NONLINEAR LOG-PERIODOGRAM REGRESSIOIN FOR PERTURBED FRACTIONAL PROCESSES

BY

YIXIAO SUN
AND
PETER C.B. PHILLIPS

COWLES FOUNDATION PAPER NO. 1077

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281
2004

Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Econometrics 115 (2003) 355 – 389
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

Nonlinear log-periodogram regression for perturbed fractional processes☆

Yixiao Sun[a,*], Peter C.B. Phillips[b]

[a] Department of Economics 0508, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA

[b] Cowles Foundation, Box 208281, Yale University, University of Auckland & University of York, New Haven, CT 06520-8281, USA

Accepted 6 February 2003

Abstract

This paper studies fractional processes that may be perturbed by weakly dependent time series. The model for a perturbed fractional process has a components framework in which there may be components of both long and short memory. All commonly used estimates of the long memory parameter (such as log periodogram (LP) regression) may be used in a components model where the data are affected by weakly dependent perturbations, but these estimates can suffer from serious downward bias. To circumvent this problem, the present paper proposes a new procedure that allows for the possible presence of additive perturbations in the data. The new estimator resembles the LP regression estimator but involves an additional (nonlinear) term in the regression that takes account of possible perturbation effects in the data. Under some smoothness assumptions at the origin, the bias of the new estimator is shown to disappear at a faster rate than that of the LP estimator, while its asymptotic variance is inflated only by a multiplicative constant. In consequence, the optimal rate of convergence to zero of the asymptotic MSE of the new estimator is faster than that of the LP estimator. Some simulation results demonstrate the viability and the bias-reducing feature of the new estimator relative to the LP estimator in finite samples. A test for the presence of perturbations in the data is given.

JEL classification: C13; C14; C22; C51

Keywords: Asymptotic bias; Asymptotic normality; Bias reduction; Fractional components model; Perturbed fractional process; Rate of convergence; Testing perturbations

☆ The authors thank Feng Zhu for his careful reading of the first draft and helpful comments. Sun acknowledges the Cowles Foundation for support under a Carl Anderson fellowship. Phillips thanks the NSF for support under Grant No. SES 0092509.

* Corresponding author. Tel.: +1-858-534-4692; fax: +1-858-534-7040.
E-mail addresses: yisun@ucsd.edu (Y. Sun), peter.phillips@yale.edu (P.C.B. Phillips).

0304-4076/03/$ - see front matter
doi:10.1016/S0304-4076(03)00115-5

1. Introduction

Fractional processes have been gaining increasing popularity with empirical researchers in economics and finance. In part, this is because fractional processes can capture forms of long run behavior in economic variables that elude other models, a feature that has proved particularly important in modelling inter-trade durations and the volatility of financial asset returns. In part also, fractional processes are attractive to empirical analysts because they allow for varying degrees of persistence, including a continuum of possibilities between weakly dependent and unit root processes.

For a pure fractional process, short run dynamics and long run behavior are driven by the same innovations. This may be considered restrictive in that the innovations that drive long run behavior may arise from quite different sources and therefore differ from those that determine the short run fluctuations of a process. To accommodate this possibility, the model we consider in the present paper allows for perturbations in a fractional process and has a components structure that introduces different sources and types of variation. Such models provide a mechanism for simultaneously capturing the effects of persistent and temporary shocks on the realized observations. They seem particularly realistic in economic and financial applications when there are many different sources of variation in the data and both long run behavior and short run fluctuations need to be modeled.

Specifically, a perturbed fractional process z_t is defined as a fractional process y_t that is perturbed by a weakly dependent process u_t as follows

$$z_t = y_t + \mu + u_t, \quad t = 1, 2, \ldots, n, \tag{1}$$

where μ is a constant and

$$y_t = (1-L)^{-d_0} w_t = \sum_{k=0}^{\infty} \frac{\Gamma(d_0 + k)}{\Gamma(d_0)\Gamma(k+1)} w_{t-k}, \quad 0 < d_0 < 1/2. \tag{2}$$

Here, y_t is a pure fractional process and u_t and w_t are independent Gaussian processes with zero means and continuous spectral densities $f_u(\lambda)$ and $f_w(\lambda)$, respectively. We confine attention to the case where the memory parameter $d_0 \in (0, \frac{1}{2})$ largely for technical reasons that will become apparent later. The case is certainly the most relevant in empirical practice, at least for stationary series, but the restriction is an important one. To maintain generality in the short run components of z_t, we do not impose specific functional forms on $f_u(\lambda)$ and $f_w(\lambda)$. Instead, we allow them to belong to a family that is characterized only by regularity conditions near the zero frequency. This formulation corresponds to the conventional semiparametric approach to modelling long range dependence.

By allowing for the presence of two separate stochastic components, the model (1) captures mechanisms in which different factors may come into play in determining long run and short run behaviors. Such mechanisms may be expected to occur in the generation of macroeconomic and financial data for several reasons. For example, time series observations of macroeconomic processes often reflect short run competitive forces as well as long run growth determinants. Additionally, economic and financial time series frequently arise from processes of aggregation and involve errors of measurement, so that the presence of an additive, short memory disturbance is quite realistic. For instance, if the underlying volatility of stock returns follows a fractional process, then realized volatility may follow a perturbed fractional process because the presence of a bid-ask bounce adds a short memory component to realized returns, with consequent effects on volatility.

Some empirical models now in use are actually special cases of perturbed fractional processes. Among these, the long memory stochastic volatility model (LMSV) is growing in popularity for modelling the volatility of financial time series (see Anderson and Bollerslev, 1997; Breidt et al., 1998; and Deo and Hurvich, 2001). This model assumes that $\log r_t^2 = y_t + \mu + u_t$, where r_t is the return, y_t is an underlying fractional process and $u_t = iid(0, \sigma^2)$, thereby coming within the framework of (1). Another example is a rational expectation model in which the ex ante variable follows a fractional process, so that the corresponding ex post variable follows (1) with u_t being a martingale difference sequence. Sun and Phillips (2002) used this framework to model the real rate of interest and inflation as perturbed fractional processes and found that this model helped explain the empirical incompatibility of memory parameter estimates of the components in the ex post Fisher equation. The study by Granger and Marmol (1997) provides a third example, addressing the frequently observed property of financial time series that the autocorrelogram can be low but positive for many lags. Granger and Marmol explained this phenomenon by considering time series that consist of a long memory component combined with a white noise component that has a much larger variance, again coming within the framework of (1).

The main object in the present paper is to develop a suitable estimation procedure for the memory parameter d_0 in (1). As we will show, existing procedures for estimating d_0 typically suffer from serious downward bias in models where there are additive perturbations like (1). The present paper therefore proposes a new procedure that allows for the possible presence of such perturbations in the data.

The spectral density $f_z(\lambda)$ of z_t can be written as $f_z(\lambda) = (2\sin\lambda/2)^{-2d_0} f^*(\lambda)$, where $f^*(\lambda) = f_w(\lambda) + (2\sin\lambda/2)^{2d_0} f_u(\lambda)$ is a continuous function over $[0, \pi]$. So, $f_z(\lambda)$ satisfies a power law around the origin of the form $f_z(\lambda) \sim G_0 \lambda^{-2d_0}$ as $\lambda \to 0+$, for some positive constant G_0. Therefore, we can estimate d_0 by using the linear log-periodogram (LP) regression introduced by Geweke and Porter-Hudak (1983). Building on the earlier work of Künsch (1986), Robinson (1995a) established the asymptotic normality of the LP estimator. Subsequently, Hurvich et al. (1998) (hereafter HDB) computed the mean square error of the LP estimator and provided an MSE-optimal rule for bandwidth selection.

The LP estimator has undoubted appeal. It is easy to implement in practice and has been commonly employed in applications. However, when the spectral density of u_t dominates that of w_t in a neighborhood of the origin, the estimator may be biased downward substantially, especially in small samples. One source of the bias is the error of approximating the logarithm of $f^*(\lambda)$ by a constant in a shrinking neighborhood of the origin. This crude approximation also restricts the rate of convergence. The rate of convergence of the LP estimator will be shown to be $n^{-2d_0/(4d_0+1)}$, which is quite slow, especially when d_0 is close to zero.

To alleviate these problems, we take advantage of the structure of our model and propose to estimate the logarithm of $f^*(\lambda)$ locally by $c + \beta\lambda^{2d_0}$. Our new estimator is defined as the minimizer of the average-squared-errors (ASE) in a nonlinear log-periodogram (NLP) regression of the form

$$\log I_{zj} = \alpha - 2d \log \lambda_j + \beta\lambda_j^{2d} + error, \quad j = 1, 2, \ldots, m, \tag{3}$$

where

$$I_{zj} = I_z(\lambda_j) = \frac{1}{2\pi n}\left|\sum_{t=0}^{n-1} z_t \exp(\mathrm{i}t\lambda_j)\right|^2, \quad \lambda_j = \frac{2\pi j}{n}, \tag{4}$$

and m is a positive integer smaller than the sample size n. We will call our estimator NLP estimator hereafter.

Let $(\hat{d}, \hat{\beta})$ denote the NLP estimator that minimizes the concentrated ASE in which the intercept α has been concentrated out. We show the consistency of $\hat{d}$ by proving that the concentrated ASE converges uniformly over $(d, \beta)' \in \Theta$ to a function which has a unique minimizer d_0, where Θ is the parameter space to be defined later. To establish the asymptotic normality of $\hat{d}$, a typical argument would first establish the consistency of $\hat{\beta}$. But showing that $\hat{\beta}$ is consistent is not straightforward, because the concentrated ASE becomes flat as a function of β as $n \to \infty$. To circumvent this problem, we first show that $\hat{d}$ converges to d_0 at some rate k_n (meaning $\hat{d} - d_0 = O_p(k_n)$) without using the consistency of $\hat{\beta}$. We then show that, when $|d - d_0| \leqslant Ck_n$, the flatness problem disappears if the ASE is recentered and normalized by k_n^2.

We investigate both the asymptotic and finite sample properties of $\hat{d}$. Asymptotic bias, variance, asymptotic mean squared error (AMSE), and asymptotic normality are determined. We find that the asymptotic bias of $\hat{d}$ is of order m^{4d_0}/n^{4d_0}, provided that $f_w(\cdot)$ and $f_u(\cdot)$ are boundedly differentiable around the origin, whereas that of the LP estimator $\hat{d}_{\mathrm{LP}}$ has the larger order m^{2d_0}/n^{2d_0}. The asymptotic variances of $\hat{d}$ and $\hat{d}_{\mathrm{LP}}$ are both of order m^{-1}. In consequence, the optimal rate of convergence to zero of $\hat{d}$ is of order $n^{-4d_0/(8d_0+1)}$, whereas that of $\hat{d}_{\mathrm{LP}}$ is of the larger order $n^{-2d_0/(4d_0+1)}$. But when d_0 is close to zero, the rate of convergence of $\hat{d}$ will still be quite slow. We find that $\hat{d}$ is asymptotically normal with mean zero, provided that $m^{8d_0+1}/n^{8d_0} \to 0$, whereas $\hat{d}_{\mathrm{LP}}$ is asymptotically normal only under the more stringent condition $m^{4d_0+1}/n^{4d_0} \to 0$.

When the underlying process is a pure fractional process, we encounter a nonstandard estimation problem as the true parameter β_0 is on the boundary of the parameter set. In this case, the limiting distribution of $\hat{\beta}$ is truncated normal, and the limiting distribution of $\hat{d}$ is more complicated, involving a mixture distribution with the mixing probabilities that depend on the component distributions. We find that the asymptotic bias and variance of $\hat{d}$ are of the same orders as those of $\hat{d}_{\mathrm{LP}}$. However, it is difficult to obtain exact expressions for the asymptotic bias and variance. In consequence, it is hard to evaluate the performance of $\hat{d}$ relative to that of $\hat{d}_{\mathrm{LP}}$ in this case.

Some Monte Carlo simulations show that the asymptotic results of the paper capture the finite sample properties of the NLP estimator quite well. For the fractional component processes considered in the simulations, the NLP estimator $\hat{d}$ has a lower bias, a higher standard deviation, and a lower RMSE compared to the LP estimator $\hat{d}_{\mathrm{LP}}$, as the asymptotic results suggest. The lower bias leads to better coverage probabilities for $\hat{d}$ over a wider range of m than for $\hat{d}_{\mathrm{LP}}$. On the other hand, the lower standard deviation of $\hat{d}_{\mathrm{LP}}$ leads to shorter confidence intervals (CI) than CI based on $\hat{d}$.

The properties of the NLP estimator are investigated under the assumption of Gaussian errors. Gaussianity is usually assumed in the log-periodogram regression literature (e.g., Robinson, 1995a and Andrews and Guggenberger, 2003) and the present paper is no exception. Nevertheless, Gaussianity is restrictive in some empirical applications and could be relaxed following the lines of recent work by Velasco (2000) and Deo and Hurvich (2001), although we have not done so here.

This paper is related to the paper by Andrews and Guggenberger (2003). They considered the conventional fractional model (i.e., $var(u_t)=0$) and proposed to approximate $\log f_w(\lambda)$ by a constant plus a polynomial of even order. Andrews and Sun (2002) investigated the same issue in the context of a local Whittle estimator. The paper is also related to estimators introduced by Robinson and Henry (2003). They considered a general class of semiparametric M-estimators of d_0 that utilize higher-order kernels to obtain bias-reduction. As they stated, their results are heuristic in nature, whereas the results of this paper are established rigorously under specific regularity conditions. Other related papers include Henry and Robinson (1996), Hurvich and Deo (1999) and Henry (1999). These papers consider approximating $\log f^*(\lambda)$ by a more sophisticated function than a constant for the purpose of obtaining a data-driven choice of m. The present paper differs from those papers in that a nonlinear approximation is used in order to achieve bias reduction and to increase the rate of convergence in the estimation of d_0. Also, the nonlinear polynomial function used here depends on the memory parameter d_0 (whereas this is not so in the work just mentioned) and the estimation procedure for d_0 utilizes this information.

The rest of the paper is organized as follows. Section 2 formally defines the NLP estimator. Section 3 outlines the asymptotics of discrete Fourier transforms and log-periodogram ordinates, which are used extensively in later sections. Section 4 establishes consistency and derives limiting distribution results for the NLP estimator. This section also proposes a test for the pure fractional process against a perturbed fractional process. Section 5 investigates the finite sample performance of the NLP estimator by simulations. Section 6 concludes. Proofs are collected in Appendix A.

Throughout the paper, $\{E\}$ is defined to be the indicator function for event E. C is a generic constant.

2. NLP regression

This section motivates the NLP estimator that explicitly accounts for the additive perturbations in (1). Throughout, (1) is taken as the data generating process and then

$$f_z(\lambda) = \left(2\sin\frac{\lambda}{2}\right)^{-2d_0} f^*(\lambda). \tag{5}$$

Taking the logarithms of (5) leads to

$$\log(f_z(\lambda)) = -2d_0 \log\lambda + \log f^*(\lambda) - 2d_0 \log\left(2\lambda^{-1}\sin\left(\frac{\lambda}{2}\right)\right). \tag{6}$$

Replacing $f_z(\lambda)$ by periodogram ordinates $I_z(\lambda)$ evaluated at the fundamental frequencies $\lambda_j,\ j=1,2,\dots,m$ yields

$$\log(I_{zj}) = -c_0 - 2d_0 \log\lambda_j + \log f^*(\lambda_j) + U_j + \mathrm{O}(\lambda_j^2), \tag{7}$$

where $c_0 = 0.577216\ldots$ is the Euler constant and $U_j = \log[I_z(\lambda_j)/f_z(\lambda_j)] + c_0$.

By virtue of the continuity of $f^*(\lambda)$, we can approximate $\log f^*(\lambda_j)$ by a constant over a shrinking neighborhood of the zero frequency. This motivates log-periodogram regression on the equation

$$\log(I_{zj}) = \text{constant} - 2d\log\lambda_j + \mathit{error}. \tag{8}$$

The LP estimator $\hat{d}_{\mathrm{LP}}$ is then given by the least squares estimator of d in this regression. If $\{U_j\}_{j=1}^m$ behave asymptotically like independent and identically distributed random variables, then the LP estimator is a reasonable choice. In fact, under assumptions to be stated below, we establish that $\sqrt{m}(\hat{d}_{\mathrm{LP}} - d_0) \sim \mathrm{N}(b_{\mathrm{LP}}, \pi^2/24)$ where $b_{\mathrm{LP}} = \mathrm{O}(m^{2d_0+1/2}/n^{2d_0})$ and '$\sim$' signifies 'asymptotically distributed'. The 'asymptotic bias' of $\hat{d}_{\mathrm{LP}}$ itself is therefore of order $\mathrm{O}(m^{2d_0}/n^{2d_0})$, which can be quite large. To reduce the bias, we can approximate $\log f^*(\lambda_j)$ by a simple nonlinear function of frequency under the following assumptions:

Assumption 1. Either (a) $\sigma_u = var^{1/2}(u_t) = 0$ for all t, so $f_u(\lambda) \equiv 0$, for $\lambda \in [-\pi,\pi]$ or: (b) $\sigma_u > 0$ and $f_u(\lambda)$ is continuous on $[-\pi,\pi]$, bounded above and away from zero with bounded first derivative in a neighborhood of zero.

Assumption 2. $f_w(\lambda)$ is continuous on $[-\pi,\pi]$, bounded above and away from zero. When $\sigma_u = 0$, $f_w(\lambda)$ is three times differentiable with bounded third derivative in a neighborhood of zero. When $\sigma_u > 0$, $f_w(\lambda)$ is differentiable with bounded derivative in a neighborhood of zero.

Assumptions 1(b) and 2 are local smoothness conditions and hold for many models in current use, including ARMA models. They allow us to develop a Taylor expansion of $\log f^*(\lambda)$ about $\lambda=0$ with an error of the order of the first omitted term. Specifically, when $\sigma_u = 0$,

$$\log f^*(\lambda_j) = \log f_w(0) + \mathrm{O}(\lambda_j^2). \tag{9}$$

When $\sigma_u > 0$,

$$\begin{aligned}
\log f^*(\lambda_j) &= \log f_w(\lambda_j) + \log\left[1 + \left(2\sin\frac{\lambda_j}{2}\right)^{2d_0} \frac{f_u(\lambda_j)}{f_w(\lambda_j)}\right] \\
&= \log f_w(\lambda_j) + \log\left\{1 + \lambda_j^{2d_0}(1+\mathrm{O}(\lambda_j^2))\left(\frac{f_u(0)}{f_w(0)} + \mathrm{O}(\lambda_j^2)\right)\right\} \\
&= \log f_w(0) + \frac{f_u(0)}{f_w(0)}\lambda_j^{2d_0} + \mathrm{O}(\lambda_j^{4d_0}).
\end{aligned} \tag{10}$$

So, in either case

$$\log f^*(\lambda_j) = \log f_w(0) + \frac{f_u(0)}{f_w(0)}\lambda_j^{2d_0} + O(\lambda_j^r) \tag{11}$$

where $O(\cdot)$ holds uniformly over $j = 1, 2, \ldots, m$, $r = 4d_0\{\sigma_u > 0\} + 2\{\sigma_u = 0\}$.

Combining (7) with (11) produces the NLP regression model:

$$\log(I_{zj}) = -2d_0 \log \lambda_j + \alpha_0 + \lambda_j^{2d_0}\beta_0 + U_j + \varepsilon_j, \tag{12}$$

where

$$\alpha_0 = \log f_w(0) - c_0, \qquad \beta_0 = f_u(0)/f_w(0),$$

and

$$\varepsilon_j = \log f^*(\lambda_j) - \log f_w(0) - \beta_0\lambda_j^{2d_0} - 2d_0\left[\log\left(2\sin\frac{\lambda_j}{2}\right) - \log\lambda_j\right]. \tag{13}$$

The NLP estimator is then defined as the minimizer of the average-squared-errors in this model, i.e.

$$(\hat{\alpha}, \hat{d}, \hat{\beta}) = \arg\min_{\alpha, d, \beta} ASE(\alpha, d, \beta), \tag{14}$$

where

$$ASE(\alpha, d, \beta) = \frac{1}{m}\sum_{j=1}^{m}[\log(I_{zj}) - \alpha + 2d\log\lambda_j - \lambda_j^{2d}\beta]^2. \tag{15}$$

Concentrating (15) with respect to α, we obtain

$$(\hat{d}, \hat{\beta}) = \arg\min_{d\in D, \beta\in B} Q(d, \beta), \tag{16}$$

with

$$Q(d,\beta) = \frac{1}{m}\sum_{j=1}^{m}\left\{\left(\log I_{zj} - \frac{1}{m}\sum_{k=1}^{m}\log I_{zk}\right)\right.$$
$$\left. + 2d\left(\log\lambda_j - \frac{1}{m}\sum_{k=1}^{m}\log\lambda_k\right) - \beta\left(\lambda_j^{2d} - \frac{1}{m}\sum_{k=1}^{m}\lambda_k^{2d}\right)\right\}^2, \tag{17}$$

where B and D are parameter sets. We write $\theta = (d, \beta)'$, $\Theta = D \times B$ for convenience and make the following assumption on the parameter space:

Assumption 3. (a) $D = [d_1, d_2]$ where $0 < d_1 < d_2 < \frac{1}{2}$ and $B = [0, \bar{b}]$ where $\bar{b} > 0$;
(b) The true parameter $(d_0, \beta_0) \in (d_1, d_2) \times [0, \bar{b})$.

In the above assumption, d_1 and d_2 can be chosen arbitrarily close to 0 and $\frac{1}{2}$, respectively, and $\bar{b}$ can be chosen arbitrarily large. When $\beta_0 = 0$, the model becomes nonstandard in the sense that the true parameter is on the boundary of the parameter set. Section 4 explores the implication of the boundary problem.

3. LP asymptotics and useful lemmas

To establish the asymptotic properties of the NLP estimator, we need to characterize the asymptotic behavior of the LP ordinates $U_j = \log[I_z(\lambda_j)/f_z(\lambda_j)] + c_0$. Define

$$A_{zj} = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} z_t \cos \lambda_j t \quad \text{and} \quad B_{zj} = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} z_t \sin \lambda_j t, \tag{18}$$

then

$$U_j = \ln\left(\frac{A_{zj}^2}{f_{zj}} + \frac{B_{zj}^2}{f_{zj}}\right) + c_0, \quad j = 1, \ldots, m. \tag{19}$$

In view of the Gaussianity of A_{zj} and B_{zj}, we can evaluate the means, variances, and covariances of U_j, if the asymptotic behavior of the vector $(A_{zj}/f_{zj}^{1/2}, B_{zj}/f_{zj}^{1/2}, A_{zk}/f_{zk}^{1/2}, B_{zk}/f_{zk}^{1/2})$ is known. The properties of this vector depend in turn on those of the discrete Fourier transforms of z_t, defined as $w(\lambda) = (2\pi n)^{-1/2} \sum_1^n z_t \mathrm{e}^{\mathrm{i}t\lambda}$.

The asymptotic behavior of $w(\lambda)$ is given in the following lemma, which is a variant of results given earlier by several other authors (Robinson, 1995a, HDB, 1998, Andrews and Guggenberger, 2003).

Lemma 1. *Let Assumptions* 1 *and* 2 *hold. Then, uniformly over* j *and* k, $1 \leqslant k < j \leqslant m$, $m/n \to 0$,

(a) $E[w(\lambda_j)\bar{w}(\lambda_j)/f_z(\lambda_j)] = 1 + \mathrm{O}(j^{-1} \log j)$,
(b) $E[w(\lambda_j)w(\lambda_j)/f_z(\lambda_j)] = \mathrm{O}(j^{-1} \log j)$,
(c) $E[w(\lambda_j)\bar{w}(\lambda_k)/(f_z(\lambda_j)f_z(\lambda_k))^{1/2}] = \mathrm{O}(k^{-1} \log j)$,
(d) $E[w(\lambda_j)w(\lambda_k)/(f_z(\lambda_j)f_z(\lambda_k))^{1/2}] = \mathrm{O}(k^{-1} \log j)$.

It follows directly from Lemma 1 that for $1 \leqslant k < j \leqslant m$,

$$EA_{zj}^2/f_{zj} = \frac{1}{2} + \mathrm{O}\left(\frac{\log j}{j}\right), \quad EB_{zj}^2/f_{zj} = \frac{1}{2} + \mathrm{O}\left(\frac{\log j}{j}\right),$$
$$EA_{zj}B_{zj}/f_{zj} = \mathrm{O}\left(\frac{\log j}{j}\right), \quad EA_{zj}B_{zk}/(f_{zj}f_{zk})^{1/2} = \mathrm{O}\left(\frac{\log j}{k}\right). \tag{20}$$

Using these results and following the same line of derivation as in HDB (1998), we can prove Lemma 2 below. Since the four parts of this lemma are proved in a similar way to Lemmas 3, 5, 6 and 7 in HDB, the proofs are omitted here.

Lemma 2. *Let Assumptions* 1 *and* 2 *hold. Then*

(a) $Cov(U_j, U_k) = \mathrm{O}(\log^2 j/k^2)$, *uniformly for* $\log^2 m \leqslant k < j \leqslant m$,
(b) $\lim_n \sup_{1 \leqslant j \leqslant m} EU_j^2 < \infty$,
(c) $E(U_j) = \mathrm{O}(\log j/j)$, *uniformly for* $\log^2 m \leqslant j \leqslant m$,
(d) $Var(U_j) = \pi^2/6 + \mathrm{O}(\log j/j)$, *uniformly for* $\log^2 m \leqslant j \leqslant m$.

With the asymptotic behavior of U_j in hand, we can proceed to show that the normalized sums $m^{-1}\sum_{j=1}^{m} c_j U_j$ are uniformly negligible under certain conditions on the coefficients c_j. Quantities of this form appear in the normalized Hessian matrix below.

Lemma 3. *Let $\{c_j(d,\beta)\}_{j=1}^{m}$ be a sequence of functions such that, for some $p \geqslant 0$,*

$$\sup_{(d,\beta)' \in \Theta} |c_j| = \mathrm{O}(\log^p m) \text{ uniformly for } 1 \leqslant j \leqslant m, \tag{21}$$

and for some $q \geqslant 0$,

$$\sup_{(d,\beta)' \in \Theta} |c_j - c_{j-1}| = \mathrm{O}(j^{-1}\log^q m) \text{ uniformly for } 1 \leqslant j \leqslant m. \tag{22}$$

Then

$$\sup_{(d,\beta)' \in \Theta} \left| \frac{1}{m} \sum_{j=1}^{m} c_j U_j \right| = \mathrm{O}_\mathrm{p}\left(\frac{\log^{\max(p,q)} m}{\sqrt{m}} \right). \tag{23}$$

We can impose additional conditions to get a tighter bound. For example, if we also require that $\sup_{(d,\beta)' \in \Theta} |c_m| = \mathrm{O}(1)$, then $\sup_{(d,\beta)' \in \Theta} |(1/m)\sum_{j=1}^{m} c_j U_j| = \mathrm{O}_\mathrm{p}(\log^q m/\sqrt{m})$, as is readily seen from the proof of the lemma. Further, the lemma remains valid if we remove the 'sup' operator from both the conditions and the conclusion.

Let $V_j(d,\beta) = 2(d-d_0)\log\lambda_j - \beta\lambda_j^{2d} + \beta_0\lambda_j^{2d_0}$, and $\bar{V}(d,\beta) = 1/m\sum_{j=1}^{m} V_j(d,\beta)$. We can use an argument similar to the proof of Lemma 3 to establish the following corollary. The proof is omitted.

Corollary 1. *Let $D^0 = \{d\colon d \in D,\ |d-d_0| \leqslant C(m/n)^\gamma\}$, for some constants $C \in R^+$, $\gamma \in [0, 2d_0]$, then*

$$\sup_{(d,\beta)' \in D^0 \times B} \left| \frac{1}{m}\sum_{j=1}^{m} U_j(V_j(d,\beta) - \bar{V}(d,\beta)) \right| = \mathrm{O}_\mathrm{p}\left(\left(\frac{m}{n}\right)^{\gamma} \frac{1}{\sqrt{m}} \right) \tag{24}$$

as $1/n + m/n \to 0$.

The following lemma assists in establishing the asymptotic normality of the NLP regression estimator.

Lemma 4. *Let $a_{kn} = a_k$ be a triangular array for which*

$$\max_k |a_k| = \mathrm{o}(m), \qquad \sum_{k=[1+m^{0.5+\delta}]}^{m} a_k^2 \sim \rho m, \qquad \sum_{k=[1+m^{0.5+\delta}]}^{m} |a_k|^p = \mathrm{O}(m), \tag{25}$$

for all $p \geqslant 1$, and $0 < \delta < 0.5$. Then,

$$\frac{1}{\sqrt{m}} \sum_{k=[1+m^{0.5+\delta}]}^{m} a_k U_k \xrightarrow{d} \mathrm{N}\left(0, \frac{\pi^2}{6}\rho\right), \tag{26}$$

where $[\,\cdot\,]$ denotes the integer part.

The proof of this lemma is based on the method of moments and involves a careful exploration of the dependence structure of the discrete Fourier transforms. Robinson's argument (1995a, pp. 1067–1070) forms the basis of this development and can be used here with some minor modifications to account for differences in the models. Details are omitted here and are available upon request.

4. Asymptotic properties of the LP and NLP estimators

4.1. Asymptotic properties of the LP estimator

We establish the asymptotic properties for the LP estimator in the context of the components model (1). Theorem 1 gives the limit theory and provides a benchmark for later comparisons.

Theorem 1. *Let Assumptions* 1 *and* 2 *hold. Let* $m = m(n) \to \infty$ *and*

$$\frac{m^{r'+1/2}}{n^{r'}} \to K'_\sigma\{\sigma_u > 0\} + K'_0\{\sigma_u = 0\} \tag{27}$$

as $n \to \infty$, *where* $r' = 2d_0\{\sigma_u > 0\} + 2\{\sigma_u = 0\}$ *and* $K'_\sigma, K'_0 > 0$ *are positive constants. Then*

$$\sqrt{m}(\hat{d}_{\mathrm{LP}} - d_0) \;\Rightarrow\; \mathrm{N}\left(b_{\mathrm{LP}}, \frac{\pi^2}{24}\right), \tag{28}$$

where

$$\begin{aligned} b_{\mathrm{LP}} = &-(2\pi)^{2d_0}\frac{f_u(0)}{f_w(0)}\frac{d_0}{(2d_0+1)^2}K'_\sigma\{\sigma_u > 0\} \\ &- \frac{2\pi^2}{9}\left(\frac{f''_w(0)}{f_w(0)} + \frac{d_0}{6}\right)K'_0\{\sigma_u = 0\}. \end{aligned} \tag{29}$$

When $\sigma_u > 0$, the ratio $m^{r'+1/2}/n^{r'} = m^{2d_0+1/2}/n^{2d_0} \to K'_\sigma$ in (27). This delivers an upper bound of order $\mathrm{O}(n^{4d_0/(1+4d_0)})$ on the rate at which m can increase with n and allows for larger choices of m for larger values of d_0. Intuitively, as d_0 increases, the contamination from perturbations at frequencies away from the origin becomes relatively smaller and we can expect to be able to employ a wider bandwidth in the regression. To eliminate the asymptotic bias b_{LP} in (28) altogether, we use a narrower band and set $m = \mathrm{o}(n^{4d_0/(1+4d_0)})$ in place of (27). Deo and Hurvich (2001) established a similar result under the assumption that u_t is iid, but not necessarily Gaussian. Their assumption that $m^{4d_0+1}\log^2 m/n^{4d_0} = \mathrm{o}(1)$ is slightly stronger than the assumption made here.

When $\sigma_u > 0$, the limit distribution (28) involves the bias

$$b_{\mathrm{LP}} = -(2\pi)^{2d_0}\frac{f_u(0)}{f_w(0)}\frac{d_0}{(2d_0+1)^2}K'_\sigma < 0, \tag{30}$$

which is always negative, as one would expect, because of the effect of the short memory perturbations. Correspondingly, the dominating bias term of $\hat{d}_{\mathrm{LP}}$ has the form

$$b_{n,\mathrm{LP}} = -(2\pi)^{2d_0} \frac{f_u(0)}{f_w(0)} \frac{d_0}{(2d_0+1)^2} \frac{m^{2d_0}}{n^{2d_0}} < 0. \tag{31}$$

The magnitude of the bias obviously depends on the quantity $f_w(0)/f_u(0)$, which is the ratio of the long run variance of the short memory input of y_t to that of the perturbation component u_t. The ratio can be interpreted as a long run signal–noise ratio (SNR), measuring the strength in the long run of the signal from the y_t inputs relative to the long run signal in the perturbations. The stronger the long run signal in the perturbations, the greater the downward bias and the more difficult it becomes to estimate the memory parameter accurately. One might expect these effects to be exaggerated in small samples where the capacity of the data to discriminate between long run and short run effects is reduced.

When $\sigma_u > 0$, the AMSE of $\hat{d}_{\mathrm{LP}}$ satisfies

$$AMSE(\hat{d}_{\mathrm{LP}}) = \mathrm{O_p}\left(\frac{m}{n}\right)^{4d_0} + \mathrm{O_p}\left(\frac{1}{m}\right). \tag{32}$$

So the AMSE-optimal bandwidth has the form $m_{\mathrm{LP}}^{\mathrm{opt}} = C_{\mathrm{LP}} n^{4d_0/(4d_0+1)}$ for some constant C_{LP}. When $m = m_{\mathrm{LP}}^{\mathrm{opt}}$, $AMSE(\hat{d}_{\mathrm{LP}}) = \mathrm{O_p}(n^{-4d_0/(4d_0+1)})$. In contrast, in the case $\sigma_u = 0$, it is well known that when $m = m_{\mathrm{LP}}^{\mathrm{opt}}$, $AMSE(\hat{d}_{\mathrm{LP}}) = \mathrm{O_p}(n^{-4/5})$. Due to the presence of the perturbations, the optimal AMSE of $\hat{d}_{\mathrm{LP}}$ converges to zero at a slower rate.

When $\sigma_u=0$, the theorem contains essentially the same results proved in HDB. In this case, the dominating bias of $\hat{d}_{\mathrm{LP}}$ is given by $b_{n,\mathrm{LP}} = -2\pi^2/9(f_w''(0)f_w^{-1}(0)+d_0/6)m^2/n^2$. HDB showed that the dominating bias of $\hat{d}_{\mathrm{LP}}$ in the case of pure fractional process regression is given by the expression $-2\pi^2/9(f_w''(0)f_w^{-1}(0))m^2/n^2$. The presence of the additional factor $d_0/6$ in the second term of our expression arises from the use of a slightly different regressor in the LP regression. In particular, we employ $-2\log\lambda_j$ as one of the regressors in (3), while HDB use $-2\log(2\sin\lambda_j/2)$. These regressors are normally considered to be asymptotically equivalent. However, while the use of $-2\log\lambda_j$ rather than $-2\log(2\sin\lambda_j/2)$ has no effect on the asymptotic variance, it does affect the asymptotic bias.

4.2. Consistency of the NLP estimator

To establish the limiting distribution of the NLP estimator, we first prove the consistency of the NLP estimator.

Theorem 2. *Let Assumptions* 1 *and* 2 *hold.*

(a) *If* $1/m + m/n \to 0$ *as* $m, n \to \infty$, *then* $\hat{d} - d_0 = \mathrm{o_p}(1)$.
(b) *If for some arbitrary small* $\Delta > 0, m/n + n^{4d_0(1+\Delta)}/m^{4d_0(1+\Delta)+1} \to 0$, *as* $m, n \to \infty$, *then* $\hat{d} - d_0 = \mathrm{O_p}((m/n)^{2d_0})$ *and* $\hat{\beta} - \beta_0 = \mathrm{o_p}(1)$.

Theorem 2 shows that $\hat{d}$ is consistent under mild conditions. All that is needed is that m approaches infinity slower than the sample size n. As shown by HDB, trimming out low frequencies is not necessary. This point is particularly important in the present case because, in seeking to reduce contamination from the perturbations, the lowest frequency ordinates are the most valuable in detecting the long memory effects.

It is not straightforward to establish the consistency of $\hat{\beta}$, because, as $n \to \infty$, the objective function becomes flat as a function of β. The way we proceed is, in fact, to show first that $\hat{d}$ converges to d_0 at some slower rate, more precisely, $\hat{d}-d_0=\mathrm{O_p}((m/n)^{2d_0})$. We prove this rate of convergence stepwise. We start by showing that $\hat{d}-d_0=\mathrm{o_p}((m/n)^{d_1/2})$ for $0<d_1<d_0$, using the fact that $\beta\lambda_j^{2d}=\mathrm{O}(m/n)^{2d_1}$ uniformly in $(d,\beta)'\in\Theta$. We can then deduce that $\hat{d}-d_0=\mathrm{o_p}((m/n)^{d_0(1+\Delta)})$. With this faster rate of convergence, we have better control over some quantities and can obtain an even faster rate of convergence for $\hat{d}$. Repeating this procedure leads to $\hat{d}-d_0=\mathrm{O_p}((m/n)^{2d_0})$, as desired. With this result, we observe that $(m/n)^{-4d_0}(Q(d,\beta)-Q(d_0,\beta_0))$ is no longer flat as a function of β for any value of d such that $|d-d_0|\leqslant C(m/n)^{2d_0}$. This observation can be readily seen from the proof of the theorem. This approach to overcoming the problem of apparent flatness in the objective function is likely to be applicable in other nonlinear estimation contexts when the involved variables are integrated of different orders or have different stochastic orders.

We also prove the rate of convergence of $\hat{d}$ without using the consistency of $\hat{\beta}$. This is unusual because in most nonlinear estimation problems it is common to prove the consistency of all parameters first in order to establish rates of convergence. The approach is successful in the present case because when d is close to d_0, the regressor λ_j^{2d} evaporates as $n\to\infty$ and approaches zero approximately at the rate of $(m/n)^{2d_0}$.

4.3. Asymptotic distribution of the NLP estimator

The asymptotic distribution of the NLP estimator depends on whether β_0 is on the boundary of the parameter set. In this section, we first establish the asymptotic properties of the gradient and Hessian functions. These asymptotic results hold for any value of $\beta_0\in B$. Using these results, we then investigate the asymptotic distribution of the NLP estimator for the cases $0<\beta_0<\bar{b}$ and $\beta_0=0$, respectively. We impose a somewhat stronger assumption:

Assumption 4. $n^{4d_0(1+\Delta)}/m^{4d_0(1+\Delta)+1}\to 0$ for some arbitrary small $\Delta>0$ and

$$m^{r+1/2}/n^r \to K_\sigma\{\sigma_u>0\}+K_0\{\sigma_u=0\} \tag{33}$$

as $m,n\to\infty$, where $r=4d_0\{\sigma_u>0\}+2\{\sigma_u=0\}$.

The two conditions in Assumption 4 are always compatible because $r\geqslant 4d_0$ and Δ is arbitrarily small. The lower bound on the growth rate of m ensures the consistency of $\hat{d}$ and $\hat{\beta}$. The upper bound on the growth rate of m guarantees that the normalized gradient of $Q(d,\beta)$ is $O_p(1)$, which is required for deriving the asymptotic distribution of $(\hat{d},\hat{\beta})$.

When $\sigma_u = 0$, the upper bound becomes $m = O(n^{4/5})$, which is the same as the upper bound for asymptotic normality of the LP estimator for a pure fractional process. When $\sigma_u > 0$, the upper bound becomes $m^{8d_0+1}/n^{8d_0} = O(1)$, which is less stringent than the upper bound given in Theorem 1. It therefore allows us to take m larger than in conventional LP regression applied to the fractional components model. In consequence, by an appropriate choice of m, we have asymptotic normality for $\hat{d}$ with a faster rate of convergence than is possible in LP regression. However, for any $0 < d_0 < \frac{1}{2}$, the upper bound is more stringent than $m = O(n^{4/5})$, the upper bound for asymptotic normality of LP regression in a pure fractional process model. Hence, the existence of the weakly dependent perturbations in (1) requires the use of a narrower bandwidth than LP regression for a pure fractional process. Interestingly, as d_0 approaches $\frac{1}{2}$, the upper bound becomes arbitrarily close to $m = O(n^{4/5})$.

We now proceed to establish the asymptotic distribution of the NLP estimator. The consistency result and Assumption 3 ensure that we only need to consider the constraint $\beta \geqslant 0$. Therefore, the first order conditions for (16) are:

$$S_n(d,\beta) = (0,\Lambda)', \tag{34}$$

$$\Lambda\beta = 0, \tag{35}$$

where Λ is the Lagrangian multiplier for the constraint $\beta \geqslant 0$,

$$S_n(d,\beta) = -\sum_{j=1}^{m} \begin{pmatrix} x_{1j}(d,\beta) - \bar{x}_1(d,\beta) \\ x_{2j}(d,\beta) - \bar{x}_2(d,\beta) \end{pmatrix} e_j(d,\beta), \tag{36}$$

$$x_{1j}(d,\beta) = -2\log\lambda_j(1-\beta\lambda_j^{2d}), \quad \bar{x}_1(d,\beta) = \frac{1}{m}\sum_{k=1}^{m} x_{1k},$$

$$x_{2j}(d,\beta) = \lambda_j^{2d}, \quad \bar{x}_2(d,\beta) = \frac{1}{m}\sum_{k=1}^{m} x_{2k}, \tag{37}$$

and

$$e_j(d,\beta) = \log I_{zj} - \frac{1}{m}\sum_{k=1}^{m}\log I_{zk} + 2d\left(\log\lambda_j - \frac{1}{m}\sum_{k=1}^{m}\log\lambda_k\right) \\ - \beta(x_{2j}(d,\beta) - \bar{x}_2(d,\beta)). \tag{38}$$

Expanding $S_n(\hat{d},\hat{\beta})$ about $S_n(d_0,\beta_0)$, we have

$$(0,\hat{\Lambda})' = S_n(d_0,\beta_0) + H_n(d^*,\beta^*)(\hat{d}-d_0,\hat{\beta}-\beta_0)', \tag{39}$$

where $H_n(d,\beta)$ is the Hessian matrix, (d^*,β^*) is between (d_0,β_0) and $(\hat{d},\hat{\beta})$. The elements of the Hessian matrix are:

$$
\begin{aligned}
H_{n,11}(d,\beta) &= \sum_{j=1}^{m}(x_{1j}-\bar{x}_1)^2 - \beta\sum_{j=1}^{m} e_j(\log\lambda_j^2)^2\lambda_j^{2d},\\
H_{n,12}(d,\beta) &= \sum_{j=1}^{m}(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2) - \sum_{j=1}^{m} e_j(\log\lambda_j^2)\lambda_j^{2d},\\
H_{n,22}(d,\beta) &= \sum_{j=1}^{m}(x_{2j}-\bar{x}_2)^2.
\end{aligned}
\tag{40}
$$

Define the diagonal matrix $D_n = diag(\sqrt{m}, \lambda_m^{2d_0}\sqrt{m})$. We show in the following lemma that the normalized Hessian $D_n^{-1}H_n(d_0,\beta_0)D_n^{-1}$ converges in probability to a 2×2 matrix defined by

$$
\Omega = \begin{pmatrix} 4 & -4d_0/(2d_0+1)^2 \\ -4d_0/(2d_0+1)^2 & 4d_0^2/((4d_0+1)(2d_0+1)^2) \end{pmatrix}, \tag{41}
$$

and the ‘asymptotic bias’ of the normalized score $D_n^{-1}S_n(d_0,\beta_0)$ is $-b$, where

$$
b = \{\sigma_u > 0\}b_\sigma + \{\sigma_u = 0\}b_0, \tag{42}
$$

and

$$
\begin{aligned}
b_\sigma &= \frac{(2\pi)^{4d_0} f_w^2(0)K_\sigma}{2f_u^2(0)}\left(\frac{8d_0}{(4d_0+1)^2}, -\frac{8d_0^2}{(2d_0+1)(4d_0+1)(6d_0+1)}\right)',\\
b_0 &= (2\pi)^2K_0\left(\frac{f_w''(0)}{f_w(0)}+\frac{d_0}{6}\right)\left(-\frac{2}{9}, \frac{2d_0}{3(2d_0+3)(2d_0+1)}\right)'.
\end{aligned}
\tag{43}
$$

Before stating the lemma, we need the following notation. Let $J_n(d,\beta)$ be a 2×2 matrix whose (i,j)th element is

$$
J_{n,ij} = \sum_{k=1}^{m}(x_{ik}(d,\beta)-\bar{x}_i(d,\beta))(x_{jk}(d,\beta)-\bar{x}_j(d,\beta)), \tag{44}
$$

and let Θ_n be a set defined by

$$
\Theta_n = \{(d,\beta)' \colon |\lambda_m^{-d_0}(d-d_0)| < \varepsilon \text{ and } |\beta-\beta_0| < \varepsilon\}. \tag{45}
$$

Lemma 5. *Let Assumptions* 1–4 *hold. Then*

(a) $\sup_{(d,\beta)'\in\Theta_n}\|D_n^{-1}(H_n(d,\beta)-J_n(d,\beta))D_n^{-1}\| = o_p(1)$,
(b) $\sup_{(d,\beta)'\in\Theta_n}\|D_n^{-1}[J_n(d,\beta)-J_n(d_0,\beta_0)]D_n^{-1}\| = o_p(1)$,
(c) $D_n^{-1}J_n(d_0,\beta_0)D_n^{-1} \to \Omega$,
(d) $D_n^{-1}S_n(d_0,\beta_0) \Rightarrow N(-b, \frac{\pi^2}{6}\Omega)$.

We now consider the asymptotic distribution when $\sigma_u > 0$. In this case, the true parameter $(d_0, \beta_0)'$ is an interior point of the parameter space. Hence $\hat{\Lambda} = 0$. A Taylor expansion of the first order condition (34) combined with the convergence results of Lemma 5 gives the following Theorem:

Theorem 3. *Let Assumptions* 1–4 *hold. If* $\sigma_u > 0$, *then*

$$D_n \begin{pmatrix} \hat{d} - d_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} \Rightarrow \mathrm{N}\left(b_{\mathrm{NLP}}, \frac{\pi^2}{6}\, \Omega^{-1} \right) \tag{46}$$

where $b_{\mathrm{NLP}} = \Omega^{-1} b_\sigma$ *and*

$$\Omega^{-1} = \begin{pmatrix} (2d_0+1)^2/(16d_0^2) & (2d_0+1)^2(4d_0+1)/(16d_0^3) \\ (2d_0+1)^2(4d_0+1)/(16d_0^3) & (4d_0+1)(2d_0+1)^4/(16d_0^4) \end{pmatrix}. \tag{47}$$

Remark 1. From the above theorem, we deduce immediately that when $\sigma_u > 0$, the asymptotic variance of $\sqrt{m}(\hat{d} - d_0)$ is $\pi^2 C_d/24$, where $C_d = 1 + (4d_0+1)/(4d_0^2) > 1$. Approximating $\log f^*(\cdot)$ locally by a nonlinear function instead of a constant therefore inflates the usual asymptotic variance of the LP regression estimator in a pure fractional model by the factor C_d. This is to be expected, as adding more variables in regression usually inflates variances.

Remark 2. When $\sigma_u > 0$, the limiting distribution (46) involves the bias b_{NLP}. The dominating bias term of $(\hat{d}, \hat{\beta})'$ is thus equal to

$$D_n^{-1}\Omega^{-1} b_n = -\frac{(2\pi)^{4d_0} f_w^2(0)}{f_u^2(0)} \left(\frac{m}{n}\right)^{4d_0} \begin{pmatrix} d_0(2d_0+1)/((4d_0+1)^2(6d_0+1)) \\ 2(2d_0+1)^2/((4d_0+1)(6d_0+1)) \end{pmatrix}. \tag{48}$$

Remark 3. When $\sigma_u > 0$, according to (48) the asymptotic bias of $\hat{d}$ is of order m^{4d_0}/n^{4d_0}. In contrast, the asymptotic bias of the LP estimator is of order m^{2d_0}/n^{2d_0}, as shown above in (31). The asymptotic bias of the NLP estimator is therefore smaller than that of the LP estimator by order m^{2d_0}/n^{2d_0}.

Remark 4. Following the previous remarks, the AMSE of $\hat{d}$ has the form $AMSE(\hat{d}) = K^2(m/n)^{8d_0} + \pi^2 C_d/(24m)$, where

$$K = (2\pi)^{4d_0} \beta_0^2 \frac{d_0(2d_0+1)}{(4d_0+1)^2(6d_0+1)}. \tag{49}$$

Straightforward calculations yield the value of m that minimizes $AMSE(\hat{d})$, viz.

$$m^{\mathrm{opt}} = \left[\left(\frac{\pi^2 C_d}{192 d_0 K^2} \right)^{1/(8d_0+1)} n^{8d_0/(8d_0+1)} \right], \tag{50}$$

where $[\,\cdot\,]$ denotes the integer part. When $m = m^{\mathrm{opt}}$, the AMSE of $\hat{d}$ converges to zero at the rate of $n^{-8d_0/(8d_0+1)}$. In contrast, when $m = m_{\mathrm{LP}}^{\mathrm{opt}}$, the AMSE of $\hat{d}_{\mathrm{LP}}$ converges to zero only at the rate of $n^{-4d_0/(4d_0+1)}$. Thus, the optimal AMSE of $\hat{d}$ converges faster to zero than that of $\hat{d}_{\mathrm{LP}}$.

Remark 5. When d_0 is close to zero, the asymptotic bias of $\hat{d}$ is of order m^{4d_0}/n^{4d_0}, which is close to the order of the asymptotic bias of $\hat{d}_{\mathrm{LP}}$. In addition, when d_0 is close to zero, the asymptotic variance of the $\hat{d}$ will be large. Therefore, as d_0 approaches zero, the advantage of $\hat{d}$ over $\hat{d}_{\mathrm{LP}}$ diminishes. This is expected as when d_0 is close to zero, the downward bias of $\hat{d}_{\mathrm{LP}}$ will be small and there is not much scope for $\hat{d}$ to manifest its bias-reducing capacity.

Remark 6. $\hat{\beta}$ converges more slowly by a rate of $(m/n)^{-2d_0}$ than $\hat{d}$. Heuristically, the information contents of the two regressors ($\log \lambda_j$ and $\lambda_j^{2d_0}$) are different. We have $\sum_{j=1}^{m} (\log \lambda_j - \sum_{k=1}^{m} \log \lambda_k/m)^2 = \mathrm{O}(m)$ whereas $\sum_{j=1}^{m} (\lambda_j^{2d_0} - \sum_{k=1}^{m} \lambda_k^{2d_0}/m)^2 = \mathrm{O}(m\lambda_m^{2d_0})$. The sums of the squares thus differ by an order of magnitude $\lambda_m^{2d_0}$. This difference leads to the slower rate of convergence of $\hat{\beta}$.

Next, we consider the asymptotic distribution when $\sigma_u = 0$. In this case, the parameter β_0 lies on the boundary of the parameter space. As a consequence, Λ may not equal zero and we have a different limiting distribution.

Theorem 4. *Let Assumptions* 1–4 *hold. If* $\sigma_u = 0$, *then*

$$\sqrt{m}(\hat{d} - d_0) \Rightarrow -(\tilde{\Omega}_{11}\eta_1 + \tilde{\Omega}_{12}\eta_2)\{\tilde{\Omega}_{12}\eta_1 + \tilde{\Omega}_{22}\eta_2 \leqslant 0\} - \Omega_{11}^{-1}\eta_1\{\tilde{\Omega}_{12}\eta_1 + \tilde{\Omega}_{22}\eta_2 > 0\}, \tag{51}$$

$$\sqrt{m}\lambda_m^{2d_0}(\hat{\beta} - \beta_0) \Rightarrow -(\tilde{\Omega}_{12}\eta_1 + \tilde{\Omega}_{22}\eta_2)\{\tilde{\Omega}_{12}\eta_1 + \tilde{\Omega}_{22}\eta_2 \leqslant 0\}, \tag{52}$$

where $\tilde{\Omega} = (\tilde{\Omega}_{ij}) = \Omega^{-1}$, *and* $\eta = (\eta_1, \eta_2)' \sim \mathrm{N}(-b_0, \pi^2\Omega/6)$.

Remark 7. Theorem 4 shows that when the true parameter $\beta_0 = 0$, i.e. in the case of a pure fractional process, the limiting distribution of $\hat{\beta}$ is truncated normal. The truncation arises because the true parameter is on the boundary of the parameter set. Since the Hessian matrix is not diagonal, the truncation also affects the limiting distribution of $\hat{d}$, which becomes a mixture distribution with the mixing probabilities depending on the component distributions.

Remark 8. It follows easily from Theorem 4 that the dominating bias term of $\hat{d}$ is of order $\mathrm{O_p}(m^2/n^2)$, the same order as that of $\hat{d}_{\mathrm{LP}}$. Since its limiting distribution is a complicated function of normal random variables, it is not easy to derive an exact expression for the dominating bias term. Hence, it is quite difficult to compare the dominating bias term of $\hat{d}$ with that of $\hat{d}_{\mathrm{LP}}$ in this case.

The following corollary follows from Theorem 4 by using a narrower frequency band. The proof is straightforward and is omitted.

Corollary 2. *Let Assumptions* 1–3 *hold. If* $n^{4d_0(1+\Delta)}/m^{4d_0(1+\Delta)+1}+m^5/n^4 \to 0$ *for some arbitrary small* $\Delta > 0$, *then*

$$\sqrt{m}\lambda_m^{2d_0}(\hat{\beta}-\beta_0) \;\Rightarrow\; v(d_0)\tau\{\tau \geqslant 0\} \tag{53}$$

where $\tau \equiv \mathrm{N}(0,1)$ *and* $v(d_0) = \pi d_0^{-2}(2d_0+1)^2\sqrt{(4d_0+1)/96}$.

4.4. A test for perturbations

The properties of the LP and NLP estimators depend on whether short memory perturbations are present in the data. The limit theory can be used to construct a test for the presence of perturbations, which can be formulated in terms of the hypotheses

$$H_0 : \beta_0 = 0 \text{ vs. } H_1 : \beta_0 > 0,$$

with no perturbations under H_0, and with short memory perturbations present under H_1 whose intensity increases with β_0.

Using Corollary 2, we can construct the t-statistic

$$t_{\hat{\beta}} = \sqrt{m}\lambda_m^{2\hat{d}}\hat{\beta}/v(\hat{d}). \tag{54}$$

If Assumptions 1–3 hold and $n^{4d_0(1+\Delta)}/m^{4d_0(1+\Delta)+1} + m^5/n^4 \to 0$, then under the null hypothesis,

$$t_{\hat{\beta}} \;\Rightarrow\; \tau\{\tau \geqslant 0\}. \tag{55}$$

If Assumptions 1–3 hold and $n^{4d_0(1+\Delta)}/m^{4d_0(1+\Delta)+1} + m^{4d_0+1/2}/n^{4d_0} \to 0$, then under the alternative hypothesis that $\beta_0 = \beta_A > 0$, we have

$$\begin{aligned} t_{\hat{\beta}} &= \sqrt{m}\lambda_m^{2\hat{d}}(\hat{\beta}-\beta_A)/v(\hat{d}) + \sqrt{m}\lambda_m^{2\hat{d}}\beta_A/v(\hat{d}) \\ &= \mathrm{O_p}(1) + \sqrt{m}\lambda_m^{2\hat{d}}\beta_A/v(\hat{d}) \end{aligned} \tag{56}$$

by Theorem 3. Since $\sqrt{m}\lambda_m^{2\hat{d}} \to \infty$ when $n^{4d_0(1+\Delta)}/m^{4d_0(1+\Delta)+1} \to 0$, we deduce that $t_{\hat{\beta}} \to \infty$ in probability and the test is therefore consistent. Note that $m^5/n^4 \to 0$ implies that $m^{4d_0+1/2}/n^{4d_0} \to 0$. We collect the results in the following corollary:

Corollary 3. *If Assumptions* 1–3 *hold and* $n^{4d_0(1+\Delta)}/m^{4d_0(1+\Delta)+1} + m^5/n^4 \to 0$, *then*

$t_{\hat{\beta}} \Rightarrow \tau\{\tau \geqslant 0\}$ *under* H_0 *and* $t_{\hat{\beta}} \to \infty$ *in probability under* H_1.

Simulations, not reported here, indicate that for sample sizes less than 2048 the power of the test is quite low.

5. Simulations

5.1. Experimental design

This section investigates the finite sample performance of the NLP estimator in comparison with conventional LP regression. The chosen data generating process is

$$z_t = (1-L)^{-d_0}w_t + u_t, \tag{57}$$

where $\{w_t\colon t = \ldots, -1, 0, 1, \ldots, n\}$ are iid $\mathrm{N}(0,1)$, $\{u_t\colon t = \ldots, -1, 0, 1, \ldots, n\}$ are iid $\mathrm{N}(0, \sigma_u^2)$ and $\{w_t\}$ are independent of $\{u_t\}$. To simulate the fractional process $y_t = (1-L)^{-d_0} w_t = \sum_{k=0}^{\infty} \Gamma(d_0+k)\Gamma^{-1}(d_0)\Gamma^{-1}(k+1) w_{t-k}$, we truncate the infinite sum after M iterations. To be more specific, we generate y_t according to $\sum_{k=0}^{M} \Gamma(d_0 + k)\Gamma^{-1}(d_0)\Gamma^{-1}(k+1) w_{t-k}$ with $M = 5000$. To reduce the initialization effects, we generate a time series of length $2n$ and trim the first n observations to get the simulated sample.

We consider the following constellation of parameter combinations

$$d_0 = 0.25, 0.45, 0.65, 0.85 \tag{58}$$

and

$$\sigma_u^2 = 0, 4, 8, 16. \tag{59}$$

In view of the fact that the LP estimator is consistent for both stationary fractional processes ($d_0 < 0.5$) and nonstationary fractional processes ($0.5 \leqslant d_0 < 1$) (see Kim and Phillips, 2000), we expect the NLP estimator to work well for nonstationary fractional component processes for this range of values of d_0 as well as for stationary fractional component processes over ($0 < d_0 < 0.5$). Hence it is of interest to include some values of d_0 that fall in the nonstationary zone.

The value of σ_u^2 determines the strength of the noise from the perturbations. The long run SNR increases as σ_u^2 decreases. When $\sigma_u^2 = 0$, z_t is a pure fractional process with an infinite long-run SNR. The inverse of the long run SNR, viz. $f_u(0)/f_w(0)$, takes the values $0, 4, 8, 16$. These are close to the values in Deo and Hurvich (2001). In their simulation study, the ratio $f_u(0)/f_w(0)$ takes the values 6.17 and 13.37.

We consider sample sizes $n = 128$, 512, and 2048. For each sample size and parameter combination, 2000 replications are performed from which we calculate the biases, standard deviations and root mean square errors of $\hat{d}$ and $\hat{d}_{\mathrm{LP}}$, for different selections of the bandwidth m. Then, for each parameter combination, we graph each of these quantities as functions of m. The results are shown in panels (a)–(c) of Figs. 1–6.

In addition, we compute the coverage probabilities, as functions of m, of the nominal 90% CI that are obtained using the asymptotic normality results of Theorems 1 and 3. We calculate the average lengths of the CI as functions of m. For some data generating processes, the coverage probabilities and the average lengths are graphed against m in panels (d) and (e) of Figs. 1 and 2. When constructing the CI, we estimate the standard errors of $\hat{d}$ and $\hat{d}_{\mathrm{LP}}$ using finite sample expressions rather than the limit expressions, because the former yield better finite sample performance for all parameter combinations and for both estimators. Specifically, the standard error of $\hat{d}$ is estimated by $SE_{HJ} = SE_J + (SE_H - SE_J)\{H(\hat{d}, \hat{\beta}) > 0\}$ where $\{H(\hat{d}, \hat{\beta}) > 0\} = 1$ if $H(\hat{d}, \hat{\beta})$ is positive definite, and

$$SE_H = \pi/\sqrt{6} H_{22,n}^{1/2}(\hat{d}, \hat{\beta})(H_{11}(\hat{d}, \hat{\beta}) H_{22}(\hat{d}, \hat{\beta}) - H_{12}^2(\hat{d}, \hat{\beta}))^{-1/2}, \tag{60}$$

$$SE_J = \pi/\sqrt{6} J_{22,n}^{1/2}(\hat{d}, \hat{\beta})(J_{11}(\hat{d}, \hat{\beta}) J_{22}(\hat{d}, \hat{\beta}) - J_{12}^2(\hat{d}, \hat{\beta}))^{-1/2}. \tag{61}$$

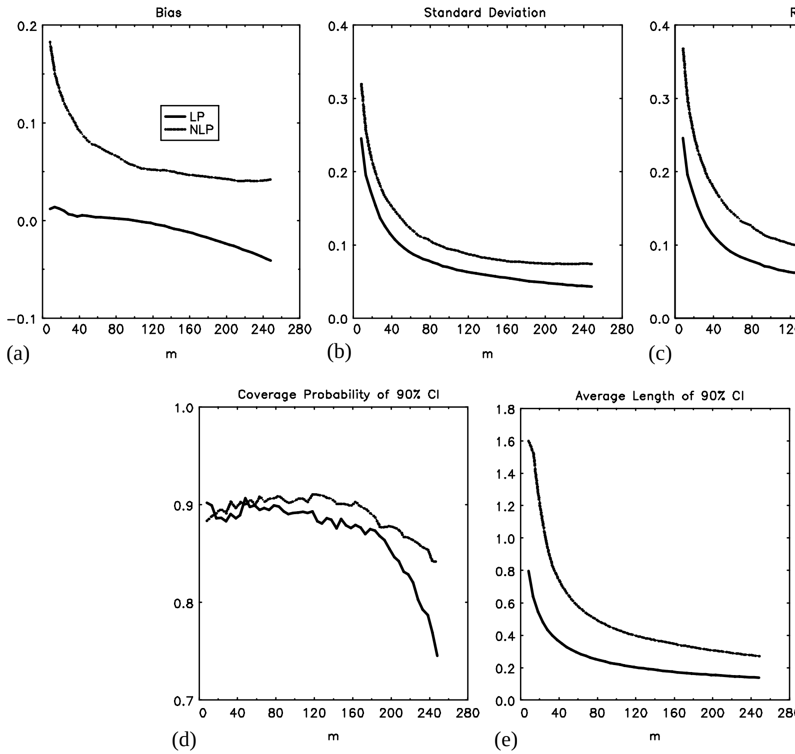

Fig. 1. Performances of the NLP estimator and the LP estimator with $d_0 = 0.45$ and $\sigma_u^2 = 0$ for sample size 512.

The standard errors of $\hat{d}_{\mathrm{LP}}$ is estimated by

$$\pi/\sqrt{24}\left(\sum_{j=1}^{m}\left(\log\lambda_j - \frac{1}{m}\sum_{k=1}^{m}\log\lambda_k\right)^2\right)^{-1/2}. \tag{62}$$

The idea of using the finite sample expression instead of the asymptotic expression has been used in many papers (e.g. Andrews and Guggenberger, 2003; Andrews and Sun, 2002). Simulation results not reported show that SE_{HJ} provides a much better approximation than both SE_J and SE_A. The approximation (62) was originally suggested by Geweke and Porter-Hudak (1983) and was used in Deo and Hurvich (2001).

5.2. Results

We report results for the cases $d_0 = 0.45$ and $d_0 = 0.85$ in details, since these are representative of the results found in the other two cases, $d_0 = 0.25$ and 0.65, respectively. Also, for each value of d_0, we discuss only the cases $\sigma_u^2 = 0$ and $\sigma_u^2 = 8$, as the results for the other values of σ_u^2 were qualitatively similar. We will concentrate on the case $n = 512$.

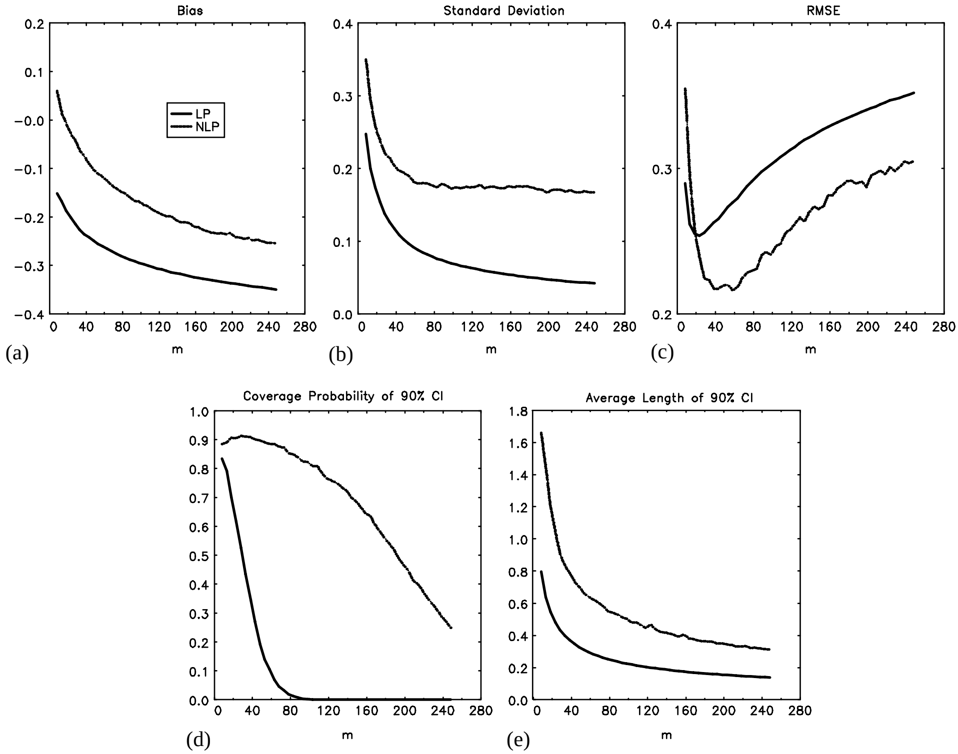

Fig. 2. Performances of the NLP estimator and the LP estimator with $d_0 = 0.45$ and $\sigma_u^2 = 8$ for sample size 512.

We first discuss the results when $d_0 = 0.45$ and $\sigma_u^2 = 0$. In this case, z_t is a pure fractional process. Fig. 1(a) shows that the bias of $\hat{d}$ is positive and larger than that of $\hat{d}_{\mathrm{LP}}$. The positiveness of the bias of $\hat{d}$ is not surprising. Intuitively, the two regressors λ_j^{2d} and $\log \lambda_j$ in the NLP regression move together. When $\sigma_u = 0$, we have $\beta_0 = 0$. But $\hat{\beta}$ is constrained to be positive, we thus expect $\hat{d}$ to be biased upward. Fig. 1(b) shows that the variance of $\hat{d}$ is larger than that of $\hat{d}_{\mathrm{LP}}$. Comparing RMSE's in Fig. 1(c), we see that the RMSE of $\hat{d}$ is larger than that of $\hat{d}_{\mathrm{LP}}$. The inferior performance of $\hat{d}$ in this case is not surprising since the LP estimator is designed for pure fractional processes, whereas our estimator $\hat{d}$ allows for additional noise in the system and is designed for perturbed fractional processes. However, it is encouraging that the LP estimator outperforms the NLP estimator only by a small margin. Apparently, the cost of including the additional regressor, even when it is not needed, is small.

Next, we discuss the results when $d_0 = 0.45$ and $\sigma_u^2 = 8$. Fig. 2(a) shows that the LP estimator $\hat{d}_{\mathrm{LP}}$ has a large downward bias in this case, whereas the NLP estimator $\hat{d}$ has a much smaller bias. Apparently, the bias-reducing feature of $\hat{d}$ established in the asymptotic theory is manifest in finite samples. Fig. 2(b) shows that the standard error of $\hat{d}_{\mathrm{LP}}$ is less than that of $\hat{d}$ for all values of m, again consistent with the asymptotic

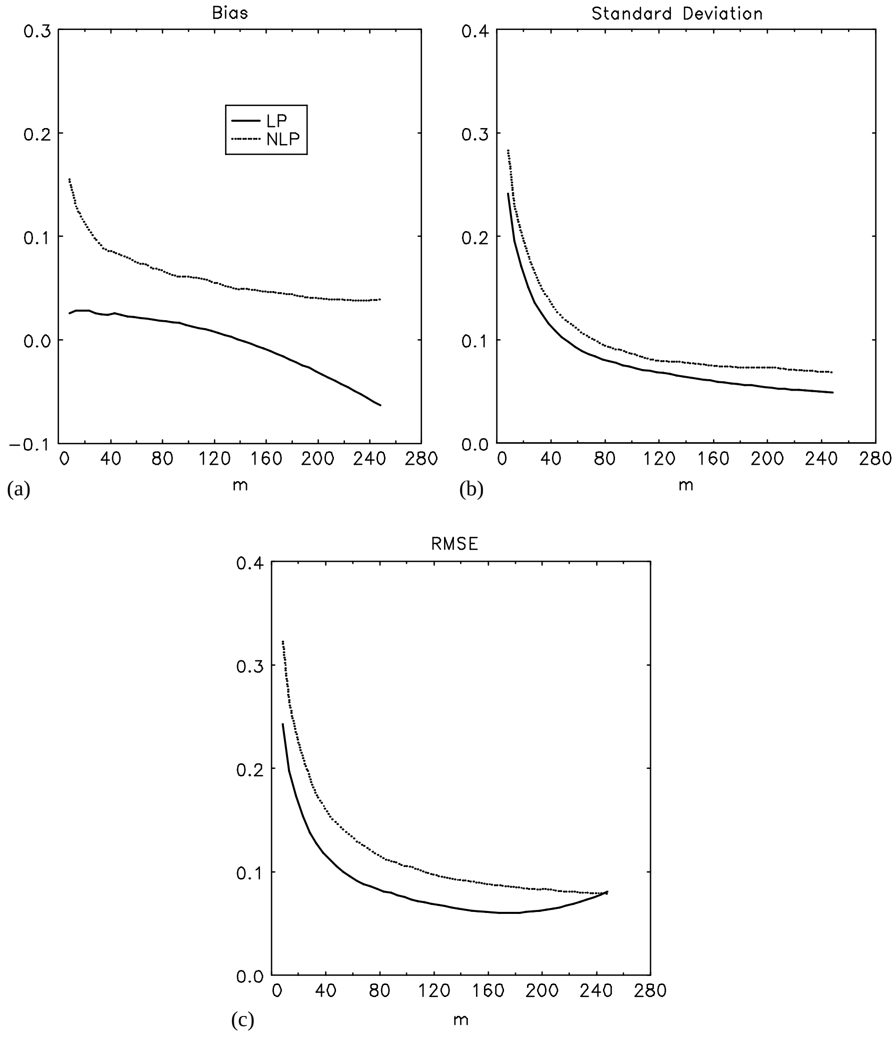

Fig. 3. Performances of the NLP estimator and the LP estimator with $d_0 = 0.85$ and $\sigma_u^2 = 0$ for sample size 512.

results. For each estimator, the standard error declines at the approximate rate $1/\sqrt{m}$ as m increases, because m is the effective sample size in the estimation of d_0. Fig. 2(c) shows that the RMSE of $\hat{d}$ is smaller than that of $\hat{d}_{\mathrm{LP}}$ over a wide range of m values. Fig. 2(d) shows that the coverage probability of $\hat{d}$ is fairly close to the nominal value of 0.9, provided that m is not taken too large. In contrast, $\hat{d}_{\mathrm{LP}}$ has a true coverage probability close to 0.9 only for very small values of m. This is due to the large bias of $\hat{d}_{\mathrm{LP}}$. However, the larger standard error of $\hat{d}$ leads to longer CI on average, and this is apparent in Fig. 2(e).

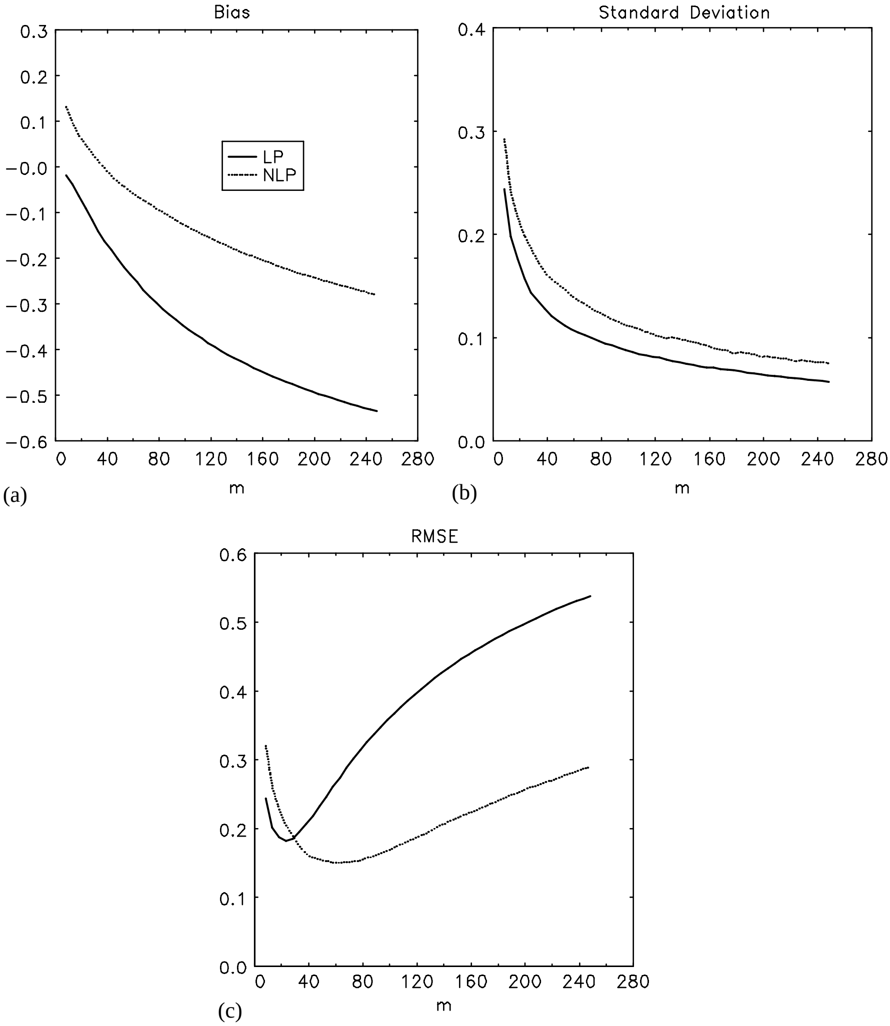

Fig. 4. Performances of the new estimator and the GPH estimator with $d_0 = 0.85$ and $\sigma_u^2 = 8$ for sample size 512.

The qualitative comparisons and conclusions made for the case $d_0 = 0.45$ remain valid for the case $d_0 = 0.25$. For brevity, we do not present the figures but we comment on these figures briefly. When $d_0 = 0.25$ and $\sigma_u^2 = 0$, the bias and standard deviation of $\hat{d}_{\mathrm{LP}}$ remain more or less the same as in Fig. 1(a) and (b). Comparing with Fig. 1, the bias curve of $\hat{d}$ remains the same, but the standard deviation curve moves up, meaning that variance inflation is more serious. When $d_0 = 0.25$ and $\sigma_u^2 = 8$, the bias reduction of $\hat{d}$ is slightly less effective and the variance inflation is slightly larger than was shown in Fig. 2. Nevertheless, the RMSE of $\hat{d}$ is still smaller than that of $\hat{d}_{\mathrm{LP}}$ for a wide range of the m values.

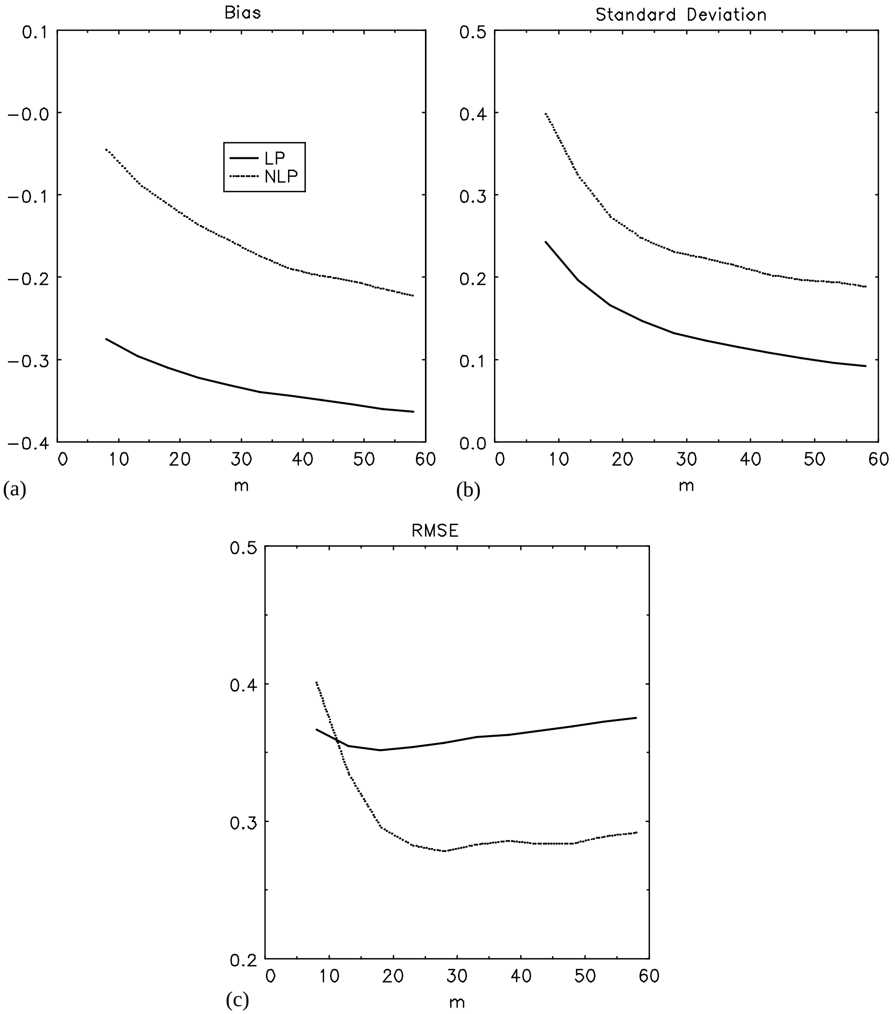

Fig. 5. Performances of the NLP estimator and the LP estimator with $d_0 = 0.45$ and $\sigma_u^2 = 8$ for sample size 128.

We now turn to the results when $d_0 = 0.85$ and $\sigma_u^2 = 0$. To save space, we only present the bias, standard deviation and RMSE graphs analogous to graphs (a), (b) and (c) in Fig. 1. Fig. 3 shows that both $\hat{d}_{\mathrm{LP}}$ and $\hat{d}$ work reasonably well for nonstationary fractional processes ($1/2 \leqslant d_0 < 1$). Compared with Fig. 1, we find that the difference in the standard errors of these two estimators becomes smaller while the difference in the biases remains more or less the same. Although $\hat{d}_{\mathrm{LP}}$ is still a better estimator than $\hat{d}$ in this case, the advantage of $\hat{d}_{\mathrm{LP}}$ has clearly diminished with the increase in d_0.

Fig. 4 provides results for the case $d_0 = 0.85$ and $\sigma_u^2 = 8$. Fig. 4(a) shows that the bias reduction from using $\hat{d}$ is substantial. For example, when $m = 40$, the bias of

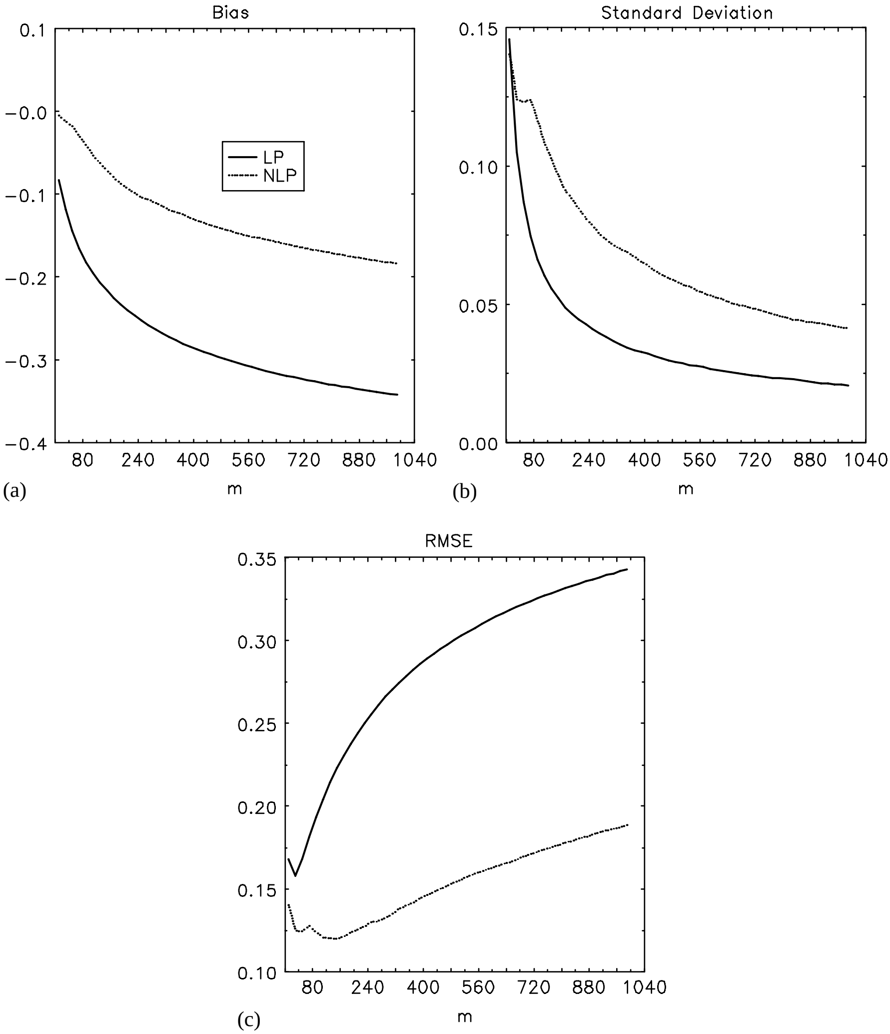

Fig. 6. Performances of the NLP estimator and the LP estimator with $d_0 = 0.45$ and $\sigma_u^2 = 8$ for sample size 2048.

$\hat{d}_{\mathrm{LP}}$ is -0.18, while that of $\hat{d}$ is only -0.02. The evidence seems to suggest that $\hat{d}$ is effective in reducing bias for stationary fractional component models as well as for nonstationary models. Fig. 4(b) shows that the standard error of $\hat{d}$ is only slightly larger than that of $\hat{d}_{\mathrm{LP}}$. The large bias reduction and small variance inflation lead to a smaller RMSE for $\hat{d}$ over a wide range of m values, as shown in Fig. 4(c). Other simulations results (not reported in Fig. 4) show that the coverage probability based on $\hat{d}_{\mathrm{LP}}$ decreases very rapidly as m increases, whereas that based on $\hat{d}$ decreases much more slowly. In fact, the coverage probability based on $\hat{d}$ is close to 0.9 over a wide range of m values.

The simulation results for $d_0 = 0.65$ are qualitatively similar to those found for the case $d_0 = 0.85$. We omit the detailed discussion. Comparing the simulation results for different values of d_0, we find that $\hat{d}$ is more effective in bias reduction for larger values of d_0. Intuitively, when d_0 is small, the bias of $\hat{d}_{\mathrm{LP}}$ is small no matter what value σ_u may take. For a large value of σ_u, the perturbation component dominates the fractional component, so that $\hat{d}_{\mathrm{LP}}$ would be around 0. In this case, the bias of $\hat{d}_{\mathrm{LP}}$ is small only because the true value of d_0 itself is small. Also, for small values of σ_u, the bias from contamination is naturally going to be small. Therefore, in both cases, the bias of $\hat{d}_{\mathrm{LP}}$ will be small when d_0 is small and there is not much scope for $\hat{d}$ to manifest its bias-reducing capacity.

We present a representative figure for the cases $n=2048$ and $n=512$. The qualitative comparisons made and conclusions reached for the $n = 512$ sample size continue to apply to $n=2048$ and $n=128$. The simulation results show that $\hat{d}$ is more effective in bias reduction when the sample size is smaller. This is because a smaller sample size implies a larger finite sample bias of $\hat{d}_{\mathrm{LP}}$ and there is some scope for $\hat{d}$ to manifest its bias-reducing capacity.

To sum up, the simulations show that, for fractional component processes, the NLP estimator $\hat{d}$ has a lower bias, a higher standard deviation, and a lower RMSE in comparison to the LP estimator $\hat{d}_{\mathrm{LP}}$, corroborating the asymptotic theory. The lower bias generally leads to improved coverage probability in CI based on $\hat{d}$ over a wide range of m. On the other hand, the lower standard deviation of $\hat{d}_{\mathrm{LP}}$ leads to shorter CI than those based on $\hat{d}$.

6. Conclusion

In empirical applications it has become customary practice to investigate the order of integration of the variables in a model when nonstationarity is suspected. This practice is now being extended to include analyses of the degree of persistence using fractional models and estimates of long memory parameters. Nonetheless, for many time series, and particularly macroeconomic variables for which there is limited data, the actual degree of persistence in the data continues to be a controversial issue. The empirical resolution of this problem inevitably relies on our capacity to separate low-frequency behavior from high-frequency fluctuations and this is particularly difficult when short run fluctuations have high variance. Actual empirical results often depend critically on the discriminatory power of the statistical techniques being employed to implement the separation.

The model used in the present paper provides some assistance in this regard. It allows for an explicit components structure in which there are different sources and types of variation, thereby accommodating a separation of short and long memory components and allowing for fractional processes that are perturbed by weakly dependent effects. Compared to the conventional formulation of a pure fractional process like (2), perturbed fractional processes allow for multiple sources of high-frequency variation and, in doing so, seem to provide a richer setting for uncovering latent persistence in an observed time series. In particular, the model provides a mechanism for simultaneously capturing the effects of persistent and temporary shocks and seems realistic in economic and financial applications when there are many different sources of variation in the data. The new econometric methods we have introduced for estimating the fractional parameter in such models take account of the presence of additive disturbances, and help to achieve bias reduction and attain a faster rate of convergence. The asymptotic theory is easy to use and seems to work reasonably well in finite samples.

The methods of the paper can be extended in a number of directions. First, the nonlinear approximation approach can be used in combination with other estimators, such as the local Whittle estimator (Robinson, 1995b), which seems natural in the present context because the procedure already uses optimization methods. Second, the idea of using a nonlinear approximation can be applied to nonstationary fractional component models and used to adapt the methods which have been suggested elsewhere (e.g., Phillips, 1999; Shimotsu and Phillips, 2001) for estimating the memory parameter in such models to cases where there are fractional components.

Appendix A. Proofs

Proof of Lemma 1. A spectral density satisfying Assumptions 1 and 2 also satisfies Assumptions 1 and 2 of Robinson (1995a). In consequence, the lemma follows from Theorem 2 of Robinson (1995a). Since we normalize the discrete Fourier transform by the spectral density $f_z^{1/2}(\lambda)$ instead of the power function $C_g^{-1/2}\lambda^{-d}$, (4.2) of Robinson (1995a) is always zero and the extra term $(j/n)^{\min(\alpha,\beta)}$ in Robinson (1995a) does not arise in our case. □

Proof of Lemma 3. Note that

$$m^{-1}\sum_{j=1}^{m} c_j U_j = m^{-1}\sum_{j=1}^{[\log^2 m]} c_j U_j + m^{-1}\sum_{j=[\log^2 m]+1}^{m} c_j U_j \equiv F_1 + F_2. \tag{A.1}$$

But $E\sup_{(d,\beta)'\in\Theta}|F_1|$ is less than

$$Em^{-1}\sum_{j=1}^{[\log^2 m]} \sup_{(d,\beta)'\in\Theta}|c_j||U_j| \leqslant m^{-1}\log^p m \sum_{j=1}^{[\log^2 m]}(EU_j^2)^{1/2} = \mathrm{O}(\log^{p+2} m/m) \tag{A.2}$$

by Lemma 2(b). Hence

$$\sup_{(d,\beta)'\in\Theta}|F_1| = \mathrm{O_p}(\log^{p+2} m/m) = \mathrm{O_p}(\log^p m/\sqrt{m}). \tag{A.3}$$

Let $s_r = \sum_{k=[\log^2 m]+1}^{r} U_r$, $r = [\log^2 m]+1,\ldots,m$ and $s_{[\log^2 m]} = 0$. Then, from Lemma 2(a), (c) and (d), it follows that

$$Es_r^2 = \sum_{k=[\log^2 m]+1}^{r} EU_k^2 + 2\sum_{[\log^2 m]+1\leqslant k<j\leqslant r} EU_j U_k$$

$$= \sum_{k=[\log^2 m]+1}^{r} \left(\frac{\pi^2}{6} + k^{-1} \log k \right) + 2 \sum_{[\log^2 m+1] \leqslant k < j < r}^{r} \mathrm{O}(k^{-2} \log^2 j)$$

$$= \mathrm{O}(r) + \mathrm{O}(r \log^2 r / \log^2 m), \tag{A.4}$$

which implies $s_r = \mathrm{O_p}(r^{1/2})$. Using this result and partial summation, we have:

$$\sup_{(d,\beta)' \in \Theta} |F_2| \leqslant \sup_{(d,\beta)' \in \Theta} \left| m^{-1} \sum_{j=[\log^2 m]+1}^{m} c_j U_j \right|$$

$$= \sup_{(d,\beta)' \in \Theta} m^{-1} \left| \sum_{j=[\log^2 m]+1}^{m} s_{j-1}(c_{j-1} - c_j) \right| + \sup_{(d,\beta)' \in \Theta} m^{-1} |s_m c_m|$$

$$= m^{-1} \sum_{j=[\log^2 m]+1}^{m} \mathrm{O_p}(j^{1/2}) \mathrm{O}(j^{-1} \log^q m) + \mathrm{O_p}(\log^p m / \sqrt{m})$$

$$= \log^q m / m \sum_{j=[\log^2 m]+1}^{m} \mathrm{O_p}(j^{-1/2}) + \mathrm{O_p}(\log^p m / \sqrt{m})$$

$$= \mathrm{O_p}((\log^{\max(p,q)} m) / \sqrt{m}). \tag{A.5}$$

Combine (A.3) with (A.5) to complete the proof. □

Proof of Theorem 1. When $\sigma_u = 0$, the theorem is essentially the same as results already established in HDB. Only one modification is needed. HDB use $-2 \log(2 \sin \lambda_j / 2)$ as one of the regressors while we employ $-2 \log \lambda_j$. The use of $-2 \log \lambda_j$ rather than $-2 \log(2 \sin \lambda_j / 2)$ has no effect on the asymptotic variance, but it does affect the asymptotic bias. This is because the asymptotic bias comes from the dominating term in ε_j and this term is different for different regressors. Using $-2 \log(2 \sin \lambda_j / 2)$ as the regressor yields

$$\varepsilon_j = \log f_w(\lambda_j) - \log f_w(0) = \left(\frac{f_w''(0)}{2 f_w'(0)} \right) \lambda_j^2 (1 + \mathrm{o}(1)). \tag{A.6}$$

In contrast, using $-2 \log \lambda_j$ as the regressor yields

$$\varepsilon_j = \left(\frac{f_w''(0)}{2 f_w'(0)} + \frac{d_0}{12} \right) \lambda_j^2 (1 + \mathrm{o}(1)). \tag{A.7}$$

With this adjustment, the arguments in HDB go through without further change.

Now consider the case $\sigma_u > 0$. Rewrite the spectral density of z_t as $f_z(\lambda) = \lambda^{-2d_0} g(\lambda)$, where $g(\lambda) = (\lambda^{-1} 2 \sin \lambda / 2)^{-2d_0} f^*(\lambda)$. Since

$$g(\lambda) - g(0) = (1 + \mathrm{O}(\lambda^2))(f_w(0) + \lambda^{2d_0} f_u(0) + \mathrm{O}(\lambda^2)) - f_w(0) = \mathrm{O}(\lambda^{2d_0}) \tag{A.8}$$

as $\lambda \to 0+$, $g(\lambda)$ is smooth of order $2d_0$. Combining this with our assumption that $m \to \infty$ and $m^{4d_0+1}/n^{4d_0} = \mathrm{O}(1)$ verifies Assumptions 1 and 2 of Andrews and Guggenberger

(2003). Hence their Theorem 1 is valid with $r=0$, $s=2d_0$ and $q=2d_0$. It is easy to show that the term $\mathrm{O}(m^q/n^q)$ in their theorem is actually $-f_u(0)/f_w(0)d_0(2d_0+1)^{-2}\lambda_m^{2d_0}$. Andrews and Guggenberger established asymptotic normality under their Assumption 3 that $m^{4d_0+1}/n^{4d_0}=\mathrm{o}(1)$. In fact, asymptotic normality holds under our assumption $m^{4d_0+1}/n^{4d_0}=\mathrm{O}(1)$ as long as an asymptotic bias of order $\mathrm{O}(1)$ is allowed. □

Proof of Theorem 2. Decompose $Q(d,\beta)-Q(d_0,\beta_0)$ into two parts as follows:

$$Q(d,\beta)-Q(d_0,\beta_0)=\frac{1}{m}\sum_{j=1}^{m}(V_j-\bar V)^2+\frac{2}{m}\sum_{j=1}^{m}(U_j+\varepsilon_j)(V_j-\bar V) \tag{A.9}$$

where the dependence on (d,β) has been suppressed for notational simplicity.

(a) We prove part (a) by showing that $1/m\sum_{j=1}^{m}(U_j+\varepsilon_j)(V_j-\bar V)=\mathrm{o_p}(1)$ uniformly in $(d,\beta)'$ and $1/m\sum_{j=1}^{m}(V_j-\bar V)^2$ converges uniformly to a function, which has a unique minimizer d_0.

First, using Corollary 1 with $\gamma=0$, we have

$$\sup_{(d,\beta)'\in\Theta}\left|1/m\sum_{j=1}^{m}U_j(V_j-\bar V)\right|=\mathrm{O_p}(1/\sqrt{m}). \tag{A.10}$$

Next, we show that $\sup_{(d,\beta)'\in\Theta}|\frac{1}{m}\sum_{j=1}^{m}\varepsilon_j(V_j(d,\beta)-\bar V(d,\beta))|=\mathrm{O_p}(\lambda_m^{4d_0})$. Under Assumptions 1 and 2, $\varepsilon_j=\mathrm{O}(\lambda_j^r)=\mathrm{O}(\lambda_j^{4d_0})$. So we have, using the fact that $\sup_{(d,\beta)'\in\Theta}|V_{j-1}(d,\beta)-V_j(d,\beta)|=\mathrm{O}(1/j)$ and $\sup_{(d,\beta)'\in\Theta}|V_m(d,\beta)-\bar V(d,\beta)|=\mathrm{O}(1)$,

$$\begin{aligned}
&\sup_{(d,\beta)'\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\varepsilon_j(V_j(d,\beta)-\bar V(d,\beta))\right|\\
&\leqslant\sup_{(d,\beta)'\in\Theta}\left(\frac{1}{m}\left|\sum_{j=1}^{m}\sum_{r=1}^{j-1}\varepsilon_r(V_{j-1}(d,\beta)-V_j(d,\beta))\right|\right.\\
&\qquad\left.+\frac{1}{m}\left|\sum_{j=1}^{m}\varepsilon_j\right|\left|V_m(d,\beta)-\bar V(d,\beta)\right|\right)\\
&=\lambda_m^{4d_0}\frac{1}{m}\left|\sum_{j=1}^{m}\sum_{r=1}^{j-1}\mathrm{O_p}\left(\frac{r}{m}\right)^{4d_0}\left(\frac{1}{j}\right)\right|+\mathrm{O_p}(\lambda_m^{4d_0})=\mathrm{O_p}(\lambda_m^{4d_0})=\mathrm{o_p}(1).
\end{aligned} \tag{A.11}$$

Finally,

$$\begin{aligned}
\frac{1}{m}\sum_{j=1}^{m}(V_j-\bar V)^2&=\frac{1}{m}\sum_{j=1}^{m}\left(2(d-d_0)\left(\log\left(\frac{j}{m}\right)-\frac{1}{m}\sum_{k=1}^{m}\log\left(\frac{k}{m}\right)\right)+\mathrm{o}(1)\right)^2\\
&=4(d-d_0)^2(1+\mathrm{o}(1)),
\end{aligned} \tag{A.12}$$

where $\mathrm{o}(\cdot)$ holds uniformly over $(d,\beta)'\in\Theta$.

In view of (A.10), (A.11) and (A.12), we can complete the proof by using a standard textbook argument.

(b) We proceed by showing first that $\hat{d}$ converges to d_0 at some preliminary rate and then go on to show that $\hat{d} - d_0 = O_p((m/n)^{2d_0})$. We obtain this rate sequentially.

First, we show that $\hat{d} - d_0 = o_p((m/n)^{d_1/2})$, where d_1 is the lower bound of the interval D. From $Q(\hat{d}, \hat{\beta}) - Q(d_0, \beta_0) \leqslant 0$, we get

$$\frac{1}{m}\sum_{j=1}^{m}(V_j(\hat{d},\hat{\beta}) - \bar{V}(\hat{d},\hat{\beta}))^2 \leqslant -\frac{2}{m}\sum_{j=1}^{m}(U_j + \varepsilon_j)(V_j(\hat{d},\hat{\beta}) - \bar{V}(\hat{d},\hat{\beta}))$$

$$= O_p\left(\frac{1}{\sqrt{m}}\right) + O_p(\lambda_m^{4d_0}) = o_p\left(\left(\frac{m}{n}\right)^{2d_1}\right), \quad \text{(A.13)}$$

where the last equality follows from the assumptions that $n^{4d_0(1+\Delta)}/m^{4d_0(1+\Delta)+1} = o(1)$ and that $d \geqslant d_1 > 0$. But $\frac{1}{m}\sum_{j=1}^{m}(V_j(\hat{d},\hat{\beta}) - \bar{V}(\hat{d},\hat{\beta}))^2$ equals

$$\frac{1}{m}\sum_{j=1}^{m}\left(2(\hat{d}-d_0)\left(\log\left(\frac{j}{m}\right) - \frac{1}{m}\sum_{j=1}^{m}\log\left(\frac{j}{m}\right)\right) + O(\lambda_m^{2\hat{d}}) + O(\lambda_m^{2d_0})\right)^2$$

$$= 4(\hat{d}-d_0)^2(1+o(1)) + O(\lambda_m^{2d_0}) + O(\lambda_m^{2\hat{d}})$$

$$= 4(\hat{d}-d_0)^2(1+o(1)) + O\left(\left(\frac{m}{n}\right)^{2d_1}\right). \quad \text{(A.14)}$$

Therefore,

$$4(\hat{d}-d_0)^2(1+o(1)) + O_p\left(\left(\frac{m}{n}\right)^{2d_1}\right) \leqslant o_p\left(\left(\frac{m}{n}\right)^{2d_1}\right), \quad \text{(A.15)}$$

which implies that $\hat{d} - d_0$ is at most $O_p((m/n)^{d_1})$. Thus $\hat{d} - d_0 = o_p((m/n)^{d_1/2})$.

Second, we show that $\hat{d}-d_0=o_p((m/n)^{d_0(1+\Delta)})$. Since $\hat{d}-d_0=o_p((m/n)^{d_1/2})$, we only need consider $d \in D_n' = \{d\colon |d-d_0| < \varepsilon(m/n)^{d_1/2}\}$ for some small $\varepsilon > 0$. Approximating sums by integrals, we deduce that, for $d \in D_n'$,

$$\frac{1}{m}\sum_{j=1}^{m}V_j^2(d,\beta) - (\bar{V}(d,\beta))^2 = \mathscr{I}_1 + \mathscr{I}_2 \quad \text{(A.16)}$$

where

$$\mathscr{I}_1 = \left(4(d-d_0)^2 + \left(\frac{2d\beta\lambda_m^{2d}}{(2d+1)\sqrt{4d+1}} - \frac{2d_0\beta_0\lambda_m^{2d_0}}{(2d_0+1)\sqrt{4d_0+1}}\right)^2\right)(1+o(1)), \quad \text{(A.17)}$$

and

$$\mathscr{I}_2 = \frac{8dd_0\beta\beta_0\lambda_m^{2d+2d_0}}{(2d+1)(2d_0+1)}\left(\frac{1}{\sqrt{(4d+1)(4d_0+1)}} - \frac{1}{2d+2d_0+1}\right)(1+o(1)). \quad \text{(A.18)}$$

Therefore

$$\frac{1}{m}\sum_{j=1}^{m} V_j^2(d,\beta) - (\bar{V}(d,\beta))^2 = 4(d-d_0)^2 + \mathrm{O}(\lambda_m^{4d_0}), \tag{A.19}$$

where the o(·) and O(·) terms in the above three equations hold uniformly over $(d,\beta)' \in D_n' \times B$. Using $Q(\hat{d},\hat{\beta}) - Q(d_0,\beta_0) \leqslant 0$ again, we have

$$\frac{1}{m}\sum_{j=1}^{m}(V_j(\hat{d},\hat{\beta}) - \bar{V}(\hat{d},\hat{\beta}))^2 \leqslant \mathrm{O_p}\left(\frac{1}{\sqrt{m}}\right) + \mathrm{O_p}(\lambda_m^{4d_0}) = \mathrm{o_p}\left(\left(\frac{m}{n}\right)^{2d_0(1+\Delta)}\right), \tag{A.20}$$

where the equality follows from the assumption $n^{4d_0(1+\Delta)}/m^{4d_0(1+\Delta)+1} = \mathrm{o}(1)$. Combining (A.19) and (A.20), we get

$$4(\hat{d}-d_0)^2 + \mathrm{o}(\lambda_m^{2d_0(1+\Delta)}) \leqslant \mathrm{o_p}\left(\left(\frac{m}{n}\right)^{2d_0(1+\Delta)}\right). \tag{A.21}$$

Hence $\hat{d} - d_0 = \mathrm{o_p}((m/n)^{d_0(1+\Delta)})$. Repeating the above procedure a finite number of times, we obtain: $\hat{d} - d_0 = \mathrm{O_p}((m/n)^{2d_0})$. Details are omitted here and are available upon request.

Now, since $(2d+2d_0+1)^2 - (4d+1)(4d_0+1) = 4d^2 - 8dd_0 + 4d_0^2 = 4(d-d_0)^2 > 0$, we deduce from (A.16) that

$$\frac{1}{m}\sum_{j=1}^{m}(V_j - \bar{V})^2 \geqslant \left(\frac{2d\beta\lambda_m^{2d}}{(2d+1)\sqrt{4d+1}} - \frac{2d_0\beta_0\lambda_m^{2d_0}}{(2d_0+1)\sqrt{4d_0+1}}\right)^2 (1+\mathrm{o}(1)). \tag{A.22}$$

for $d \in D$ such that $|d-d_0| \leqslant C(m/n)^{2d_0}$. In view of $\frac{1}{m}\sum_{j=1}^{m}(V_j(\hat{d},\hat{\beta}) - \bar{V}(\hat{d},\hat{\beta}))^2 \leqslant \mathrm{o_p}(\lambda_m^{4d_0})$, we obtain

$$\left(\frac{2\hat{d}\hat{\beta}\lambda_m^{2\hat{d}}}{(2\hat{d}+1)\sqrt{4\hat{d}+1}} - \frac{2d_0\beta_0\lambda_m^{2d_0}}{(2d_0+1)\sqrt{4d_0+1}}\right)^2 (1+\mathrm{o}(1)) \leqslant \mathrm{o_p}(\lambda_m^{4d_0}). \tag{A.23}$$

Some algebraic manipulations show that when $\hat{d} - d_0 = \mathrm{O_p}((m/n)^{2d_0})$,

$$\frac{4\hat{d}\lambda_m^{4(\hat{d}-d_0)}}{(2\hat{d}+1)^2(4\hat{d}+1)}(\hat{\beta}-\beta_0)^2 \leqslant \mathrm{o_p}(\lambda_m^{4d_0}), \tag{A.24}$$

from which we deduce that $\hat{\beta} - \beta_0 = \mathrm{o_p}(1)$. □

Proof of Lemma 5. (a) The (2,2) element of $\sup_{\theta\in\Theta_n}\|D_n^{-1}(H_n(d,\beta) - J_n(d,\beta))D_n^{-1}\|$ is zero, so it suffices to consider the (1,1) and (1,2) elements. Since

$$I_{z_j} + 2d\log\lambda_j - \beta\lambda_j^{2d} = \alpha_0 + U_j + \varepsilon_j + (d-d_0)\log\lambda_j^2 + \beta_0\lambda_j^{2d_0} - \beta\lambda_j^{2d},$$

$\sup_{\theta\in\Theta_n} |(\beta/m)\sum_{j=1}^m e_j(\log\lambda_j^2)^2\lambda_j^{2d}|$, the (1,1) element, is bounded by $L_1+L_2+L_3+L_4$, where

$$L_1 = \sup_{\theta\in\Theta_n}\left|\frac{\beta}{m}\sum_{j=1}^m\left((\log\lambda_j^2)^2\lambda_j^{2d}-\frac{1}{m}\sum_{k=1}^m(\log\lambda_k^2)^2\lambda_k^{2d}\right)U_j\right|, \tag{A.25}$$

and L_2, L_3 and L_4 are defined as L_1 is defined, but with U_j replaced by ε_j, $(d-d_0)\log\lambda_j^2$ and $\beta_0\lambda_j^{2d_0}-\beta\lambda_j^{2d}$, respectively.

We can show that $L_i = o_p(1)$ for $i=1,2,3,4$. To save space, we only present the proof of $L_1 = o_p(1)$ here. Note that $\log^2(\lambda_j^2)\lambda_j^{2d}-(1/m)\sum_{k=1}^m\log^2(\lambda_k^2)\lambda_k^{2d}$ equals

$$4\log^2\lambda_m\left(\lambda_j^{2d}-\frac{1}{m}\sum_{k=1}^m\lambda_k^{2d}\right)+8\log\lambda_m\left(\log\left(\frac{j}{m}\right)\lambda_j^{2d}-\frac{1}{m}\sum_{k=1}^m\log\left(\frac{k}{m}\right)\lambda_k^{2d}\right)$$
$$+4\log^2\left(\frac{j}{m}\right)\lambda_j^{2d}-\frac{4}{m}\sum_{k=1}^m\log^2\left(\frac{k}{m}\right)\lambda_k^{2d}. \tag{A.26}$$

L_1 is thus bounded by $\sup_{\theta\in\Theta_n}|4\beta\lambda_m^{2d}|(\log^2\lambda_m L_{11}+2|\log\lambda_m|L_{12}+L_{13})$, where

$$L_{1i+1} = \sup_{\theta\in\Theta_n}\left|\frac{1}{m}\sum_{j=1}^m\left(\left(\frac{j}{m}\right)^{2d}\log^i\left(\frac{j}{m}\right)-\frac{1}{m}\sum_{k=1}^m\left(\frac{k}{m}\right)^{2d}\log^i\left(\frac{k}{m}\right)\right)U_j\right|,$$
$$i=0,1,2. \tag{A.27}$$

It follows from Lemma 3 that $L_{1i+1} = O_p(\log^i m/\sqrt{m})$. Therefore

$$L_1 = O_p\left(\frac{\log^2\lambda_m}{\sqrt{m}}\lambda_m^{2d_1}+\frac{|\log\lambda_m|\log m}{\sqrt{m}}\lambda_m^{2d_1}+\frac{\log^2 m}{\sqrt{m}}\lambda_m^{2d_1}\right)=o(1). \tag{A.28}$$

Hence the (1,1) element of $\sup_{\theta\in\Theta_n}\|D_n^{-1}(H_n(d,\beta)-J_n(d,\beta))D_n^{-1}\|$ is $o_p(1)$.

Following the same procedure, we can show that the (1,2) element is $o_p(1)$. The details are omitted.

(b) and (c) hold by using $x_{1j}(d_0,\beta_0) = -2\log\lambda_j(1+o(1))$ and $x_{2j}=\lambda_j^{2d_0}$ and approximating sums by integrals.

(d) Let $\xi_j = (\xi_{1j},\xi_{2j})'$, where

$$\xi_{1j} = -2\log\frac{j}{m}+\frac{2}{m}\sum_{k=1}^m\log\frac{k}{m},\quad \xi_{2j}=\left(\frac{j}{m}\right)^{2d_0}-\frac{1}{m}\sum_{k=1}^m\left(\frac{k}{m}\right)^{2d_0}. \tag{A.29}$$

Then, we can rewrite $D_n^{-1}S_n(d_0,\beta_0)$ as

$$D_n^{-1}S_n(d_0,\beta_0) = -\frac{1}{\sqrt{m}}\sum_{j=1}^m\xi_j(U_j+\varepsilon_j)(1+o(1)). \tag{A.30}$$

Note that $\sum_{j=1}^{m} \xi_{1j}\varepsilon_j$ equals

$$\{\sigma_u > 0\}\lambda_m^{4d_0} \sum_{j=1}^{m} \left(-2\log\frac{j}{m} + \frac{2}{m}\sum_{k=1}^{m}\log\frac{k}{m}\right)\left(-\frac{f_w^2(0)}{2f_u^2(0)}\left(\frac{j}{m}\right)^{4d_0}\right)(1+\mathrm{o}(1))$$

$$+\{\sigma_u = 0\}\lambda_m^2 \sum_{j=1}^{m}\left(-2\log\frac{j}{m} + \frac{2}{m}\sum_{k=1}^{m}\log\frac{k}{m}\right)$$

$$\times\left(\left(\frac{j}{m}\right)^2\left(\frac{f_w''(0)}{2f_w(0)} + \frac{d_0}{12}\right)\right)(1+\mathrm{o}(1))$$

$$=\{\sigma_u > 0\}m\lambda_m^{4d_0}\frac{f_w^2(0)}{2f_u^2(0)}\frac{8d_0}{(4d_0+1)^2}(1+\mathrm{o}(1))$$

$$-\{\sigma_u = 0\}m\lambda_m^2\left(\frac{f_w''(0)}{f_w(0)} + \frac{d_0}{6}\right)\frac{2}{9}(1+\mathrm{o}(1)) \tag{A.31}$$

and $\sum_{j=1}^{m} \xi_{2j}\varepsilon_j$ equals

$$=\{\sigma_u > 0\}\lambda_m^{4d_0}\sum_{j=1}^{m}\left(\left(\frac{j}{m}\right)^{2d_0} - \frac{1}{m}\sum_{k=1}^{m}\left(\frac{k}{m}\right)^{2d_0}\right)\left(-\frac{f_u^2(0)}{2f_w^2(0)}(\frac{j}{m})^{4d_0}\right)(1+\mathrm{o}(1))$$

$$+\{\sigma_u = 0\}\lambda_m^2\sum_{j=1}^{m}\left(\left(\frac{j}{m}\right)^{2d_0} - \frac{1}{m}\sum_{k=1}^{m}\left(\frac{k}{m}\right)^{2d_0}\right)$$

$$\times\left(\left(\frac{j}{m}\right)^2\left(\frac{f_w''(0)}{2f_w(0)} + \frac{d_0}{12}\right)\right)(1+\mathrm{o}(1))$$

$$=-\{\sigma_u > 0\}m\lambda_m^{4d_0}\frac{f_w^2(0)}{2f_u^2(0)}\frac{8d_0^2}{(2d_0+1)(4d_0+1)(6d_0+1)}(1+\mathrm{o}(1))$$

$$+\{\sigma_u = 0\}m\lambda_m^2\left(\frac{f_w''(0)}{f_w(0)} + \frac{d_0}{6}\right)\frac{2d_0}{3(2d_0+3)(2d_0+1)}(1+\mathrm{o}(1)). \tag{A.32}$$

Therefore

$$D_n^{-1}S_n(d_0,\beta_0) + b = \frac{1}{\sqrt{m}}\sum_{j=1}^{m}\xi_j U_j + \mathrm{o}(1). \tag{A.33}$$

We now prove that for any vector $v=(v_1,v_2)'$, $(1/\sqrt{m})\sum_{j=1}^{m} v'\xi_j U_j \Rightarrow \mathrm{N}(0,(\pi^2/6)\, v'\Omega v)$. Write

$$\frac{1}{\sqrt{m}}\sum_{j=1}^{m} v'\xi_j U_j = T_1 + T_2 + T_3, \tag{A.34}$$

where

$$T_1 = \frac{1}{\sqrt{m}} \sum_{j=1}^{[\log^8 m]} a_j U_j, \quad T_2 = \frac{1}{\sqrt{m}} \sum_{j=[\log^8 m]+1}^{[m^{0.5+\delta}]} a_j U_j$$

$$T_3 = \frac{1}{\sqrt{m}} \sum_{j=[m^{0.5+\delta}]}^{m} a_j U_j, \quad a_j = v'\xi_j, \tag{A.35}$$

for some $0 < \delta < 0.5$.

Since $\max_{1\leqslant j\leqslant m} |\xi_{1j}| = \mathrm{O}(\log m)$ and $\max_{1\leqslant j\leqslant m} |\xi_{2j}| = \mathrm{O}(\log m)$, we have $\max_{1\leqslant j\leqslant m} |a_j| = \mathrm{O}(\log m)$. Therefore the proofs in HDB that $T_1 = \mathrm{o_p}(1)$ and $T_2 = \mathrm{o_p}(1)$ are also valid in the present case. We now show that $T_3 \to \mathrm{N}(0, (\pi^2/6)v'\Omega v)$ by verifying that the sequence $\{a_j\}$ satisfies (25) with $\rho = v'\Omega v$. The first condition of (25) holds as $\max_{1\leqslant j\leqslant m} |a_j| = \mathrm{O}(\log m) = \mathrm{o}(m)$. The second condition holds because

$$\begin{aligned}\sum_{j=[m^{0.5+\delta}+1]}^{m} a_j^2 &= \sum_{j=1}^{m} a_j^2 - \sum_{j=1}^{[m^{0.5+\delta}]} a_j^2 = \sum_{j=1}^{m} a_j^2 + \mathrm{o}(m)\\ &= mv'\left(\frac{1}{m}\sum_{j=1}^{m} \xi_j'\xi_j\right) v + \mathrm{o}(m) \sim mv'\Omega v.\end{aligned} \tag{A.36}$$

The last equality follows because we can show that $\lim_{m\to\infty} (1/m) \sum_{j=1}^{m} \xi_j'\xi_j = \Omega$ by approximating the sums by integrals. The third condition holds because

$$\begin{aligned}\sum_{j=[m^{0.5+\delta}]+1}^{m} |a_j|^p &\leqslant 2^p|v_1| \sum_{j=[m^{0.5+\delta}]+1}^{m} |\xi_{1j}|^p + 2^p|v_2| \sum_{j=[m^{0.5+\delta}]+1}^{m} |\xi_{2j}|^p\\ &= \mathrm{O}(m) + 2^p|v_1| \sum_{j=[m^{0.5+\delta}]+1}^{m} |\xi_{2j}|^p\\ &= \mathrm{O}\left(\sum_{j=[m^{0.5+\delta}]+1}^{m} \left|\left(\frac{2\pi j}{m}\right)^{2d_0}\right|^p\right)\\ &\quad + \mathrm{O}\left\{\sum_{j=[m^{0.5+\delta}]+1}^{m} \left[\frac{1}{m}\sum_{j=1}^{m}\left(\frac{2\pi j}{m}\right)^{2d_0}\right]^p\right\} + \mathrm{O}(m)\\ &= \mathrm{O}(m) + \mathrm{O}(m) + \mathrm{O}(m) = \mathrm{O}(m).\end{aligned} \tag{A.37}$$

Here we have employed $\sum_{j=[m^{0.5+\delta}]+1}^{m} |\xi_{1j}|^p = \mathrm{O}(m)$. See (A18) in HDB (1998).

The above results combine to establish part (d). □

Proof of Theorem 4. Let $\delta_n = (\delta_{n1}, \delta_{n2})' = -H_n^{-1}(d^*, \beta^*)S_n(d_0, \beta_0)$. It is easy to show that

(a) when $\delta_{n2} \geqslant 0$,

$$D_n(\hat{d} - d_0, \hat{\beta} - \beta_0)' = -D_n H_n^{-1}(d^*, \beta^*)S_n(d_0, \beta_0)$$
$$= -\Omega^{-1} D_n^{-1} S_n(d_0, \beta_0)(1 + \mathrm{o_p}(1));$$

(b) when $\delta_{n2} < 0$, $\hat{\beta} - \beta_0 = 0$ and

$$\sqrt{m}(\hat{d} - d_0) = -\sqrt{m} H_{n,11}^{-1}(d^*, \beta^*)S_{n,1}(d_0, \beta_0)$$
$$= -\Omega_{11}^{-1} m^{-1/2} S_{n,1}(d_0, \beta_0)(1 + \mathrm{o_p}(1)).$$

Let $\eta_n = (\eta_{n,1}, \eta_{n,2})' = D_n^{-1} S_n(d_0, \beta_0)$, then

$$\sqrt{m}(\hat{d} - d_0) = -(\tilde{\Omega}_{11}\eta_{n1} + \tilde{\Omega}_{12}\eta_{n2})\{\tilde{\Omega}_{12}\eta_{n1} + \tilde{\Omega}_{22}\eta_{n2} \leqslant 0\}(1 + \mathrm{o_p}(1))$$
$$-\Omega_{11}^{-1}\eta_{n1}\{\tilde{\Omega}_{12}\eta_{n1} + \tilde{\Omega}_{22}\eta_{n2} > 0\}(1 + \mathrm{o_p}(1)) \tag{A.38}$$

$$\sqrt{m}\lambda_m^{2d_0}(\hat{\beta} - \beta_0) = -(\tilde{\Omega}_{12}\eta_{n1} + \tilde{\Omega}_{22}\eta_{n2})\{\tilde{\Omega}_{12}\eta_{n1} + \tilde{\Omega}_{22}\eta_{n2} \leqslant 0\}(1 + \mathrm{o_p}(1)). \tag{A.39}$$

The proof is completed by invoking the continuous mapping theorem. □

References

Anderson, T., Bollerslev, T., 1997. Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns. Journal of Finance 52 (3), 975–1005.

Andrews, D.W.K., Guggenberger, P., 2003. A bias-reduced log-periodogram regression estimator for the long-memory parameter. Econometrica 71 (2), 675–712.

Andrews, D.W.K., Sun, Y., 2002. Adaptive Local polynomial Whittle estimation of long range dependence. Cowles Foundation Discussion Paper No. 1384, Yale University.

Breidt, F.J., Crato, N., De Lima, P., 1998. The detection and estimation of long memory in stochastic volatility. Journal of Econometrics 83, 32–348.

Deo, R.S., Hurvich, C.M., 2001. On the log periodogram regression estimator of the long memory parameter in the long memory stochastic volatility models. Econometric Theory 17 (4), 686–710.

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

Granger, C.W.J., Marmol, F., 1997. The correlogram of a long memory process plus a simple noise. Department of Economics, University of California, San Diego.

Henry, M., 1999. Robust automatic bandwidth for long memory. Department of Economics, Columbia University.

Henry, M., Robinson, P.M., 1996. Bandwidth choice in Gaussian semiparametric estimation of long range dependence. In: Robinson, P.M., Rosenblatt, M. (Eds.), Athens Conference on Applied Probability and Time Series in memory of Hannan, E.J. Vol. II.

Hurvich, C.M., 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., 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.

Kim, C., Phillips, P.C.B., 2000. Modified log periodogram regression. Department of Economics, Yale University.
Künsch, H., 1986. Discrimination between monotonic trends and long-range dependence. Journal of Applied Probability 23, 1025–1030.
Phillips, P.C.B., 1999. Discrete Fourier transforms of fractional processes. Cowles Foundation, Yale University.
Robinson, P.M., 1995a. Log-periodogram regression of time series with long range dependence. Annals of Statistics 23, 1048–1072.
Robinson, P.M., 1995b. Gaussian semiparametric estimation of long range dependence. Annals of Statistics 23, 1630–1661.
Robinson, P.M., Henry, M., 2003. Higher-order kernel semiparametric M-estimation of long memory. Journal of Econometrics 114 (1), 1–27.
Shimotsu, K., Phillips, P.C.B., 2001. Exact local Whittle estimation of fractional integration. Cowles Foundation, Yale University.
Sun, Y., Phillips, P.C.B., 2002. Understanding the Fisher Equation. Department of Economics, University of California, San Diego.
Velasco, C., 2000. Non-Gaussian log-periodogram regression. Econometric Theory 16, 44–79.