	$n^{.15}$	$.054_{(.086)}$	$.036_{(.099)}$	$.028_{(.111)}$
	$n^{.25}$	$.049_{(.093)}$	$.031_{(.112)}$	$.022_{(.128)}$
1,000	1	$.035_{(.057)}$	$.014_{(.056)}$	$.011_{(.064)}$
	$n^{.15}$	$.030_{(.056)}$	$.014_{(.060)}$	$.011_{(.068)}$
	$n^{.25}$	$.028_{(.060)}$	$.016_{(.067)}$	$.013_{(.074)}$
5,000	1	$.023_{(.039)}$	$.009_{(.034)}$	$.010_{(.039)}$
	$n^{.15}$	$.018_{(.036)}$	$.010_{(.036)}$	$.011_{(.040)}$
	$n^{.25}$	$.016_{(.036)}$	$.013_{(.040)}$	$.012_{(.044)}$
Adaptive m				
100	1	$-.035_{(.165)}$	$-.079_{(.207)}$	$-.080_{(.217)}$
	$n^{.15}$	$-.035_{(.165)}$	$-.079_{(.207)}$	$-.080_{(.217)}$
	$n^{.25}$	$-.097_{(.237)}$	$-.111_{(.247)}$	$-.116_{(.253)}$
1,000	1	$.022_{(.076)}$	$-.013_{(.091)}$	$-.018_{(.103)}$
	$n^{.15}$	$.009_{(.086)}$	$-.019_{(.104)}$	$-.027_{(.118)}$
	$n^{.25}$	$-.012_{(.114)}$	$-.032_{(.140)}$	$-.038_{(.148)}$
5,000	1	$.023_{(.048)}$	$.003_{(.047)}$	$.003_{(.053)}$
	$n^{.15}$	$.015_{(.049)}$	$.001_{(.051)}$	$.001_{(.057)}$
	$n^{.25}$	$.010_{(.055)}$	$.003_{(.061)}$	$.001_{(.066)}$

We next consider the variance of the estimators. We focus on the case of fixed bandwidth, $m=n^{.7}$, and limit ourselves to the cases ($p=0$, $\ell=1$), ($p=1$, $\ell=n^{.25}$), and ($p=2$, $\ell=n^{.15}$), so that Theorem 2 applies in the second two cases. Table 3 first presents some theoretical variances, for the same values of n and m used in the simulations. In Table 3, the asymptotic variance, based on Theorem 2, is $\Phi_p/(4m)$, where $\Phi_p=(4p)!(p!)^4((2p)!)^{-4}$ and the finite-sample corrected variance is given by

$$\frac{\Phi_p}{4\sum_{k=\ell}^{m}[\log(k+p/2)-(m-\ell+1)^{-1}\sum_{j=\ell}^{m}\log(j+p/2)]^2}.$$

The corrected expression, which is asymptotically equivalent to the asymptotic expression, was justified heuristically by Hurvich and Chen (2000). Finally, Table 3 presents the variances from the corresponding simulations, with m fixed, and the mean-0 function.

For $p=0$, the simulation variances do not agree well with either of the theoretical variance expressions. Perhaps these

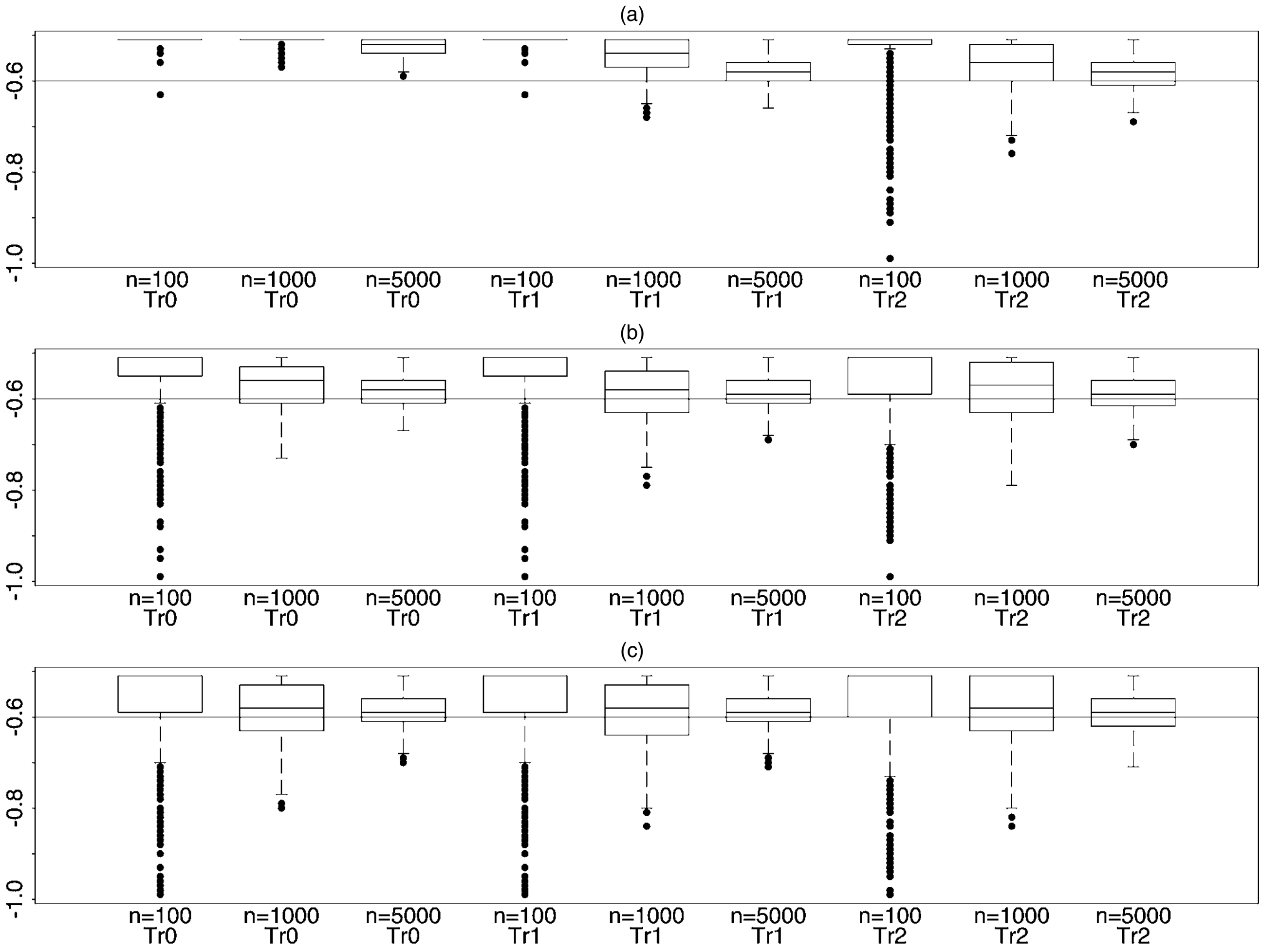

Figure 1. Estimates of d_0, Quartic Mean Function: (a) $p = 0$; (b) $p = 1$; (c) $p = 2$.

difficulties are due to the noninvertibility of the differences ($d_0 < -.5$), combined with the failure to use a taper. There is no contradiction with Theorem 2, because the theorem assumes that $p \geq 1$. Furthermore, for $p = 0$ and $n = 100$, the estimator was often at the boundary of $-.51$, thereby reducing the observed variance. For $p = 1$, $\ell = n^{.25}$ and $p = 2$, $\ell = n^{.15}$, by the time we reach $n = 5{,}000$, the simulation variances agree well with the corresponding finite-sample corrected variances, but differ substantially from the asymptotic variances.

Although our theoretical results require both tapering and trimming, we do not know whether either of these is required to ensure the $\sqrt{m}$-consistency and asymptotic normality of $\hat{d}_n$, based on differenced data. Nevertheless, the simulation results shown here indicate that both tapering and trimming can be helpful in reducing the finite-sample bias and RMSE of $\hat{d}_n$. Next, we explore empirically the effects of differencing.

4.2 Comparison With a Direct Estimator of the Memory Parameter

Robinson (1997) proved the $\log^{1/2}(n)$-consistency of a Gaussian semiparametric estimator based directly on the original series of levels $\{X_t\}$ given by (1). This rate was sufficient for the purpose at hand and may be improvable. We discuss this point further at the end of this section. Our focus here is the performance of the direct estimator in practice. For notational convenience, we use the superscript "$'$" to denote the addition of 1 to a value of d. The direct estimator $\hat{d}'^{\mathrm{Dir}}$ of $d'_0 \in (0, 1/2)$, which uses neither differencing nor tapering, nor a preliminary estimate of the mean function, can be obtained by minimizing with respect to d' a modification of $\hat{J}_{1,m}(d')$ in (9), in which $I_{Y,k}$ is replaced by $I_{X,k}$. Note that $d'_0 = d_0 + 1$ is the memory parameter of $\{\epsilon_t\}$ in (1). In our simulation study, we constructed values of $\hat{d}'^{\mathrm{Dir}}$ using the same simulated realizations of $\{X_t\}$ of length $n + 1$ used in the previous section. Whereas we differenced and tapered the simulated realizations of $\{X_t\}$ in computing our estimator $\hat{d}_n$, we now use $\{X_t\}$ directly for computing $\hat{d}'^{\mathrm{Dir}}$. In our study, we took $\hat{d}'^{\mathrm{Dir}}$ to be the minimizer of the profile likelihood for $d' \in [.01, .49]$. We also considered trimmed versions of the direct estimator, $\hat{d}'^{\mathrm{Dir}}_{\mathrm{Tr1}}$ and $\hat{d}'^{\mathrm{Dir}}_{\mathrm{Tr2}}$, with lower trimming $\ell = n^{.15}$ and $\ell = n^{.25}$. Table 4 gives the bias and RMSE for $\hat{d}'^{\mathrm{Dir}}$, $\hat{d}'^{\mathrm{Dir}}_{\mathrm{Tr1}}$, and $\hat{d}'^{\mathrm{Dir}}_{\mathrm{Tr2}}$. Both fixed and adaptive bandwidths were used, as in the study of $\hat{d}_n$.

We focus first on the case of the quartic mean function. Here $\hat{d}'^{\mathrm{Dir}}$ and its trimmed versions are severely positively biased in almost all cases, with the overwhelming majority of all realizations leading to an estimate on the boundary, .49. Trimming does not improve the bias of the direct estimator, in contrast to what we found earlier for $\hat{d}_n$ with $p = 0$. A possible explanation for this is that for the quartic mean function, the low-frequency contribution to the DFT of the mean is smaller if differences are

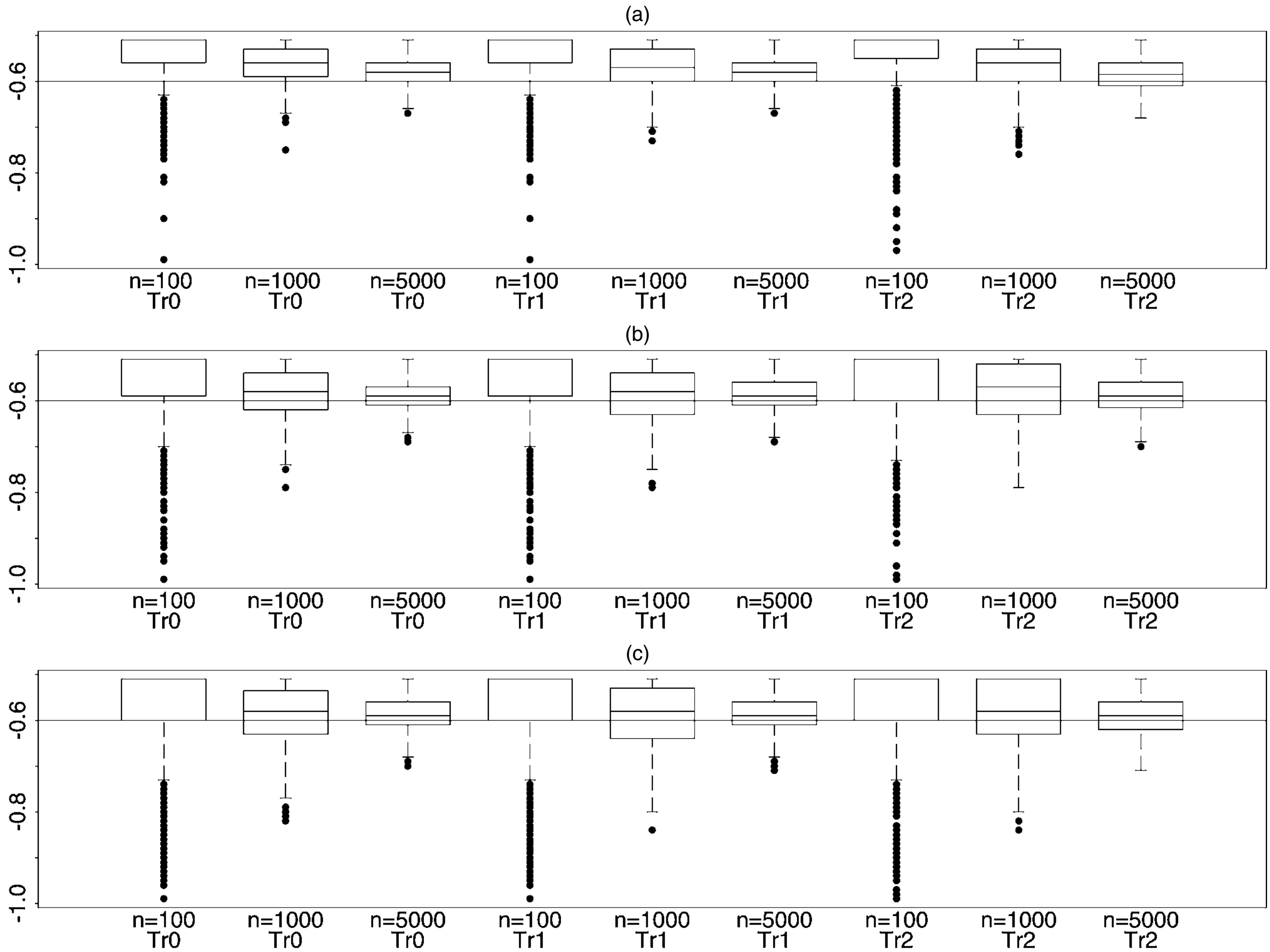

Figure 2. Estimates of d_0, Constant Mean Function: (a) $p = 0$; (b) $p = 1$; (c) $p = 2$.

used than if the nondifferenced data are used. Presumably (cf. Prop. 1), an even stronger degree of trimming would improve the bias of the direct estimator, at the expense of a considerable inflation in variance. The direct estimators often had a higher RMSE than the corresponding estimators $\hat{d}_n$, although in some cases the direct estimator had lower RMSE due to its extremely small variance, which was in turn due to the fact that most of the estimated values were on the boundary, .49.

For the constant mean function, the direct estimators performed better in terms of RMSE than the corresponding $\hat{d}_n$. This is not surprising, because the variance of $\hat{d}_n$ is inflated by noninvertibility if $p = 0$ and by the taper if $p = 1$ or $p = 2$.

Table 3. Asymptotic, Finite-Sample Corrected, and Simulation Variances for $\hat{d}_n$

n		$p = 0$, $\ell = 1$	$p = 1$, $\ell = n^{.25}$	$p = 2$, $\ell = n^{.15}$
100	Asymptotic: $\Phi_p/(4m)$	.0100	.0150	.0194
	Corrected	.0150	.0525	.0413
	Simulation	.004471	.011634	.011547
1,000	Asymptotic: $\Phi_p/(4m)$	.00200	.00300	.00389
	Corrected	.00229	.00562	.00574
	Simulation	.002073	.004309	.004487
5,000	Asymptotic: $\Phi_p/(4m)$	.000644	.000966	.00125
	Corrected	.000685	.00145	.00158
	Simulation	.001012	.001474	.001488

Table 4. Bias (RMSE) for $\hat{d}'^{Dir}$

n	ℓ	Quartic mean	Constant mean
$m = n^{.7}$			
100	1	$.090_{(.090)}$	$.025_{(.096)}$
	$n^{.15}$	$.090_{(.090)}$	$.025_{(.096)}$
	$n^{.25}$	$.089_{(.089)}$	$.028_{(.103)}$
1,000	1	$.090_{(.090)}$	$.016_{(.048)}$
	$n^{.15}$	$.090_{(.090)}$	$.016_{(.050)}$
	$n^{.25}$	$.089_{(.090)}$	$.017_{(.058)}$
5,000	1	$.090_{(.090)}$	$.008_{(.029)}$
	$n^{.15}$	$.090_{(.090)}$	$.009_{(.030)}$
	$n^{.25}$	$.088_{(.088)}$	$.010_{(.034)}$
Adaptive m			
100	1	$.089_{(.090)}$	$-.065_{(.186)}$
	$n^{.15}$	$.089_{(.090)}$	$-.065_{(.186)}$
	$n^{.25}$	$.022_{(.151)}$	$-.102_{(.240)}$
1,000	1	$.090_{(.090)}$	$-.003_{(.080)}$
	$n^{.15}$	$.090_{(.090)}$	$-.007_{(.087)}$
	$n^{.25}$	$.089_{(.089)}$	$-.024_{(.122)}$
5,000	1	$.090_{(.090)}$	$.005_{(.041)}$
	$n^{.15}$	$.090_{(.090)}$	$.004_{(.046)}$
	$n^{.25}$	$.089_{(.089)}$	$.004_{(.055)}$

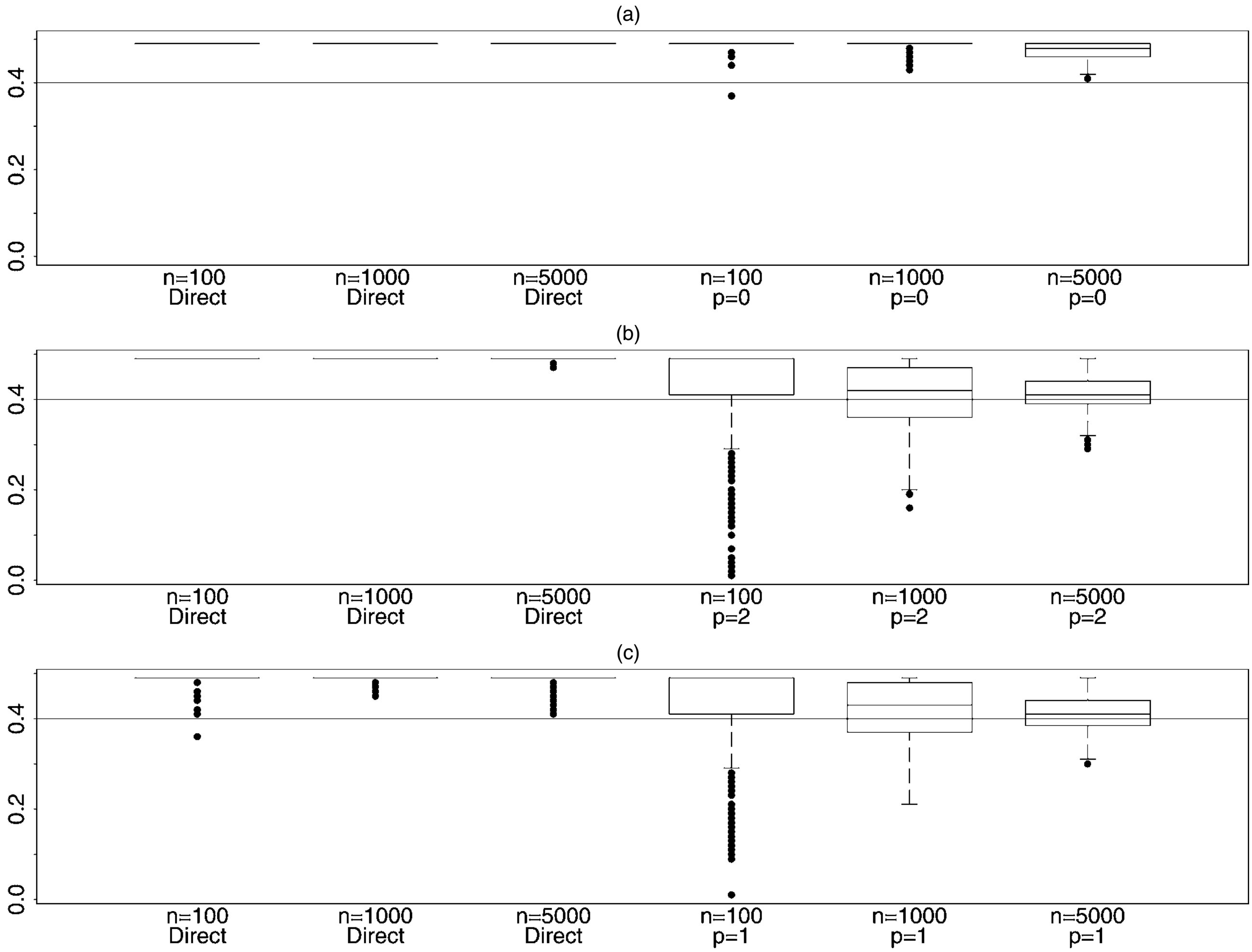

Figure 3. Estimates of $d_0 + 1$, Quartic Mean Function: (a) No Trimming; (b) Trimming Tr1; (c) Trimming Tr2.

Figures 3 and 4, for the quartic and constant mean functions, present boxplots of the direct estimates (based on levels) and our proposed estimates (based on differences) of d_0' for the three degrees of trimming and fixed bandwidth. Our proposed estimators are given by $\hat{d}_n + 1$, and we used $p = 0$, $p = 2$, and $p = 1$ for the three degrees of trimming Tr0, Tr1, and Tr2. The results are consistent with the foregoing discussion. The effect of using the adaptive choice of m was to often inflate the RMSE, as was found earlier for $\hat{d}_n$.

Overall, our study suggests that $\hat{d}_n$ is preferred to the direct estimator in practice, because for many mean functions there will be a strong bias reduction in $\hat{d}_n$ resulting from the combination of differencing, tapering, and trimming that is sufficient to offset any variance inflation. Furthermore, in theory, although it may be possible to establish the same rate of convergence for the direct estimator without trimming that we have shown in Theorem 2 for $\hat{d}_n$ with tapering, trimming, and differencing, we have been able to obtain this rate only for the direct estimator with a very large degree of trimming, in Proposition 1. This clearly would produce a very strong inflation in the variance, because far fewer frequencies would be used in the estimator. For example, when $n = 5{,}000$, we have $n^{.6} = 165$ and $n^{.7} = 388$, so even a lower trimming of $\ell = n^{.6}$, which is still not sufficient for Proposition 1 to apply, would require us to trim 42% of the frequencies used in the nontrimmed estimator.

4.3 Feasible Inference for the Regression Function

Robinson (1997) derived the asymptotic properties of the nonparametric regression estimator of $r(z)$ given by

$$\hat{r}(z) = \frac{1}{nb}\sum_{t=0}^{n} K\left(\frac{nz - t}{nb}\right) X_t$$

for $z \in (0, 1)$, where $\{X_t\}$ is given by (1), b is a bandwidth, and K is a suitable kernel function. We use the same kernel used in the simulations of Robinson (1997),

$$K(v) = (1/2)\{1 + \cos(\pi v)\}, \qquad |v| \le 1.$$

Robinson (1997) assumed that $r(z) \in \mathrm{Lip}(\tau)$, $0 < \tau \le 1$, or $r'(z) \in \mathrm{Lip}(\tau - 1)$, $1 < \tau \le 2$, and showed that if $(nb)^{-1} + n^{1/2-d_0'} b^{1/2-d_0'+\tau} \to 0$, where $d_0' = d_0 + 1 \in (0, 1/2)$, then for any $z \in (0, 1)$,

$$(nb)^{1/2-d_0'}[\hat{r}(z) - r(z)] \xrightarrow{d} \mathrm{N}\big(0, G\rho(d_0')\big), \tag{12}$$

where $G = f^*(0)$,

$$\rho(d_0') = \theta(d_0') \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(v)K(w)|v - w|^{2d_0'-1}\, \mathrm{d}v\, \mathrm{d}w, \tag{13}$$

and

$$\theta(d_0') = 2\Gamma(1 - 2d_0')\cos\{\pi(1/2 - d_0')\}. \tag{14}$$

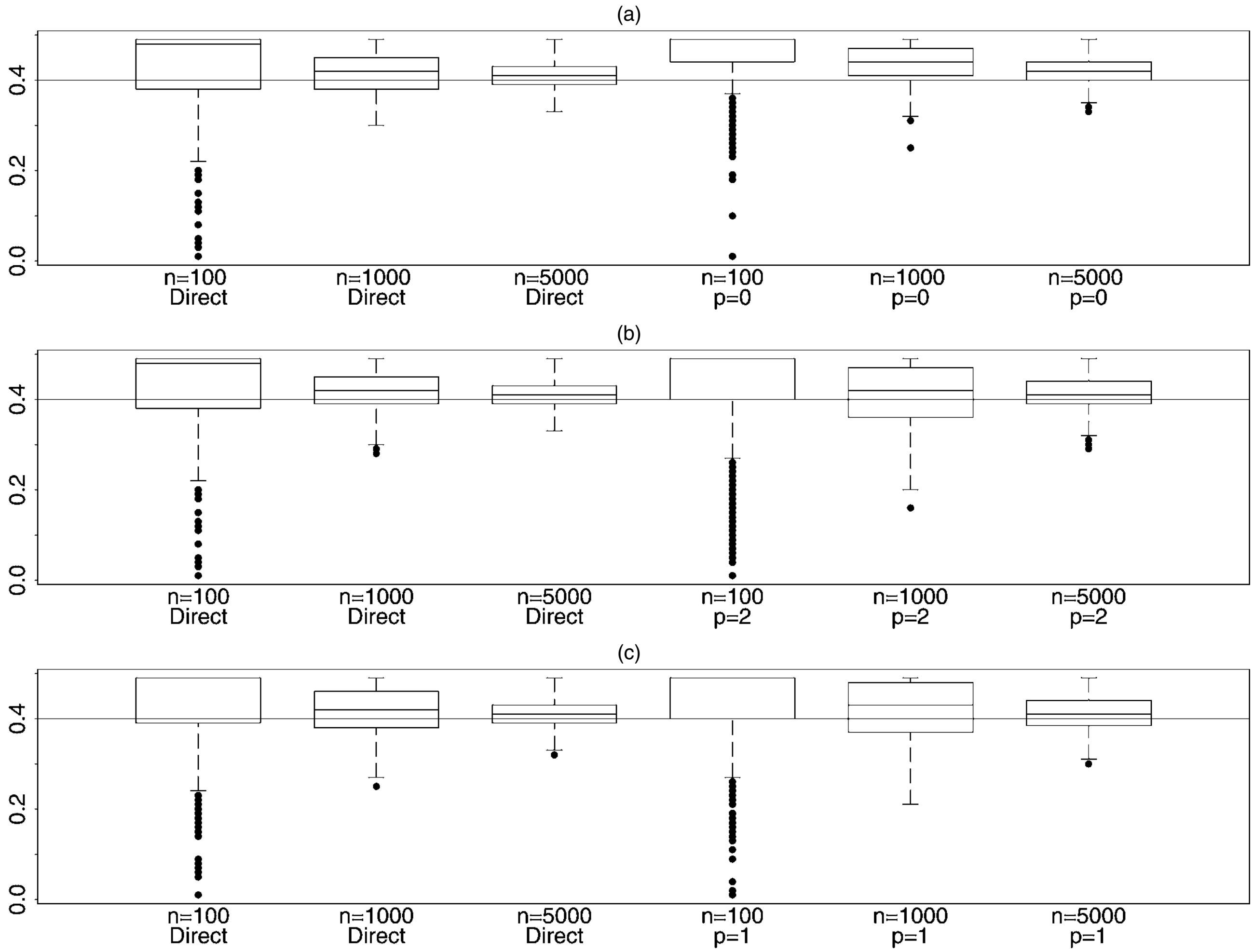

Figure 4. Estimates of $d_0 + 1$, Constant Mean Function: (a) No Trimming; (b) Trimming Tr1; (c) Trimming Tr2.

Note that the asymptotic variance in (12) depends on the unknown parameters G and d_0', but Robinson (1997) showed that if estimators $\hat{G}$ and $\hat{d}'$ are available such that $\hat{G} \xrightarrow{P} G$ and $(\log nb)(\hat{d}' - d_0') \xrightarrow{P} 0$, then

$$\{\hat{G}\rho(\hat{d}')\}^{-1/2}(nb)^{1/2-\hat{d}'}\{\hat{r}(z) - r(z)\} \xrightarrow{d} \mathrm{N}(0,1). \qquad (15)$$

Under some additional conditions, Robinson (1997, thm. 4) showed that $\hat{G}^{\mathrm{Dir}} \xrightarrow{P} G$ and $(\log nb)(\hat{d}'^{\mathrm{Dir}} - d_0') \xrightarrow{P} 0$, where

$$\hat{G}^{\mathrm{Dir}} = \frac{1}{m}\sum_{k=1}^{m} x_k^{2\hat{d}'^{\mathrm{Dir}}} I_{X,k}.$$

In this section we extend our simulation study to problems on feasible inference for r. We consider estimation of G, then ρ, and finally the properties of confidence intervals for r. For brevity, we report results only for the fixed bandwidth, $m = n^{.7}$, in the remainder of this section.

For estimation of G, we consider $\hat{G}^{\mathrm{Dir}}$ defined earlier, which is based directly on the levels $\{X_t\}$, as well as several estimators based on the differences, $\{Y_t\}$, with possible tapering. These estimators are given by

$$\hat{G}_{n,p} = \frac{1}{m - \ell + 1}\sum_{k=\ell}^{m}\{(k + p/2)/n\}^{2\hat{d}_{n,p}} I_{Y,k},$$

where $\hat{d}_{n,p}$ is our proposed estimator of the memory parameter d_0 of the differences, and for the remainder of this section we use $\ell = 1$, $n^{.25}$, and $n^{.15}$ for $p = 0, 1, 2$. We also considered trimmed direct estimators $\hat{G}_{\mathrm{Tr1}}^{\mathrm{Dir}}$ and $\hat{G}_{\mathrm{Tr2}}^{\mathrm{Dir}}$ given by

$$\hat{G}_{\mathrm{Tr1}}^{\mathrm{Dir}} = \frac{1}{m - n^{.15} + 1}\sum_{k=n^{.15}}^{m} x_k^{2\hat{d}_{\mathrm{Tr1}}'^{\mathrm{Dir}}} I_{X,k}$$

and

$$\hat{G}_{\mathrm{Tr2}}^{\mathrm{Dir}} = \frac{1}{m - n^{.25} + 1}\sum_{k=n^{.25}}^{m} x_k^{2\hat{d}_{\mathrm{Tr2}}'^{\mathrm{Dir}}} I_{X,k}.$$

Because the AR(1) parameter in the simulations is .2, we have $G = f^*(0) = 1/[2\pi(1 - .2)^2] = .2487$. Boxplots of the various estimators of G are given in Figures 5 and 6, and the statistical summaries are given in Tables 5 and 6. For the quartic mean function the estimates based on $\hat{d}_n$ are generally the least biased, especially as p increases, and trimming sometimes improves the bias of the direct methods. For the zero mean, the direct estimators and the proposed estimator with $p = 0$ are less biased than in the case of the quartic mean. For both mean functions, the variance of the estimators based on $\hat{d}_n$ with $p = 1$ or $p = 2$ is larger than that of the direct methods.

Next, we consider estimation of ρ. The true value of ρ in the situation under study is $\rho = 11.4973$, obtained by evaluating

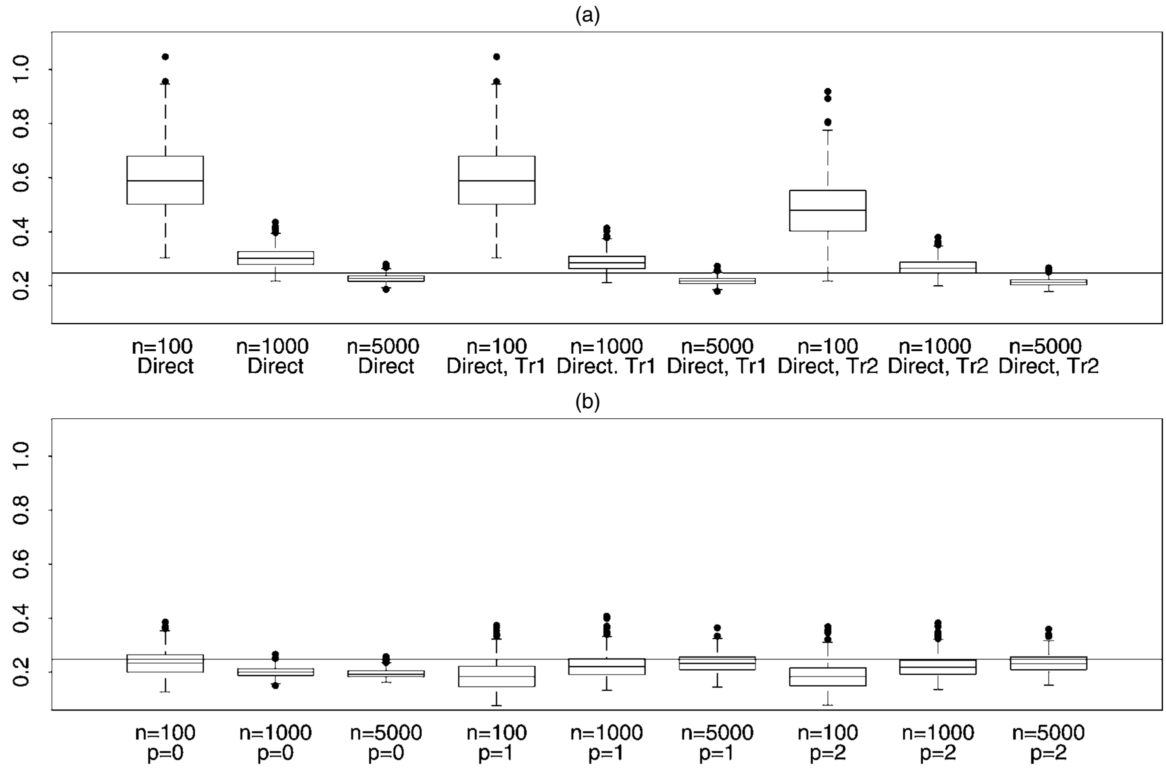

Figure 5. Estimates of $G = .2487$, Quartic Mean Function: (a) Direct Estimates; (b) Tapered Estimates.

$\rho(d_0')$ in (13) using numerical integration. Estimates of ρ were obtained by plugging estimated values of d_0' into (13) and using numerical integration. The means and variances of the estimates of ρ are given in Tables 7 and 8, and the corresponding boxplots are given in Figures 7 and 8 for the quartic and zero mean functions. For the quartic mean, estimates based on $\hat{d}_n$ are far

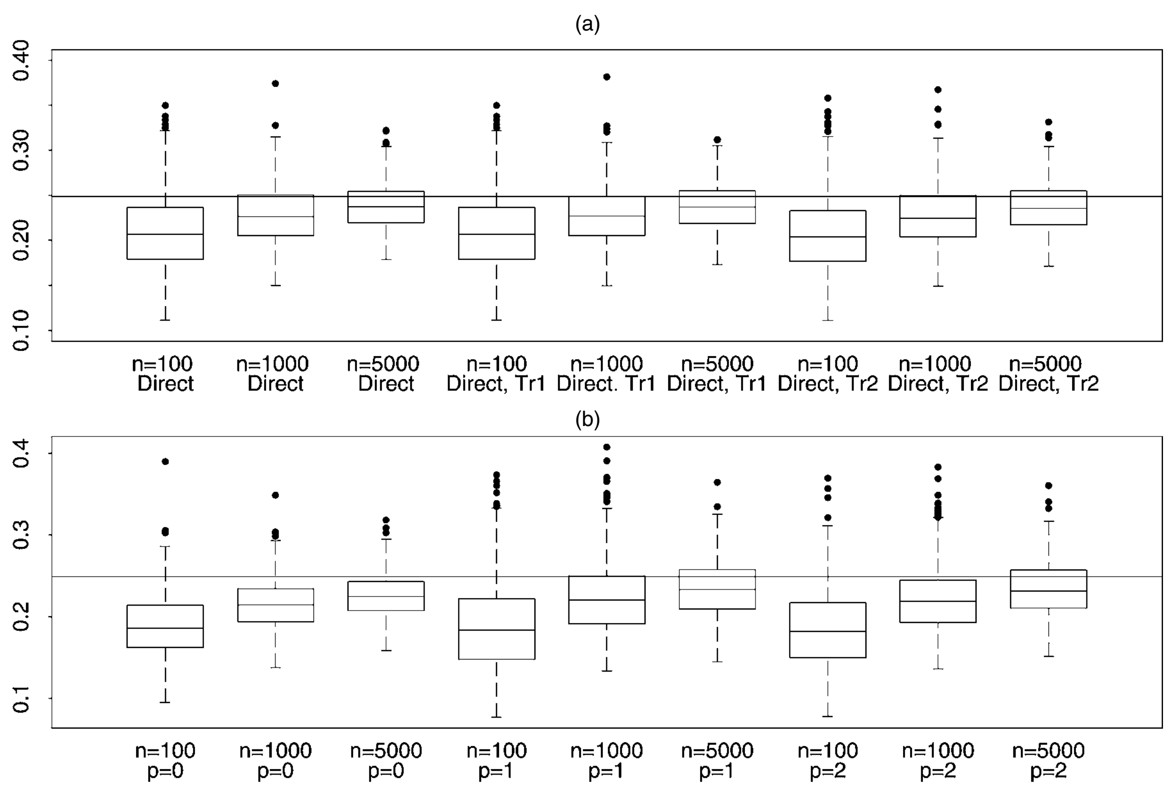

Figure 6. Estimates of $G = .2487$, Constant Mean Function: (a) Direct Estimates; (b) Tapered Estimates.

Table 5. Simulation Results for Estimates of $G = .2487$, Quartic Mean Function

					p		
n		Direct	Direct, Tr1	Direct, Tr2	0	1	2
100	Mean	.5938	.5938	.4819	.2353	.1865	.1880
	Variance	.01659	.01659	.01270	.00221	.00232	.00290
1,000	Mean	.3032	.2872	.2676	.2009	.2232	.2257
	Variance	.001215	.001123	.000951	.000349	.001661	.002102
5,000	Mean	.2262	.2176	.2127	.1950	.2332	.2340
	Variance	.000199	.000185	.000177	.000251	.001116	.001228

less biased but have substantially higher variance than estimates based on the direct estimator, because the direct estimator often has very small or zero variance, and the bias of the proposed estimator improves as p increases. For the zero mean, estimates based on $\hat{d}_n$ again have a larger variance than estimates based on the direct estimator, which is again not surprising, because $\hat{d}_n$ has a larger variance than the direct estimator in this case as well.

Finally, we present simulation results on confidence intervals for $r(1/2)$. All of the confidence intervals are based on (15), and $\hat{r}(1/2)$ was computed using a bandwidth of $b = n^{-.1}$. This is an arbitrary choice, within the constraints for asymptotic normality and unbiasedness described before (12). Table 9 gives the error rates (noncoverage rates) of nominal 95% and 99% confidence intervals based on our $\hat{d}'$, with $p = 0, 1, 2$, and corresponding trimming numbers $\ell = 1, n^{.25}$ and $n^{.15}$. These intervals for confidence level $1 - \alpha$ are of the form

$$\hat{r}(1/2) \pm z_{\alpha/2}\{\hat{G}\rho(\hat{d}')\}^{1/2}(nb)^{\hat{d}'-1/2}.$$

We also present results on intervals based on the direct methods of estimation of the forms

$$\hat{r}(1/2) \pm z_{\alpha/2}\left\{\hat{G}^{\mathrm{Dir}}\rho(\hat{d}'^{\mathrm{Dir}})\right\}^{1/2}(nb)^{\hat{d}'^{\mathrm{Dir}}-1/2},$$

$$\hat{r}(1/2) \pm z_{\alpha/2}\left\{\hat{G}^{\mathrm{Dir}}_{\mathrm{Tr1}}\rho(\hat{d}'^{\mathrm{Dir}}_{\mathrm{Tr1}})\right\}^{1/2}(nb)^{\hat{d}'^{\mathrm{Dir}}_{\mathrm{Tr1}}-1/2},$$

and

$$\hat{r}(1/2) \pm z_{\alpha/2}\left\{\hat{G}^{\mathrm{Dir}}_{\mathrm{Tr2}}\rho(\hat{d}'^{\mathrm{Dir}}_{\mathrm{Tr2}})\right\}^{1/2}(nb)^{\hat{d}'^{\mathrm{Dir}}_{\mathrm{Tr2}}-1/2}.$$

These are denoted in Table 9 by "Direct," "Direct Tr1," and "Direct Tr2." For all of the feasible confidence intervals, the estimate of d'_0 is based on the fixed bandwidth, $m = n^{.7}$. For the sake of calibration, we also include results in Table 10 for an infeasible method in which the true asymptotic standard error is used, yielding the interval

$$\hat{r}(1/2) \pm z_{\alpha/2}\{G\rho(d'_0)\}^{1/2}(nb)^{d'_0-1/2}.$$

Table 10 shows that the noncoverage rate of the infeasible confidence intervals is somewhat higher than the nominal rate for the quartic mean, but somewhat closer to the nominal rate for the zero mean. In principle, one should not expect the feasible intervals to do better than the infeasible ones. For the zero mean (see Table 9), the noncoverage rate of all of the feasible methods is typically higher than the nominal rate, but it improves as the sample size increases. For the quartic mean, the direct methods and our method with $p = 0$ almost always yielded a zero noncoverage rate, whereas our method with $p = 1$ and $p = 2$ yielded noncoverage rates that are somewhat higher than for the zero mean but become closer to the nominal rate as the sample size increases. The zero noncoverage rate described earlier is presumably due to overly wide confidence intervals as a result of overestimation of d'_0.

5. DATA ANALYSIS

We analyzed a dataset of seasonally adjusted monthly temperatures (in °C) for the northern hemisphere for the years 1854–1989 ($n = 1{,}632$). The data were given by Beran (1994, pp. 257–261). Figure 9 gives a time series plot of the data, together with the estimated regression function, using the same kernel as used by Robinson (1997), with two different arbitrarily chosen bandwidths, $b = n^{-.4}$ (the rougher curve) and $b = n^{-.2}$ (the smoother curve). From this plot, the evidence for global warming appears mixed, in that the estimated mean function is higher in the second half of the data than in the first half, although the function is relatively flat, and even decreasing, for certain reasonably long periods. Table 11 gives direct and tapered estimates of d'_0, G, and ρ, for both fixed bandwidth ($m = n^{.7}$) and adaptively chosen bandwidth, just as was done in our simulations. The estimates of d'_0 are much higher for the adaptive bandwidth than for the fixed bandwidth. This leads to corresponding large differences in the estimates of G and ρ.

Next, we constructed t-tests of the null hypothesis $r(z_2) - r(z_1) = 0$ for two particular arbitrary choices of z_1 and z_2,

Table 6. Simulation Results for Estimates of $G = .2487$, Constant Mean Function

					p		
n		Direct	Direct, Tr1	Direct, Tr2	0	1	2
100	Mean	.2085	.2085	.2071	.1898	.1865	.1877
	Variance	.001870	.001870	.001916	.001572	.002328	.002894
1,000	Mean	.2282	.2279	.2276	.2158	.2232	.2257
	Variance	.001012	.001081	.001161	.000842	.001663	.002096
5,000	Mean	.2382	.2379	.2373	.2266	.2332	.2341
	Variance	.000653	.000690	.000801	.000685	.001118	.001228

Table 7. Simulation Results for Estimates of $\rho = 11.4973$, Quartic Mean Function

					p		
n		*Direct*	*Direct, Tr1*	*Direct, Tr2*	*0*	*1*	*2*
100	Mean	100.80	100.80	99.130	100.041	70.932	67.082
	Variance	0	0	128.73	58.649	1,818.0	1,875.2
1,000	Mean	100.80	100.80	99.171	93.751	33.975	31.081
	Variance	0	0	103.28	412.65	1,285.6	1,160.2
5,000	Mean	100.80	100.57	93.592	60.629	18.596	18.073
	Variance	0	13.689	407.62	1,188.9	360.54	331.55

Table 8. Simulation Results for Estimates of $\rho = 11.4973$, Constant Mean Function

					p		
n		*Direct*	*Direct, Tr1*	*Direct, Tr2*	*0*	*1*	*2*
100	Mean	55.349	55.349	62.036	69.544	70.573	66.371
	Variance	1,902.8	1,902.8	1,919.1	1,691.2	1,820.9	1,882.0
1,000	Mean	22.311	23.606	28.266	34.970	34.011	31.227
	Variance	575.65	642.99	943.02	1,162.1	1,293.9	1,169.9
5,000	Mean	13.713	14.091	14.896	19.676	18.594	18.069
	Variance	35.777	53.226	73.952	319.23	360.56	331.58

namely (January 1864, January 1980) and (September 1900, January 1975). Because Robinson (1997, lem. 3) demonstrated the asymptotic independence of $\hat{r}(z_1)$ and $\hat{r}(z_2)$ for $z_1 \neq z_2$, we obtain the t-statistic

$$t = \frac{\hat{r}(z_2) - \hat{r}(z_1)}{\sqrt{2}\{\hat{G}\rho(\hat{d}')\}^{1/2}(nb)^{\hat{d}'-1/2}},$$

which is asymptotically standard normal under the null hypothesis. The various t-statistics are given in the remainder of Table 11, using a subscript to denote the choice of (z_1, z_2) and a superscript to denote the bandwidth b. It is seen that the t-statistics are most sensitive to the bandwidth m used in estimating d_0' (fixed vs. adaptive) and relatively insensitive to all of the other choices. For fixed m, many of the t-statistics provide strong evidence for global warming, whereas for adaptively chosen m, none of the t-statistics is greater than 1. This pattern can be explained by the fact that the estimated value

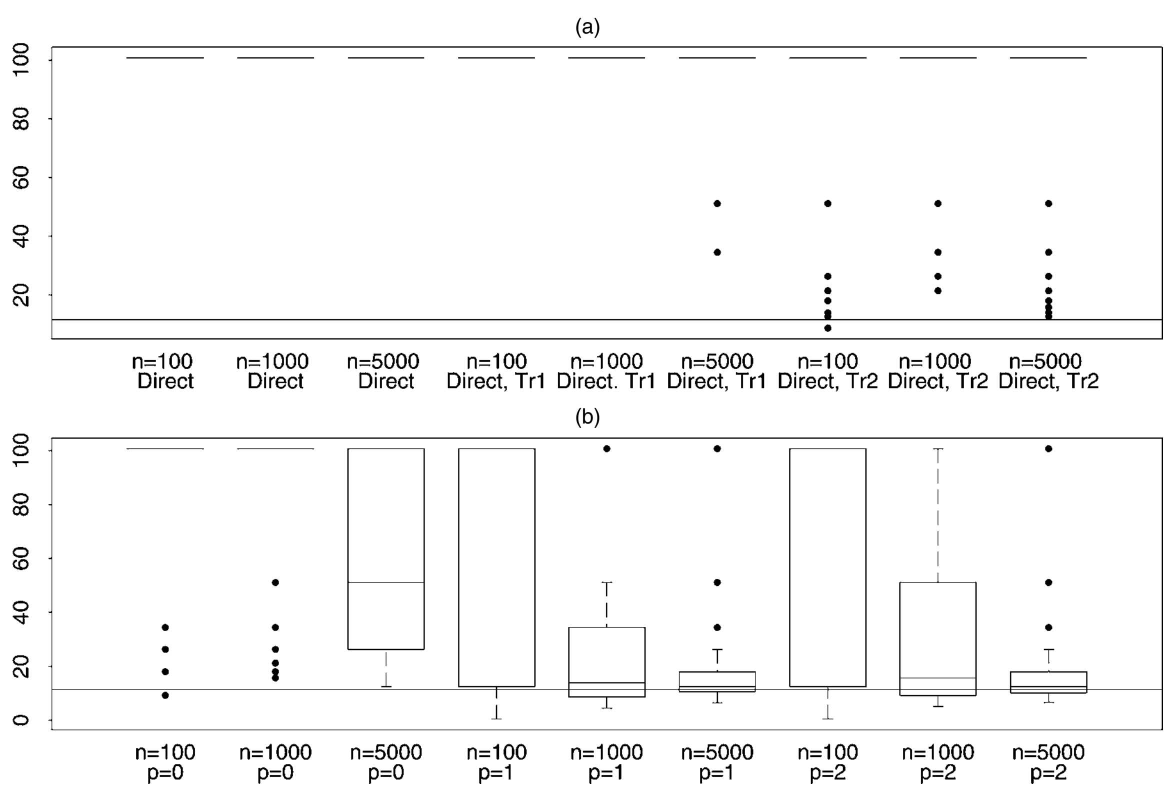

Figure 7. Estimates of $\rho = 11.4973$, Quartic Mean Function: (a) Direct Estimates; (b) Tapered Estimates.

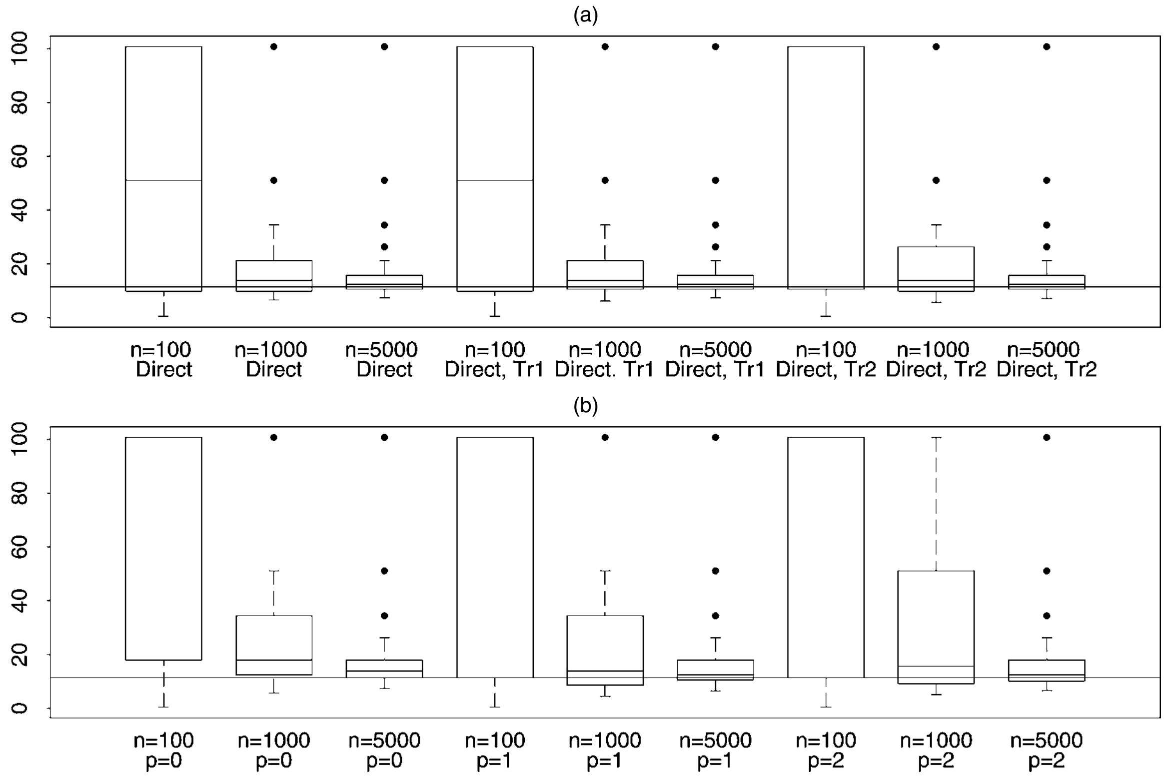

Figure 8. Estimates of $\rho = 11.4973$, Constant Mean Function: (a) Direct Estimates; (b) Tapered Estimates.

Table 9. Noncoverage Rates for Feasible Confidence Intervals for $r(1/2)$, Nominal Level $1-\alpha$

n		$\alpha = .05$, quartic mean	$\alpha = .01$, quartic mean	$\alpha = .05$, mean 0	$\alpha = .01$, mean 0
100	Direct	0	0	.074	.050
	Direct, Tr1	0	0	.074	.050
	Direct, Tr2	0	0	.078	.052
	$p=0$	0	0	.034	.016
	$p=1$	.110	.092	.092	.066
	$p=2$	.118	.098	.100	.074
1,000	Direct	0	0	.062	.030
	Direct, Tr1	0	0	.072	.030
	Direct, Tr2	0	0	.074	.038
	$p=0$	0	0	.044	.018
	$p=1$	.104	.066	.086	.048
	$p=2$	.128	.066	.098	.046
5,000	Direct	0	0	.060	.022
	Direct, Tr1	0	0	.062	.022
	Direct, Tr2	0	0	.064	.028
	$p=0$	.002	0	.032	.012
	$p=1$	.078	.030	.060	.028
	$p=2$	.074	.030	.062	.026

Table 10. Noncoverage Rates for Infeasible Confidence Intervals for $r(1/2)$, Nominal Level $1-\alpha$

n	$\alpha = .05$, quartic mean	$\alpha = .01$, quartic mean	$\alpha = .05$, mean 0	$\alpha = .01$, mean 0
100	.086	.018	.034	.010
1,000	.090	.016	.040	.010
5,000	.068	.018	.070	.016

of d_0' was always substantially larger for adaptively chosen m than for fixed m.

APPENDIX: PROOFS

The proofs of Theorems 1 and 2 are based on the decomposition of the DFT ordinates of the signal as the sum of the DFT ordinates of the unobserved process η and of the differenced mean function Δr. In the first section of this appendix, we state the results that we need for the DFT ordinates of the stationary but noninvertible process η. We present these results in the modern framework of empirical processes, because this is the standard way to prove asymptotic results for M-estimators. The proofs are more or less straightforward

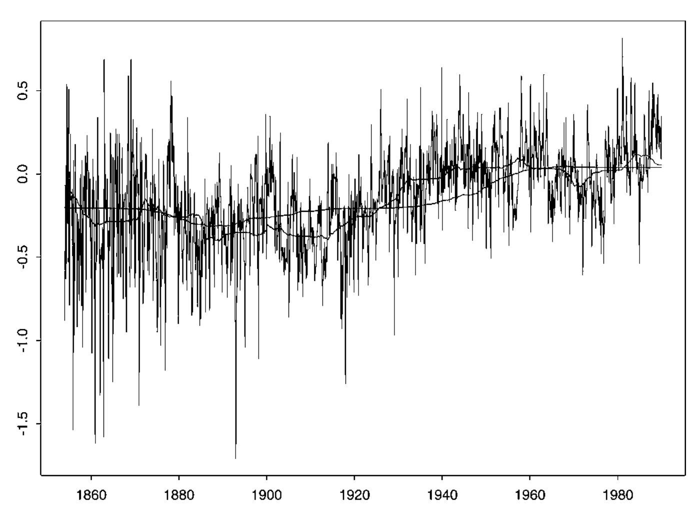

Figure 9. Seasonally Adjusted Monthly Temperatures (°C) and Estimated Means, Northern Hemisphere, 1854–1989.

Table 11. Analysis of Global Temperatures, 1854–1989

	Direct	Direct, Tr1	Direct, Tr2	p = 0	p = 1	p = 2
$m = n^{.7}$						
$\hat{d}'$	.33	.26	.23	.38	.28	.30
$\hat{G}$	.0121	.0140	.0149	.0109	.0097	.0085
$\hat{\rho}$	7.52	5.88	5.43	9.88	6.24	6.69
$t^{n^{-.4}}_{1980-1864}$	1.55	2.21	2.56	1.14	2.37	2.24
$t^{n^{-.2}}_{1980-1864}$	1.57	2.49	3.00	1.07	2.58	2.37
$t^{n^{-.4}}_{1975-1900}$	1.63	2.33	2.69	1.20	2.49	2.36
$t^{n^{-.2}}_{1975-1900}$	1.94	3.07	3.71	1.32	3.19	2.93
Adaptive m						
$\hat{d}'$	.48	.47	.46	.49	.45	.49
$\hat{G}$	.0050	.0050	.0053	.0053	.0047	.0036
$\hat{\rho}$	51.1	34.6	26.3	100.8	21.3	100.8
$t^{n^{-.4}}_{1980-1864}$	.48	.60	.70	.31	.86	.38
$t^{n^{-.2}}_{1980-1864}$	.39	.50	.59	.25	.73	.30
$t^{n^{-.4}}_{1975-1900}$	.50	.63	.74	.33	.91	.40
$t^{n^{-.2}}_{1975-1900}$	.48	.61	.73	.31	.90	.37

adaptations of existing proofs for similar results (cf., e.g., Robinson 1995b; Velasco 1999b; Hurvich and Chen 2000). Because our assumptions differ from those of these authors, we provide a streamlined but self-contained proof at the end of the Appendix. Note that even though the process η has a summable covariance function, we cannot use classical results on weakly dependent time series, such as those of Dahlhaus (1988), because such results usually preclude that the spectral density vanishes at 0, which is the case in the present context.

In the following sections, we provide a full treatment of the terms that involve the mean function r.

A.1 Preliminary Results

For $\ell \le k \le m$, denote $u_k = f^*(0)^{-1/2} x_k^{d_0} d_{\eta,k}$, $v_k = \sqrt{2\pi}\, d_{Z,k}$, and $\zeta_k = x_k^{2d_0} f^*(0)^{-1} I_{\eta,k} - 1 = |u_k|^2 - 1$, and for $\mathbf{c} \in \mathbb{R}^{m-\ell+1}$, define $\mathcal{Z}_m(\mathbf{c}) = \sum_{k=\ell}^m c_k \zeta_k$.

Theorem A.1. Assume (A1), (A2), and (A4).

1. Denote $p_{j,k} = (j \vee k)^{-1} + ((j \vee k)/n)^2$. There exists a constant C such that for all $k \le \vartheta n/\pi$,
$$\mathbb{E}[|u_k - v_k|^2] \le C p_{k,k}, \tag{A.1}$$
$$\big|\mathbb{E}[\bar{v}_k(u_j - v_j)]\big| + \big|\mathbb{E}[v_k(u_j - v_j)]\big| \le C p_{j,k}, \tag{A.2}$$
and
$$|\mathrm{cum}(\bar{v}_k, u_k - v_k, v_j, \bar{u}_j - \bar{v}_j)| \le C n^{-1/2} \sqrt{p_{j,j} p_{k,k}}. \tag{A.3}$$
2. Let ℓ and m satisfy $\ell/m \to 0$ and $m/n \to 0$. For $\epsilon \in (0, 1]$ and $K > 0$, let $\mathcal{C}_m(\epsilon, K)$ be the subset of vectors $\mathbf{c} \in \mathbb{R}^{m-\ell+1}$ such that
$$|c_k - c_{k+1}| \le K m^{-\epsilon} k^{\epsilon-2}, \qquad |c_m| \le K m^{-1}. \tag{A.4}$$
Then $\sup_{\mathbf{c} \in \mathcal{C}_m(\epsilon,K)} \mathcal{Z}_m(\mathbf{c}) = O_P(m^{-1/2} + m/n)$.
3. Let ℓ and m satisfy $\ell/m + m/n \to 0$. Let $\mathbf{c}_m$ be a sequence of vectors of $\mathbb{R}^{m-\ell+1}$ that satisfies
$$\sum_{k=1}^m c_{m,k} = 0, \qquad \sum_{k=1}^m c_{m,k}^2 = 1, \tag{A.5}$$
and
$$\lim_{n\to\infty} \left(|c_{m,m}| + \sum_{k=1}^{m-1} |c_{m,k+1} - c_{m,k}| \right)^2 \log(n) = 0. \tag{A.6}$$
Then $\mathcal{Z}_m(\mathbf{c}_m)$ converges weakly to the Gaussian law with mean-0 and variance
$$\sigma^2(h) = \left(\sum_{u=1}^{v} b_u^2 \right)^{-2} \sum_{u=-v}^{v} \left(\sum_{t=1}^{v-|u|} b_u b_{u+t} \right)^2.$$

A.2 Proof of Theorem 1

Define
$$J_{\ell,m}(d) = \log\left(\frac{1}{m-\ell+1} \sum_{k=\ell}^m k^{2d-2d_0} \right) - 2(d-d_0)\gamma_{\ell,m}$$
and
$$E_n(d) = \left(\sum_{j=\ell}^m j^{2d-2d_0} \right)^{-1} \sum_{k=\ell}^m k^{2d-2d_0} \big(x_k^{2d_0} f^*(0)^{-1} I_{Y,k} - 1 \big).$$
With these notations, we obtain $J_{\ell,m}(d_0) = 0$,
$$\begin{aligned}\hat{J}_{\ell,m}(d) &= \log(1 + E_n(d)) + J_{\ell,m}(d) \\ &\quad + \log(f^*(0)) - 2d_0 \log(2\pi/n) - 2d_0 \gamma_{\ell,m}, \end{aligned}\tag{A.7}$$
and
$$\begin{aligned}&\hat{J}_{\ell,m}(d) - \hat{J}_{\ell,m}(d_0) \\ &\quad = \log(1 + E_n(d)) - \log(1 + E_n(d_0)) + J_{\ell,m}(d) - J_{\ell,m}(d_0). \end{aligned}\tag{A.8}$$
The strict concavity of the function log implies that $J_{\ell,m}$ is minimized by d_0. Moreover, there exists a constant $C_0 > 0$ such that for all $d \in (-1, -1/2)$ and all $m \ge 1$,
$$J_{\ell,m}(d) - J_{\ell,m}(d_0) \ge C_0 (d - d_0)^2. \tag{A.9}$$
For any positive real A and positive integer n, define $\mathcal{D}_{A,n} = \{d \in (-1, -1/2)$ and $n^{\eta/2}|d - d_0| > A\}$. Applying (A.7) and (A.9), we obtain
$$\begin{aligned}\mathbb{P}(\hat{d}_n \in \mathcal{D}_{A,n}) &= \mathbb{P}\big(n^{\eta}(\hat{d}_n - d_0)^2 \ge A^2\big) \\ &\le \mathbb{P}\big(J_{\ell,m}(\hat{d}_n) - J_{\ell,m}(d_0) \ge C_0 A^2 n^{-\eta}\big) \\ &= \mathbb{P}\big(\hat{J}_{\ell,m}(\hat{d}_n) - \hat{J}_{\ell,m}(d_0) + \log(1 + E_n(d_0)) \\ &\qquad - \log(1 + E_n(\hat{d}_n)) \ge C_0 A^2 n^{-\eta}\big).\end{aligned}$$

Because $\hat{d}_n$ minimizes $\hat{J}_{\ell,m}$, it holds that $\hat{J}_{\ell,m}(\hat{d}_n) - \hat{J}_{\ell,m}(d_0) \le 0$, and hence

$$\begin{aligned}\hat{J}_{\ell,m}(\hat{d}_n) &- \hat{J}_{\ell,m}(d_0) + \log(1+E_n(d_0)) - \log(1+E_n(\hat{d}_n)) \\ &\le \log(1+E_n(d_0)) - \log(1+E_n(\hat{d}_n)) \\ &\le 2 \sup_{d\in(-1,-1/2)} \big|\log(1+E_n(d))\big|.\end{aligned}$$

Hence we obtain

$$\mathbb{P}(\hat{d}_n \in \mathcal{D}_{A,n}) \le \mathbb{P}\Big(2 \sup_{d\in(-1,-1/2)} \big|\log(1+E_n(d))\big| \ge C_0 A^2 n^{-\eta}\Big).$$

We can conclude that $\lim_{A\to\infty} \sup_{n\to\infty} \mathbb{P}(\hat{d}_n \in \mathcal{D}_{A,n}) = 0$ [which means exactly that $\hat{d}_n - d_0 = O_P(n^{-\eta/2})$] if we prove that there exists $\eta > 0$ such that

$$\sup_{d\in(-1,-1/2)} \big|\log(1+E_n(d))\big| = O_P(n^{-\eta}).$$

This follows from

$$\sup_{d\in(-1,-1/2)} |E_n(d)| = O_P(n^{-\eta}), \tag{A.10}$$

which we now prove. For $k = \ell,\dots,m$, denote $\gamma_{m,k}(d) = k^{2(d-d_0)}/\{\sum_{j=\ell}^m j^{2(d-d_0)}\}$, $\gamma_m(d) = (\gamma_{m,k}(d))_{\ell\le k\le m}$, and $w_k = f^*(0)^{-1} x_k^{2d_0} \times (I_{Y,k} - I_{\eta,k})$. Then, using the notation of Theorem A.1, we have

$$E_n(d) = \sum_{k=\ell}^m \gamma_k(d)\zeta_k + \sum_{k=\ell}^m \gamma_k(d) w_k =: \mathcal{Z}_m(\gamma_m(d)) + E_{2,n}(d).$$

The following bounds for γ_k hold (cf. Robinson 1995b). For all $d \in (-1,-1/2)$,

$$0 \le \gamma_k(d) \le Ck^{-1}\left(\frac{k}{m}\right)^{-2d_0-1} \quad \text{and} \quad |\gamma_k(d) - \gamma_{k+1}(d)| \le Ck^{-2}\left(\frac{k}{m}\right)^{-2d_0-1}. \tag{A.11}$$

Thus the assumptions of Theorem A.1 hold with $\epsilon = -2d_0 - 1$, and we have

$$\sup_{d\in(-1,-1/2)} \big|\mathcal{Z}_m(\gamma_m(d))\big| = O_P(m^{-1/2} + m/n). \tag{A.12}$$

Define $V_k = f^*(0)^{-1/2} x_k^{d_0} d_{\Delta r,k}$. Then

$$\begin{aligned}w_k &= |V_k|^2 + 2\mathrm{Re}\big(f^*(0)^{-1} x_k^{d_0} d_{\eta,k} V_k\big) \\ &= |V_k|^2 + 2\mathrm{Re}(V_k u_k) + 2\mathrm{Re}(\sqrt{2\pi}\, d_{Z,k} V_k).\end{aligned}$$

Thus we can write

$$\begin{aligned}E_{2,n}(d) &= \sum_{k=\ell}^m \gamma_k(d)|V_k|^2 + 2\sum_{k=\ell}^m \gamma_k(d)\mathrm{Re}(V_k u_k) \\ &\quad + 2\sum_{k=\ell}^m \gamma_k(d)\mathrm{Re}(\sqrt{2\pi}\, d_{Z,k} V_k) \\ &=: E_{3,n}(d) + E_{4,n}(d) + E_{5,n}(d).\end{aligned}$$

By applying the first bound in (A.11) and Lemma 1, we obtain, for all $d, d_0 \in (-1,-1/2)$,

$$\begin{aligned}|E_{3,n}(d)| &\le C\sum_{k=\ell}^m k^{-1}(k/m)^{-2d_0-1}(k/n)^{2d_0} n^{-1}(k^{-2p} + n^{-2}) \\ &= C(m/n)^{2d_0+1}\sum_{k=\ell}^m (k^{-2-2p} + k^{-2}n^{-2}) \\ &\le Cnm^{-1}(\ell^{-1-2p} + \ell^{-1}n^{-2}) \\ &\le C(\ell^{-1-2p} m^{-1} n + n^{-1}).\end{aligned} \tag{A.13}$$

By Lemma 1, (A.1), and the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}\mathbb{E}\big[|d_{\Delta r,k} u_k|\big] &\le Cn^{-1/2}(k^{-p} + n^{-1})\big(k^{-1/2} + kn^{-1}\big) \\ &\le Cn^{-1/2}\big(k^{-p-1/2} + n^{-1}\big).\end{aligned}$$

Applying this bound, (A.11), and summation by parts, we obtain

$$\begin{aligned}\mathbb{E}\Big[&\sup_{d\in(-1,-1/2)} |E_{4,n}(d)|\Big] \qquad \text{(A.14)} \\ &\le C\sum_{k=\ell}^m k^{-2}(k/m)^{-2d_0-1} \\ &\qquad \times \sum_{j=\ell}^k (j/n)^{d_0} n^{-1/2}\big(j^{-p-1/2} + n^{-1}\big) \\ &\quad + Cm^{-1}\sum_{k=\ell}^m (k/n)^{d_0} n^{-1/2}\big(k^{-p-1/2} + n^{-1}\big). \qquad \text{(A.15)}\end{aligned}$$

We first obtain the following bound:

$$\begin{aligned}\sum_{j=\ell}^k &(j/n)^{d_0} n^{-1/2}\big(j^{-p-1/2} + n^{-1}\big) \\ &\le C\big(n^{-d_0-1/2}\ell^{d_0-p+1/2} + n^{-d_0-3/2}k^{d_0+1}\big).\end{aligned} \tag{A.16}$$

Thus the right side of (A.14) is bounded by a constant times

$$\begin{aligned}m^{2d_0+1}&\big(\ell^{-d_0-p-3/2}n^{-d_0-1/2} + \ell^{-1-d_0}n^{-d_0-3/2}\big) \\ &\le \ell^{-1-p}m^{-1/2}n^{1/2} + n^{-1/2}.\end{aligned} \tag{A.17}$$

The term (A.15) is bounded by a constant times

$$\begin{aligned}\ell^{d_0-p+1/2}&m^{-1}n^{-d_0-1/2} + m^{d_0+1}n^{-d_0-3/2} \\ &\le \ell^{-p-1/2}m^{-1}n^{1/2} + n^{-1/2}.\end{aligned} \tag{A.18}$$

Because of the data taper, even though Z is a mean-0 unit variance white noise, its DFT ordinates are not uncorrelated. Nevertheless, by assumption (A2), they satisfy $\mathbb{E}[d_{Z,k}\bar{d}_{Z,k+s}] = 0$ if $s > v$ and $|\mathbb{E}[d_{Z,k}\bar{d}_{Z,k+s}]| \le 1$ if $s \le v$. Hence for any sequence of complex numbers $(a_j)_{\ell\le j\le m}$, we have the bound

$$\mathbb{E}\left[\left|\sum_{j=\ell}^m a_j d_{Z,j}\right|^2\right] \le \sum_{j=\ell}^m\sum_{s=0}^v |a_j a_{j+s}| \le v\sum_{j=\ell}^m |a_j|^2. \tag{A.19}$$

Applying (A.19) and Lemma 1, we obtain

$$\begin{aligned}\mathbb{E}&\left[\left|\sum_{j=\ell}^k x_k^{d_0} d_{\Delta r,k} d_{Z,j}\right|^2\right] \\ &\le v\sum_{j=\ell}^k x_k^{2d_0} I_{\Delta r,k} \\ &\le C(\ell^{2d_0-2p+1}n^{-2d_0-1} + n^{-2d_0-3}k^{2d_0+1}).\end{aligned} \tag{A.20}$$

Applying (A.11), (A.20), the Cauchy–Schwarz inequality, and summation by parts, we obtain

$$\begin{aligned}\mathbb{E}\Big[&\sup_{d\in(-1,-1/2)} |E_{5,n}(d)|\Big] \\ &\le Cm^{2d_0+1}\sum_{k=\ell}^m k^{-2d_0-3}\mathbb{E}^{1/2}\left[\left|\sum_{j=\ell}^k x_k^{d_0} d_{\Delta r,k} d_{Z,j}\right|^2\right]\end{aligned}$$

$$+ Cm^{-1}\mathbb{E}^{1/2}\left[\left|\sum_{j=\ell}^{m} x_k^{d_0} d_{\Delta r,k} d_{Z,j}\right|^2\right]$$
$$\le Cm^{2d_0+1}\big(\ell^{-d_0-p-3/2}n^{-d_0-1/2} + \ell^{-d_0-3/2}n^{-d_0-1/2}\big)$$
$$+ Cm^{-1}\big(\ell^{d_0-p+1/2}n^{-d_0-1/2} + n^{-d_0-3/2}m^{d_0+1/2}\big)$$
$$\le C\big(\ell^{-p-1}m^{-1/2}n^{1/2} + n^{-1/2}\big). \qquad \text{(A.21)}$$

Gathering (A.13), (A.17), (A.18), and (A.21) yields

$$\mathbb{E}\Big[\sup_{d\in(-1,-1/2)} |E_{2,n}(d)|\Big] \le C(\ell^{-2p-1}m^{-1}n + n^{-1})^{1/2},$$

which concludes the proof of (A.10).

A close inspection of the foregoing computations shows that for any real numbers c_k, $k=\ell,\dots,m$, it holds that

$$\mathbb{E}\left[\left|\sum_{k=\ell}^{m} c_k w_k\right|\right]$$
$$\le C\max_{\ell\le k\le m}|c_k|\big(\ell^{-1-2p}n + \ell^{-1/2-p}n^{1/2} + mn^{-1/2}\big). \qquad \text{(A.22)}$$

A.3 Proof of Theorem 2

Because $\hat{d}_n$ is consistent, for sufficiently large n it satisfies

$$0 = \frac{\partial \hat{J}_{\ell,m}(\hat{d}_n)}{\partial d} = \frac{2\sum_{k=\ell}^{m} k^{2\hat{d}_n}\log(k)I_{Y,k}}{\sum_{k=\ell}^{m} k^{2\hat{d}_n}I_{Y,k}} - 2\gamma_{\ell,m}.$$

This implies, by a Taylor expansion,

$$0 = \sum_{k=\ell}^{m} k^{2d_0}(\log(k)-\gamma_{\ell,m})I_{Y,k}$$
$$+ 2(\hat{d}_n - d_0)\sum_{k=\ell}^{m} k^{2\tilde{d}_n}\log(k)(\log(k)-\gamma_{\ell,m})I_{Y,k},$$

where $\tilde{d}_n$ lies between $\hat{d}_n$ and d_0. Define $\nu_k = \log(k)-\gamma_{\ell,m}$, $s_m^2 = \sum_{k=\ell}^{m}\nu_k^2$ and

$$T_m = s_m^{-2}\sum_{k=\ell}^{m} k^{2\tilde{d}_n-2d_0}\log(k)\nu_k\frac{I_{Y,k}}{x_k^{-2d_0}f^*(0)}.$$

Under the assumption $\ell/m\to 0$, it is easily seen that $\lim_{m\to\infty} s_m^2/m = 1$. Thus we prove Theorem 2 with s_m instead of $m^{1/2}$. With the notation introduced earlier, we can write

$$s_m(\hat{d}_n - d_0) = -\frac{1}{2}T_m^{-1}s_m^{-1}\sum_{k=\ell}^{m}\nu_k\frac{I_{Y,k}}{x_k^{-2d_0}f^*(0)} = -T_m^{-1}(V_n+R_n),$$

with

$$V_n = \frac{1}{2}s_m^{-1}\sum_{k=\ell}^{m}\nu_k\zeta_k \qquad\text{and}\qquad R_n = \frac{1}{2}s_m^{-1}\sum_{k=\ell}^{m}\nu_k w_k.$$

The sequence $(s_m^{-1}\nu_k)_{\ell\le k\le m}$ satisfies assumptions (A.5) and (A.6). Thus, applying Theorem A.1, part 3, V_n converges weakly to the Gaussian law with mean 0 and variance given in Theorem 2. Applying the previous bound and (A.22), we obtain

$$\mathbb{E}[|R_n|]$$
$$\le C\log(m)\big(\ell^{-1-2p}m^{-1/2}n + \ell^{-1/2-2}m^{-1/2}n^{1/2} + m^{1/2}n^{-1/2}\big).$$

Hence $R_n = o_P(1)$ under assumption (11). We now prove that T_m converges in probability to 1. By Theorem 1, $\hat{d}_n$ is $n^{-\eta}$-consistent for some $\eta>0$, and hence we need only prove that $(T_m-1)\times \mathbb{1}_{\{|\hat{d}_n-d_0|\le n^{-\eta}\}} = o_P(1)$. Write

$$T_m = 1 + s_m^{-2}\sum_{k=\ell}^{m}\big\{k^{2\hat{d}_n-2d_0}-1\big\}\log(k)\nu_k$$
$$+ s_m^{-2}\sum_{k=\ell}^{m}k^{2\hat{d}_n-2d_0}\log(k)\nu_k\left\{\frac{I_{\eta,k}}{f^*(0)x_k^{-2d_0}}-1\right\}$$
$$+ s_m^{-2}\sum_{k=\ell}^{m}\log(k)k^{2\hat{d}_n-2d_0}\nu_k w_k$$
$$=: 1 + T_{1,m} + T_{2,m} + T_{3,m}.$$

It is easily checked by a Taylor expansion that $T_{1,m}\mathbb{1}_{\{|\hat{d}_n-d_0|\le n^{-\eta}\}} = O(\log^4(m)m^{-1}n^{-\eta})$.

Denote $b_{m,k} = s_m^{-2}\log(k)\nu_k(k^{2\tilde{d}_n-2d_0}-1)\mathbb{1}_{\{|\hat{d}_n-d_0|\le n^{-\eta}\}}$ and $\mathbf{b}_m = (b_{m,k})_{\ell\le k\le m}$. Then $|b_{m,k}|\le \log^3(m)m^{-1}n^{-\eta}$ and $|b_{m,k}-b_{m,k+1}|\le Cm^{-1}k^{-1}$. Thus the assumptions of Theorem A.1, part 2, hold with $\epsilon=1$. Hence, $T_{2,m}\mathbb{1}_{\{|\hat{d}_n-d_0|\le n^{-\eta}\}} = \mathcal{Z}_m(\mathbf{b}_m) = o_P(1)$.

Finally, denote $c_k = s_m^{-2}k^{2\hat{d}_n-2d_0}\log(k)\nu_k\mathbb{1}_{\{|\hat{d}_n-d_0|\le n^{-\eta}\}}$. Then $T_{3,m} = \sum_{k=\ell}^{m}c_k w_k$ and $\max_{\ell\le k\le m}|c_k| = O(\log^2(m)m^{-1})$. Thus (A.22) yields

$$\mathbb{E}\big[|T_{3,m}|\mathbb{1}_{\{|\hat{d}_n-d_0|\le n^{-\eta}\}}\big]$$
$$\le C\log^2(m)\big(\ell^{-1-2p}m^{-1}n + \ell^{-1/2-2}m^{-1}n^{1/2} + n^{-1/2}\big)$$
$$= o(1)$$

under assumption (11). This concludes the proof of Theorem 2.

A.4 Proof of Proposition 1

The proof is exactly similar to the proof of Theorem 1. We just sketch it here. Denote

$$\tilde{E}_n(d) = \left(\sum_{j=\ell}^{m} j^{2d-2d^*}\right)^{-1}\sum_{k=\ell}^{m}k^{2d-2d^*}\big(x_k^{2d^*}f^*(0)^{-1}I_{X,k}-1\big).$$

Define $\tilde{\mathcal{D}}_{\eta,n} = \{d\in[0,1/2], |d-d^*|\ge n^{-\eta}\}$. Then, arguing as in the proof of Theorem 1, we obtain

$$\mathbb{P}(\tilde{d}_n\in\tilde{\mathcal{D}}_{\eta,n}) \le \mathbb{P}\Big(2\sup_{d\in[0,1/2]}\big|\log(1+\tilde{E}_n(d))\big| \ge Cn^{-2\eta}\Big),$$

for some constant C. This last probability tends to 0 as long as $\tilde{E}_n(d) = o_P(n^{-2\eta})$. Actually, under our assumptions, we can prove that if $m = [n^{4/5-\epsilon}]$ and $\ell = n^{3/5+\epsilon}$, then, for all $d^*\in(0,1/2)$,

$$\sup_{d\in[0,1/2]}|\tilde{E}_n(d)| = O_P\big(m^{-1/2}\big) = O_P\big(n^{-2/5+\epsilon/2}\big). \qquad \text{(A.23)}$$

We do not give a full proof of (A.23), the main ingredient of which would be summation by parts as in the proof of Theorem A.1, part 2. To understand why this is true, we prove only that $\tilde{E}_n(d^*) = O_P(n/ml)$, that is,

$$\frac{1}{m}\sum_{k=1}^{m}\big\{f^*(0)^{-1}x_k^{2d^*}I_{X,k}-1\big\} = O_P(n/ml). \qquad \text{(A.24)}$$

This in any case yields a lower bound for $\sup_{d\in[0,1/2]}|\tilde{E}_n(d)|$. Because $X_t = r(t/n)+\epsilon_t$, we have $I_{X,k} = I_{\epsilon_k} + 2\,\mathrm{Re}(d_{r,k}d_{\epsilon_k}) + I_{r,k}$. By standard arguments, under our assumptions, it holds that

$$\frac{1}{m}\sum_{k=1}^{m}\left\{f^*(0)^{-1}x_k^{2d^*}I_{\epsilon,k}-1\right\}=O_P(m^{-1/2}). \tag{A.25}$$

Now consider the term $I_{r,k}$. Because r is twice continuously differentiable, by arguments similar to the proof of Lemma 1, we have $d_{r,k}=O(n^{1/2}k^{-1})$ and $I_{r,k}=O(nk^{-2})$. This bound cannot be improved unless the periodic extension of r is smooth, which necessitates $r(0)=r(1)$. If this does not hold, then this bound is sharp; consider, for instance, $r(t)=t$. Thus we obtain

$$\frac{1}{m}\sum_{k=1}^{m}x_k^{2d^*}I_{r,k}=O(n/ml).$$

If $m=[n^{4/5-\epsilon}]$ and $\ell=[n^{3/5+2\epsilon}]$, then $n/ml=o(m^{-1/2})$. Thus the main term in $E_n(d^*)$ comes from (A.25).

A.5 Proof of Lemma 1

By a Taylor expansion, and because $\sum_{t=1}^n|h(t/n)|\le H_n^{1/2}n^{1/2}$, we have

$$\left|d_{\Delta r,k}-n^{-1}(2\pi H_n)^{-1/2}\sum_{t=1}^{n}h(t/n)r'(t/n)\exp(ix_kt)\right| \le C\|r''\|_\infty n^{-3/2}.$$

Under assumption (A1), the periodic extension of hr' is continuously differentiable. Hence its Fourier series, defined as $c_j:=\int_0^1 h(s)r'(s)\mathrm{e}^{2i\pi js}\,\mathrm{d}s$, is absolutely summable and $h(s)r'(s)=\sum_{j\in\mathbb{Z}}c_j\mathrm{e}^{-2i\pi js}$. Moreover, by assumption, the derivatives up to the order $p-1$ of hr' vanish at 0 and 1. This implies that $\sum_{j\in\mathbb{Z}}j^{2p}|c_j|^2<\infty$. Hence we have

$$n^{-1}\sum_{t=1}^{n}h(t/n)r'(t/n)\exp(2i\pi kt/n)=\sum_{j\in\mathbb{Z}}c_jn^{-1}\sum_{t=1}^{n}\exp\bigl(2i\pi(k-j)t/n\bigr)=\sum_{\lambda\in\mathbb{Z}}c_{k+\lambda n}.$$

By applying Hölder's inequality and noting that $k\le n/2$, we bound this last series

$$\begin{aligned}\sum_{\lambda\in\mathbb{Z}}|c_{k+\lambda n}| &\le\left\{\sum_{\lambda\in\mathbb{Z}}(k+\lambda n)^{2p}|c_{k+\lambda n}|^2\right\}^{1/2}\left\{\sum_{\lambda\in\mathbb{Z}}(k+\lambda n)^{-2p}\right\}^{1/2}\\ &\le Ck^{-p},\end{aligned}$$

for some constant C depending only on h, r', and their derivatives up to the order p.

A.6 Proof of Theorem A.1

Proof of Part 1. Denote $a_k=f^*(0)^{1/2}x_k^{-d_0}$ and $D_n(x)=(2\pi H_n)^{-1/2}\sum_{t=1}^n h(t/n)\mathrm{e}^{itx}$, and let $\kappa=\mathrm{cum}(Z_1,Z_1,Z_1,Z_1)$ be the fourth cumulant of Z_1. Also define

$$\begin{aligned}q_k&=\left|\int_{-\pi}^{\pi}\left(\frac{|a(x)|^2}{|a_k|^2}-1\right)|D_n(x-x_k)|^2\,\mathrm{d}x\right|,\\ q_{j,k}&=\left|\int_{-\pi}^{\pi}\left(\frac{a(x)}{a_k}-1\right)D_n(x_k-x)\overline{D_n(x_j-x)}\,\mathrm{d}x\right|\\ &\quad+\left|\int_{-\pi}^{\pi}\left(\frac{a(x)}{a_k}-1\right)D_n(x_k-x)D_n(x_j+x)\,\mathrm{d}x\right|.\end{aligned}$$

By straightforward algebra, we obtain

$$\begin{aligned}\mathbb{E}[|u_k-v_k|^2]&=\int_{-\pi}^{\pi}\left(\frac{|a(x)|^2}{|a_k|^2}-1\right)|D_n(x-x_k)|^2\,\mathrm{d}x\\ &\quad-2\,\mathrm{Re}\left(\int_{-\pi}^{\pi}\left(\frac{a(x)}{a_k}-1\right)|D_n(x-x_k)|^2\,\mathrm{d}x\right)\\ &\le p_k+2q_{k,k},\end{aligned}$$

$$\begin{aligned}\bigl|\mathbb{E}[\bar v_j(u_k-v_k)]\bigr|&=\left|\int_{-\pi}^{\pi}\left(\frac{a(x)}{a_k}-1\right)D_n(x_k-x)\overline{D_n(x_j-x)}\,\mathrm{d}x\right|\\ &\le q_{j,k},\end{aligned}$$

$$\begin{aligned}\bigl|\mathbb{E}[v_j(u_k-v_k)]\bigr|&=\left|\int_{-\pi}^{\pi}\left(\frac{a(x)}{a_k}-1\right)D_n(x_k-x)D_n(x_j+x)\,\mathrm{d}x\right|\\ &\le q_{j,k},\end{aligned}$$

and

$$\begin{aligned}&\bigl|\mathrm{cum}(\bar v_k,u_k-v_k,v_j,\bar u_j-\bar v_j)\bigr|\\ &=\left|\frac{\kappa}{2\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left(\frac{a(x)}{a_k}-1\right)\left(\frac{a(y)}{a_j}-1\right)\right.\\ &\qquad\left.\times\ D_n(x_k-x)\,D_n(x_j-y)\overline{D_n(x_k-x-y+z)D_n(x_j-z)}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\right|\\ &\le cn^{-1/2}\left(\int_{-\pi}^{\pi}\left|\frac{a(x)}{a_k}-1\right|^2|D_n(x_k-x)|^2\,\mathrm{d}x\right)^{1/2}\\ &\qquad\times\left(\int_{-\pi}^{\pi}\left|\frac{a(x)}{a_j}-1\right|^2|D_n(x_j-x)|^2\,\mathrm{d}x\right)^{1/2}\\ &\le c\sqrt{(q_j+2q_{j,j})(q_k+2q_{k,k})/n}.\end{aligned}$$

Thus there only remains to prove that there exists a constant C such that $q_k\le Cp_{k,k}$ and $q_{j,k}\le Cp_{j,k}$. Such computations have already been done by many authors, starting with Robinson (1995a,b). The key ingredient is (6), which describes the decay of the kernel D_n. Denote $\alpha=-d_0\in(1/2,1)<p$. Because a^* is assumed to be twice differentiable in a neighborhood of 0, by symmetry, it holds that

$$|a^*(x)-a^*(0)|\le Cx^2,$$

and the function $a(x)=x^\alpha a^*(x)$ has the following properties:

$$C^{-1}|x|^\alpha\le|a(x)|\le C|x|^\alpha \tag{A.26}$$

and

$$|a(x)-a_k|\le C\bigl(x^{\alpha+2}+(x\wedge x_k)^{\alpha-1}|x-x_k|\bigr). \tag{A.27}$$

Then (6) trivially implies the following bound for the kernel D_n:

$$|D_n(x)|\le C\frac{n^{1/2}}{(1+n|x|)^{p+1}}. \tag{A.28}$$

Proof of $q_k\le Cp_{k,k}$. Define $A_k(x):=(|a(x)/a_k|^2-1)\times|D_n(x_k-x)|^2$. If $|x|\in[\vartheta,\pi]$, then $|x-x_k|\ge\vartheta/2$, and by (A.28), $|D_n(x_k-x)|^2\le Cn^{-2p-1}$. Thus

$$\begin{aligned}\int_{\vartheta\le|x|\le\pi}|A_k(x)|\,\mathrm{d}x&\le Cn^{-2p-1}\left(1+x_k^{-\alpha}\int_{-\pi}^{\pi}|a(x)|\,\mathrm{d}x\right)\\ &\le Ck^{-\alpha}n^{\alpha-2p-1}\\ &\le Ck^{-1}.\end{aligned}$$

Now consider the integral over $[-\vartheta,\vartheta]$. If $x\in[2x_k,\vartheta]\cup[-\vartheta,-x_k/2]$, then $|x-x_k|\ge x/2$. Applying (A.26) and (A.28), we obtain

$$\begin{aligned}\int_{[2x_k,\vartheta]\cup[-\vartheta,-x_k/2]}|A_k(x)|\,\mathrm{d}x &\le Cn^{-2p-1}\int_{x_k/2}^{\vartheta}(1+x_k^{-2\alpha}x^{2\alpha})x^{-2p-2}\,\mathrm{d}x\\ &\le Ck^{-2p-1}.\end{aligned}$$

If $|x|\le x_k/2$, then $|x-x_k|>x_k/2$. Applying (A.28) and (A.26), we obtain

$$\begin{aligned}\int_{-x_k/2}^{x_k/2}|A_k(x)|\,\mathrm{d}x &\le \frac{Cn}{(1+nx_k/2)^{2p+2}}\left\{x_k+x_k^{-2\alpha}\int_0^{x_k}x^{2\alpha}\,\mathrm{d}x\right\}\\ &\le Ck^{-2p-1}.\end{aligned}$$

For $x\in[x_k/2,2x_k]$, applying (A.28) and (A.27), we have

$$|A_k(x)|\le Cn(x_k^{-1}|x-x_k|+x_k^2)(1+n|x-x_k|)^{-2p-2}.$$

Hence we obtain

$$\int_{x_k/2}^{2x_k}|A_k(x)|\,\mathrm{d}x\le C(k^{-1}+(k/n)^2).$$

Gathering the different terms yields $q_k\le Cp_{k,k}$.

Proof of $q_{j,k}\le Cp_{j,k}$. Denote $A_{j,k}(x)=(a_k^{-1}a(x)-1)D_n(x-x_k)\overline{D_n(x-x_j)}$. If $|x|\in(\vartheta,\pi]$, then $|x-x_k|\ge\vartheta/2$, $|x-x_j|\ge\vartheta/2$ and, by (A.28), $|D_n(x-x_k)\overline{D_n(x-x_j)}|\le Cn^{-2p-1}$. Thus, by the same arguments as before, we obtain

$$\begin{aligned}\int_{\vartheta\le|x|\le\pi}|A_{j,k}(x)|\,\mathrm{d}x &\le Cn^{-2p-1}\left(1+x_k^{-\alpha}\int_{-\pi}^{\pi}|a(x)|\,\mathrm{d}x\right)\\ &\le Ck^{-\alpha}n^{\alpha-2p-1}\le C(j\vee k)^{-1}.\end{aligned}$$

Denote $y=\min(x_k,x_j)$ and $z=\max(x_k,x_j)$. If $x\in[-\vartheta,-y]\cup[2z,\vartheta]$, then $|x-y|\ge|x|$, $|x-z|\ge z$. Hence (A.28) and $|D_n(x-y)D_n(x-z)|\le Cn^{-2p-1}z^{-p-1}|x|^{-p-1}$, and thus

$$\begin{aligned}\int_{[-\vartheta,-y]\cup[2z,\vartheta]}|A_{j,k}(x)|\,\mathrm{d}x &\le Cn^{-2p-1}z^{-p-1}\int_y^{\vartheta}(1+(x/x_k)^{\alpha})x^{-p-1}\,\mathrm{d}x\\ &\le Cn^{-2p-1}(1+(y/x_k)^{\alpha})y^{-p}z^{-p-1}\\ &\le C(ny)^{-p}(nz)^{-p-1}\le C(j\vee k)^{-1}.\end{aligned}$$

If $x\in[-y,y/2]$, then $|x-x_k|\ge x_k/2$ and $|x-x_j|\ge x_j/2$, and $|a(x)/a_k|$ is bounded because of (A.26). Thus, applying (A.28), we obtain

$$\int_{-y}^{y/2}|A_{j,k}(x)|\,\mathrm{d}x\le C(ny)^{-p}(nz)^{-p-1}.$$

If $x\in[y/2,2z]$, applying (A.27), we obtain

$$\begin{aligned}\int_{y/2}^{2z}|A_{j,k}(x)|\,\mathrm{d}x &\le Cx_k^{-\alpha}y^{\alpha-1}\int_{y/2}^{2z}|x-x_k||D_n(x-y)||D_n(x-z)|\,\mathrm{d}x\\ &\quad+Cx_k^{-\alpha}\int_{y/2}^{2z}x^{\alpha+2}|D_n(x-y)||D_n(x-z)|\,\mathrm{d}x\\ &\le C(y^{-1}|z-y|+x_k^{-\alpha}z^{\alpha+2})\int_{y/2}^{2z}|D_n(x-y)||D_n(x-z)|\,\mathrm{d}x.\end{aligned}$$

By splitting this last integral at $(y+z)/2$ and applying (A.28), we easily obtain that

$$\int_{y/2}^{2z}|D_n(x-y)||D_n(x-z)|\,\mathrm{d}x\le(1+n|z-y|)^{-p-1}.$$

Hence

$$\begin{aligned}\int_{y/2}^{2z}|A_{j,k}(x)|\,\mathrm{d}x\\ \le C\big(n^{-p-1}y^{-1}(z-y)^{-p}+x_k^{-\alpha}z^{\alpha+2}(n(z-y))^{-p-1}\big).\end{aligned}$$

By considering the cases $y<z/2$ and $z/2\le y<z$ separately, we obtain

$$\int_{y/2}^{2z}|A_{j,k}(x)|\,\mathrm{d}x\le C\big((j\vee k)^{-1}+\{(j\vee k)/n\}^2\big).$$

Thus we have proven that $|\int_{-\pi}^{\pi}A_{j,k}(x)\,\mathrm{d}x|\le Cp_{j,k}$. To prove that $|\int_{-\pi}^{\pi}(a(x)/a_k-1)D_n(x_k-x)D_n(x_j+x)\,\mathrm{d}x|\le Cp_{j,k}$, the arguments are the same as before, but the interval $[-\vartheta,\vartheta]$ is split differently. We bound the integral separately over the intervals $[-\vartheta,-2x_j]$, $[-2x_j,-x_j/2]$, $[-x_j/2,x_k/2]$, $[x_k/2,2x_k]$, and $[2x_k,\vartheta]$.

This concludes the proof of the first item of Theorem A.1.

Proof of Part 2. Define $r_k=x_k^{2d_0}f^*(0)^{-1}I_{X,k}-2\pi I_{Z,k}$. The Cauchy–Schwarz inequality and (A.1) imply that

$$\mathbb{E}[|r_k|]=\mathbb{E}\big[|x_k^{2d_0}f^*(0)^{-1}\zeta_k-2\pi I_{Z,k}|\big]\le Cp_{k,k}^{1/2}.\qquad\text{(A.29)}$$

For $\mathbf{c}\in\mathbb{R}^m$, write

$$\mathcal{Z}_m(\mathbf{c})=\sum_{k=1}^m c_kr_k+\sum_{k=1}^m c_k(2\pi I_{Z,k}-1)=:\mathcal{Z}_{1,m}(\mathbf{c})+\mathcal{Z}_{2,m}(\mathbf{c}).$$

Applying (A.29) and summation by parts, we obtain

$$\begin{aligned}\mathbb{E}\Big[\sup_{\mathbf{c}\in\mathcal{C}_m(\epsilon,K)}|\mathcal{Z}_{1,m}(\mathbf{c})|\Big]&\\ \le Km^{-\epsilon}\sum_{k=1}^{m-1}k^{\epsilon-2}\sum_{j=1}^{k}\mathbb{E}[|r_j|]+Km^{-1}\sum_{j=1}^{m}\mathbb{E}[|r_j|]&\\ \le Cm^{-\epsilon}\sum_{k=1}^{m-1}k^{\epsilon-2}\sum_{j=1}^{k}\big(j^{-1/2}+(j/n)\big)&\\ +Cm^{-1}\sum_{j=1}^{m}\big(j^{-1/2}+(j/n)\big)&\\ \le C\big(m^{-1/2}+(m/n)\big).&\end{aligned}$$

To deal with $\mathcal{Z}_{2,m}$, write

$$\begin{aligned}\mathcal{Z}_{2,m}(\mathbf{c}) &= \sum_{k=1}^m c_k\frac{1}{n}\sum_{t=1}^n(Z_t^2-1)\\ &\quad+\sum_{k=1}^m c_kH_n^{-1}\sum_{1\le s\ne t\le n}h(s/n)\bar{h}(t/n)\mathrm{e}^{i(t-s)x_k}Z_sZ_t,\\ &:= \mathcal{Z}_{3,m}(\mathbf{c})+\mathcal{Z}_{4,m}(\mathbf{c}).\end{aligned}$$

Equation (A.4) implies that $\sum_{k=1}^m|c_k|$ is uniformly bounded over the class $\mathcal{C}_m(\epsilon,K)$. Assumption (A3) implies that $\frac{1}{n}\sum_{t=1}^n(Z_t^2-1)=O_P(n^{-1/2})$. Hence, $\sup_{\mathbf{c}\in\mathcal{C}_m(\epsilon,K)}\mathcal{Z}_{3,m}(\mathbf{c})=O_P(n^{-1/2})$. Because h is bounded, there exists a constant $C(h)$ that depends only on h, such that

$$\mathbb{E}\left[\left(H_n^{-1}\sum_{1\le s\ne t\le n}h(s/n)\bar{h}(t/n)Z_sZ_t\sum_{j=1}^{k}\mathrm{e}^{i(t-s)x_j}\right)^2\right]\le C(h)k.$$

Hence, applying summation by parts, we obtain

$$\mathbb{E}\Big[\sup_{\mathbf{c}\in\mathcal{C}_m(\epsilon,K)} \mathcal{Z}_{3,m}(\mathbf{c})\Big] \le CKm^{-\epsilon}\sum_{k=1}^{m-1} k^{\epsilon-2}k^{1/2} + CKm^{-1/2}$$
$$= O(m^{-1/2}).$$

Proof of Part 3. Using the previous notations, it suffices to show that $\mathcal{Z}_{1,m}(\mathbf{c}_m)$ converges to 0 in probability and $\mathcal{Z}_{2,m}(\mathbf{c}_m)$ is asymptotically mean-0 Gaussian with variance $\sigma^2(h)$. Applying the relation $|u_k|^2 - |v_k|^2 = |u_k - v_k|^2 + 2\,\mathrm{Re}(\bar{v}_k(u_k - v_k))$, we have

$$\mathbb{E}\big[|\mathcal{Z}_{1,m}(\mathbf{c}_m)|\big] \le \sum_{k=1}^{m} |c_{m,k}|\mathbb{E}[|u_k - v_k|^2] + 2\mathbb{E}\left[\left|\sum_{k=1}^{m} c_{m,k}\bar{v}_k(u_k - v_k)\right|^2\right]^{1/2} =: R_{1,n} + 2R_{2,n}.$$

Applying (A.1), (A.6), and summation by parts, we obtain $R_{1,n} = o(1)$. To bound $R_{2,n}$, we use the identity

$$\mathbb{E}[\bar{v}_k(u_k - v_k)v_j(\bar{u}_j - \bar{v}_j)] = \mathbb{E}[|u_k - v_k|^2]\mathbb{1}_{\{j=k\}} + \mathbb{E}[\bar{v}_k(u_k - v_k)]\mathbb{E}[v_j(\bar{u}_j - \bar{v}_j)] + \mathbb{E}[\bar{v}_k(\bar{u}_j - \bar{v}_j)]\mathbb{E}[v_j(u_k - v_k)] + \mathrm{cum}(\bar{v}_k, u_k - v_k, v_j, \bar{u}_j - \bar{v}_j).$$

Applying this identity, (A.1), (A.2), and (A.3) yields

$$R_{2,n}^2 \le C\sum_{k=1}^{m} c_{m,k}^2 p_{k,k} + \sum_{1\le j\le j\le m} c_{m,k}c_{m,j}\{p_{j,j}p_{k,k} + p_{j,k}p_{k,j} + n^{-1/2}\sqrt{p_{j,j}p_{k,k}}\}.$$

The definition of $p_{j,k}$, (A.6), and summation by parts yield that $R_{2,n} = o(1)$. Thus we have proven that $\mathcal{Z}_{1,m}(\mathbf{c}_m) = o_P(1)$.

Under (A.5), we can write $\mathcal{Z}_{2,m}(\mathbf{c}_m) = \sum_{t=2}^{n}\sum_{s=1}^{t-1} C_{n,s,t}Z_sZ_t$, with

$$C_{n,s,t} := \frac{2}{H_n}\sum_{k=1}^{m} c_{m,k}\,\mathrm{Re}\left(h(s/n)\bar{h}(t/n)e^{i(t-s)x_k}\right).$$

By the martingale central limit theorem (cf., e.g., Hall and Heyde 1980), $\mathcal{Z}_{2,m}(\mathbf{c}_m)$ is asymptotically $\mathcal{N}(0, \Sigma_v)$ if we prove that

$$\sum_{t=1}^{n}\left(\sum_{s=1}^{t-1} C_{n,s,t}Z_s\right)^2 \xrightarrow{P} \Sigma_v \tag{A.30}$$

and

$$\sum_{t=1}^{n}\mathbb{E}\left[\left(\sum_{s=1}^{t-1} C_{n,s,t}Z_s\right)^4\right] \to 0. \tag{A.31}$$

To prove (A.30), write $\sum_{t=2}^{n}(\sum_{s=1}^{t-1} C_{n,s,t}Z_s)^2 =: A_n + B_n + 2C_n$, with

$$A_n := \sum_{t=2}^{n}\sum_{s=1}^{t-1} C_{n,s,t}^2,$$

$$B_n := \sum_{t=2}^{n}\sum_{s=1}^{t-1}(Z_s^2 - 1)C_{n,s,t}^2, \qquad \text{and}$$

$$C_n := \sum_{t=2}^{n}\sum_{1\le r<s<t} C_{n,r,t}C_{n,s,t}Z_rZ_s.$$

Under assumption (A2) and conditions (A.5) and (A.6), it can be shown that $\lim_{n\to\infty} A_n = \sigma^2(h)$ and, applying summation by parts, it can also be shown that

$$\max_{1\le s\le n}\sum_{t=1}^{n} C_{n,s,t}^2 = O(n^{-1}) \tag{A.32}$$

and

$$|C_{n,s,t}| \le Kc_m^*|t - s|^{-1}, \tag{A.33}$$

with $c_m^* = (|c_{m,m}| + \sum_{k=1}^{m-1}|c_{m,k+1} - c_{m,k}|)$ and for some numerical constant K. Assumption (A3) implies that $\mathbb{E}[Z_t^2 - 1] = 0$ and $\mathbb{E}[(Z_s^2 - 1)(Z_t^2 - 1)] = 0$ if $s \ne t$ and is uniformly bounded if $s = t$. Hence B_n has mean 0 and, applying (A.33), we have

$$\mathbb{E}[B_n^2] \le C\sum_{s=1}^{n-1}\sum_{s<t,u\le n} C_{n,s,t}^2C_{n,s,u}^2 \le A_n\max_{1\le s\le n}\sum_{t=1}^{n} C_{n,s,t}^2 = O(n^{-1}).$$

Under Assumption (A3), if $s < t$ and $u < v$, then $\mathbb{E}[Z_sZ_tZ_uZ_v] = 0$ if $s \ne u$ or $t \ne v$. Applying the Cauchy–Schwarz inequality, (A.32), (A.33) and condition (A.6), we obtain

$$\begin{aligned}\mathbb{E}[C_n^2] &= \sum_{1\le r<s<n}\left(\sum_{s<t\le n} C_{n,r,t}C_{n,s,t}\right)^2 \\ &\le \sum_{1\le r<s<n}\sum_{s<t\le n} C_{n,r,t}^2\sum_{s<t\le n} C_{n,s,t}^2 \\ &\le Cn^{-1}c_m^{*2}\sum_{1\le r<s<n}\sum_{s<t\le n}(t - r)^{-2} \\ &\le Cc_m^{*2}\log(n) = o(1).\end{aligned}$$

Thus (A.30) holds. Under (A3), applying Rosenthal's inequality for martingale differences (cf. Hall and Heyde 1980, thm. 2.12) and (A.32), we obtain

$$\sum_{t=1}^{n}\mathbb{E}\left[\left(\sum_{s=1}^{t-1} C_{n,s,t}Z_s\right)^4\right] \le C\sum_{t=1}^{n}\sum_{s=1}^{t-1} C_{n,s,t}^4 + C\sum_{t=1}^{n}\left(\sum_{s=1}^{t-1} C_{n,s,t}^2\right)^2 = O(n^{-1}).$$

This concludes the proof of the central limit theorem.

[Received July 2002. Revised November 2004.]

REFERENCES

Andrews, D. W. K., and Sun, Y. (2001), "Local Polynomial Whittle Estimation of Long-Range Dependence," preprint, Yale University Cowles Foundation.

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

Chen, W. (2001), Tapering and Nonstationarity/Noninvertibility in Time Series," presented at a symposium on *New Directions in Time Series Analysis*, CIRM, Luminy, France, April 2001.

Csörgo, S., and Mielniczuk, J. (1995a), "Interrupt," *The Annals of Statistics*, 23, 1000–1014.

——— (1995b), "Distant Long-Range Dependent Sums and Regression Estimation," *Stochastic Processes and Their Applications*, 59, 143–155.

Dahlhaus, R. (1988), "Empirical Spectral Processes and Their Applications to Time Series Analysis," *Stochastic Processes and Their Applications*, 30, 69–83.

Davies, R. B., and Harte, D. S. (1987), "Tests for Hurst Effect," *Biometrika*, 74, 95–101.

Delgado, M. A., and Robinson, P. M. (1996), "Optimal Spectral Bandwidth for Long Memory," *Statistica Sinica*, 6, 97–112.

Deo, R. S. (1997), "Nonparametric Regression With Long-Memory Errors," *Statistics and Probability Letters*, 33, 89–94.

Hall, P., and Hart, J. D. (1990), "Nonparametric Regression With Long-Range Dependence," *Stochastic Processes and Their Applications*, 36, 339–351.

Hall, P., and Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, New York: Academic Press.

Hart, J. D. (1989), "Differencing as an Approximate De-Trending Device," *Stochastic Processes and Their Applications*, 31, 251–259.

Henry, M. (2001), "Robust Automatic Bandwidth for Long Memory," *Journal of Time Series Analysis*, 22, 293–316.

Hurvich, C. M., and Beltrao, K. I. (1994), "Automatic Semiparametric Estimation of the Memory Parameter of a Long-Memory Time Series," *Journal of Time Series Analysis*, 15, 285–302.

Hurvich, C. M., and Chen, W. W. (2000), "An Efficient Taper for Potentially Overdifferenced Long-Memory Time Series," *Journal of Time Series Analysis*, 21, 155–180.

Hurvich, C. M., and Deo, R. S. (1999), "Plug-in Selection of the Number of Frequencies in Regression Estimates of the Memory Parameter of a Long-Memory Time Series," *Journal of Time Series Analysis*, 20, 331–341.

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," *Journal of Time Series Analysis*, 19, 19–46.

Hurvich, C. M., Moulines, E., and Soulier, P. (2001), "Estimating Long Memory in Stochastic Volatility," Prépublication 153, l'Université d'Evry Val d'Essonne.

——— (2002), "The FEXP Estimator for Linear, Possibly Nonstationary Processes," *Stochastic Processes and Their Applications*, 97, 307–340.

Künsch, H. R. (1986), "Discrimination Between Monotonic Trends and Long-Range Dependence," *Journal of Applied Probability*, 23, 1025–1030.

Moulines, E., and Soulier, P. (1999), "Broadband Log-Periodogram Regression of Time Series With Long-Range Dependence," *The Annals of Statistics*, 27, 1415–1439.

Robinson, P. M. (1994), "Semiparametric Analysis of Long-Memory Time Series," *The Annals of Statistics*, 22, 515–539.

——— (1995a), "Log-Periodogram Regression of Time Series With Long Range Dependence," *The Annals of Statistics*, 23, 1048–1072.

——— (1995b), "Gaussian Semiparametric Estimation of Long Range Dependence," *The Annals of Statistics*, 23, 1630–1661.

——— (1997), "Large-Sample Inference for Nonparametric Regression With Dependent Errors," *The Annals of Statistics*, 25, 2054–2083.

Velasco, C. (1999a), "Non-Stationary Log-Periodogram Regression," *Journal of Econometrics*, 91, 325–371.

——— (1999b), "Gaussian Semiparametric Estimation of Non-Stationary Time Series," *Journal of Time Series Analysis*, 20, 87–127.