Journal of Econometrics 182 (2014) 309–328

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Modified local Whittle estimator for long memory processes in the presence of low frequency (and other) contaminations☆

Jie Hou[a], Pierre Perron[b,*]

[a] International School of Economics and Management, Capital University of Economics and Business, Beijing 100070, PR China

[b] Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215, United States

ARTICLE INFO

Article history:
Received 15 March 2013
Received in revised form
23 January 2014
Accepted 7 May 2014
Available online 23 May 2014

JEL classification:
C22
C13
C14

Keywords:
Long memory process
Random level shifts
Short memory dynamics
Additive noise
Local-Whittle estimators

ABSTRACT

We propose a modified local-Whittle estimator of the memory parameter of a long memory time series process which has good properties under an almost complete collection of contamination processes that have been discussed in the literature, mostly separately. These contaminations include processes whose spectral density functions dominate at low frequencies such as random level shifts, deterministic level shifts and deterministic trends. We show that our modified estimator has the usual asymptotic distribution applicable for the standard local Whittle estimator in the absence of such contaminations. We also show how the estimator can be modified to further account for additive noise and that our modification for low frequency contamination reduces the bias due to short-memory dynamics. Through extensive simulations, we show that the proposed estimator provides substantial efficiency gains compared to existing semiparametric estimators in the presence of contaminations, with little loss of efficiency when these are absent.

1. Introduction

Processes that are persistent in the sense that the serial correlation between distant observations decay hyperbolically are called long memory processes. They have found extensive use in capturing the behavior of many observed series since their introduction by Hurst (1951). A long memory process is also characterized in the frequency domain by a spectral density function proportional to λ^{-2d} as the frequency λ approaches zero at a rate dictated by the memory parameter d. In terms of parametric modeling, Granger and Joyeux (1980) and Hosking (1981) introduced the fractionally integrated $ARFIMA(p, d, q)$ model, a long-memory generalization of the short-memory $ARMA(p, q)$ process.

The estimators of the memory parameter are divided into parametric and semi-parametric ones. The theory of parametric estimators was developed by Fox and Taqqu (1986) and Dahlhaus (1989), among others. Semiparametric estimators of the memory parameter have become popular since they do not require knowing the specific form of the short memory structure. They are based on the periodograms of the series, and can be categorized into two types: the log-periodogram (LP) estimator first proposed by Geweke and Porter-Hudak (1983) and the local-Whittle (LW) estimator which is credited to Kunsch (1987). The LP estimator is akin to OLS and the LW estimator to the MLE in the frequency domain. Robinson (1995a,b) analyzed the asymptotic properties of these two types of estimators. He showed that they are asymptotically normal, have the same convergence rate and that the asymptotic variance of the LW estimator is smaller than that of the LP estimator. There are, however, so-called contaminations that have an effect on the bias and efficiency of these semi-parametric estimators, either in finite samples or even asymptotically. Much of the literature so far has focused on providing methods to mitigate the effect of additive noise and/or short-memory dynamics, which have only a finite sample effect. In the case of additive noise or so-called perturbed fractional processes, although both the LW and LP estimators preserve consistency and asymptotic normality, as shown by Deo and Hurvich (2001) and Arteche (2004), they can be severely biased. Hurvich and Ray (2003), Hurvich et al. (2005) and Arteche (2006), among others, have proposed estimators that can reduce the effect of noise by introducing an

☆ We are especially grateful to Zhongjun Qu for helpful comments, as well as to two referees.

* Corresponding author.
E-mail addresses: houjie15@gmail.com (J. Hou), perron@bu.edu (P. Perron).

http://dx.doi.org/10.1016/j.jeconom.2014.05.004
0304-4076/

additive constant or a polynomial term in the spectral density function. These methods are all based on local Whittle estimators, given their flexibility in accommodating more structures in the specified data-generating process. The estimators are also strongly biased when substantial short-memory dynamics are present. Among others, Andrews and Sun (2004) considered an adaptive local polynomial Whittle estimator. By substituting a polynomial structure for the constant term used to approximate the behavior of the short memory component near frequency zero in the local Whittle estimator, they showed that their estimator has considerable efficiency gains compared to classic LW and LP estimators under the presence of short memory dynamics. Recently, Frederiksen et al. (2012) combined the two methods and proposed estimators that can simultaneously reduce the bias and mean squared error caused by short memory dynamics and noise perturbation.

There are other low frequency contaminations (denoted as LFC) that can have a more serious effect causing outright inconsistent estimates. They may be important enough to induce researchers to mistakenly conclude that a short memory process with low frequency contaminations is actually a long memory process. Such an effect is often called "spurious long memory". These low frequency contaminations include, but are not confined to, random level shifts, deterministic level shifts and deterministic trends. A short-memory process contaminated by those components will exhibit hyperbolically decaying autocorrelations as well as a pole in its spectral density function at frequency zero, which are characteristics of a long memory process. Among others, Diebold and Inoue (2001), Granger and Hyung (2004), Mikosch and Stărică (2004) and Perron and Qu (2010) provide theoretical explanations for and simulation evidence of this spurious long memory effect. It has also been argued that models incorporating a short memory process with such low frequency contaminations provide a better in-sample fit and, in particular, forecast better compared to models assuming a pure long memory process. Various studies reported evidence that these forms of data contaminations are in fact very likely present in the volatility of asset prices and considerably weakens the evidence of pure long-memory; see, e.g., Granger and Hyung (2004), Mikosch and Stărică (2004), Stărică and Granger (2005), Perron and Qu (2010), Lu and Perron (2010), Qu and Perron (2013), Varneskov and Perron (2013) and Xu and Perron (2014).

Recent work by Dolado et al. (2005), Ohanissian et al. (2008), Perron and Qu (2010) and Qu (2011) proposed tests in both the time and the frequency domain with varying degrees of success. Many have argued that the long-memory properties of many economic time series are indeed spurious. These tests focus on distinguishing between a short memory process affected by low frequency contaminations from a true long memory process. So they do not offer methods to estimate the memory parameter in the presence of low frequency contaminations when the true signal may be of long or short memory.

Recently, attention focused on providing modified LP or LW estimators to account for low frequency contaminations. McCloskey and Perron (2013) proposed trimmed LP estimators that have desirable asymptotic and finite sample properties in the presence of low frequency contaminations. Using a similar trimming technique, McCloskey and Hill (2013) proposed trimmed frequency domain quasi maximum likelihood estimators for short-memory time series models (e.g., ARMA, GARCH and stochastic volatility models) that may be contaminated by low frequency movements. McCloskey (2013) considered a trimmed frequency domain quasi maximum likelihood estimator that can be used to consistently estimate the parameters of a long-memory stochastic volatility model in the presence of low frequency contamination assuming the signal to be an $ARFIMA(p, d, q)$ process. Iacone (2010) considered trimmed LW estimators.

We propose modified LW estimators that work under all kinds of contaminations: low frequency, additive noise and short memory dynamics. Our emphasis is on accounting for low frequency contaminations and we show how to further modify the estimator to account for the other types. It adopts techniques used in Andrews and Sun (2004), Hurvich et al. (2005) and Frederiksen et al. (2012) to introduce additive terms in the frequency domain quasi maximum likelihood function to capture the effect of the low frequency contaminations, based on results of Perron and Qu (2010) and McCloskey and Perron (2013) showing the spectral density function of low frequency contaminations to be of order $O_p(T^{-1}\lambda_k^{-2})$ near frequency zero. To account for additive noise, we follow Hurvich et al. (2005). Interestingly, our modification for low frequency contaminations also reduces the finite sample bias induced by short-memory dynamics, so that no further modification is necessary for this case.

Our modified estimators have the following advantages: being semiparametric, they do not require knowing the structure of the short memory process; they do not require trimming so all data is used; unlike the trimmed LP estimator, they do not require the underlying process to be Gaussian; they have the same asymptotic variance as the standard LW estimator when no contamination is present; without low frequency contaminations, they are asymptotically equivalent to the standard LW estimator that does not account for low frequency contaminations so that no efficiency loss is incurred by incorporating our modifications; they can easily be extended to a full parametric case. When low frequency contaminations are present, it has, in most cases, the smallest bias and mean-squared error amongst all existing estimators designed to control for low frequency contaminations, whether or not other types of contaminations are present. To our knowledge, our contribution is the first to provide an estimator with good properties under all previously considered contaminations: low frequency, additive noise and short-memory dynamics.

The structure of the paper is as follows. Section 2 presents the model and some preliminary results. Section 3 motivates and introduces our modified LW estimator that accounts for possible low frequency contaminations. Section 4 presents results about the consistency and limit distribution. Section 5 discusses how to extend the estimator to account for additive noise and short-memory dynamics. Section 6 presents the results of simulations to assess the finite sample properties under a variety of possible scenarios. Section 7 provides brief concluding remarks. All technical derivations are collected in a mathematical appendix.

The following notation is used throughout: "$\xrightarrow{d}$" stands for convergence in distribution, "$\xrightarrow{p}$" for convergence in probability, "$\to$" for the limit as $T \to \infty$ (unless otherwise stated), "$a \curlyvee b$" denotes the maximum of a and b, "$x \sim y$" means that $x/y \xrightarrow{p} 1$.

2. The model and preliminary results

We start with some basic definitions of a long memory process. Let $\{y_t\}_{t=1}^T$ be a stationary time series with spectral density function $f_y(\lambda)$ at frequency λ given by

$$f_y(\lambda) = G(\lambda)\lambda^{-2d} \quad \text{as } \lambda \to 0_+ \tag{1}$$

with $G(\lambda)$ a slowly varying function as $\lambda \to 0_+$ (i.e., for any real t, $G(t\lambda)/G(\lambda) \to 1$ as $\lambda \to 0_+$). When $d > 0$, y_t is a long-memory process with a spectral density function that increases for frequencies that get close to zero. The rate of divergence to infinity depends on the parameter d. Under some general conditions, this low-frequency definition is equivalent to the following long-lag autocorrelation definition (Beran, 1995). Let $\gamma_y(\tau)$ be the autocorrelation function of y_t. If $\gamma_y(\tau) = c(\tau)\tau^{2d-1}$ as $\tau \to \infty$, with $c(\tau)$ a slowly varying function as $\tau \to \infty$, the process is said to have long memory. For $0 < d < 1/2$, this implies that the autocorrelations decrease to zero at a slow hyperbolic rate which depends on the parameter d, in contrast to the fast geometric rate of decay that applies to a short-memory process. Examples of long-memory processes include the popular class of fractionally integrated autoregressive moving average models, though in what follows we shall remain agnostic about the nature of the short-memory component imposing only high level assumptions. When $d = 0$, y_t is a short-memory process.

The Data Generating Process (DGP) considered is one where the series of interest, z_t, is a long or short-memory process plus some low frequency contamination, viz.,

$$z_t = c + y_t + u_t \tag{2}$$

where y_t is a process with memory parameter $d \in [0, 1/2)$ and c a constant. Note that the value $d = 0$ is allowed so the DGP includes a short-memory process contaminated by some low frequency component. The process u_t is the low frequency contamination which will be defined below. We suppose that a sample of size T is available. We define the periodograms of the processes $\{z_t, y_t, u_t\}$ to be, for some frequency ordinate λ_k, $I_{z,k} = I_k = I_z(\lambda_k)$, $I_{y,k} = I_y(\lambda_k)$ and $I_{u,k} = I_u(\lambda_k)$ where $I_w(\lambda) = (2\pi T)^{-1}|\Sigma_{t=1}^{T} w_t e^{it\lambda}|^2$ for $w = z, y, u$, and their spectral density functions by $f_{z,k} = f_k = f_z(\lambda_k)$, $f_{y,k} = f_y(\lambda_k)$ and $f_{u,k} = f_u(\lambda_k)$. Semiparametric frequency domain estimators for non-contaminated fractional processes are all based on the local approximation (1) and are robust to the nature of the short memory dynamics since they only use information from periodogram ordinates near the origin.

The local Whittle (LW) estimation method of Kunsch (1987) and Robinson (1995a) has become popular because of its likelihood interpretation, nice asymptotic properties (smaller asymptotic variance compared to log-periodogram estimators), mild assumptions (e.g., no need for a normality assumption) and most importantly in our case, the possibility to easily modify it to accommodate the presence of contaminations. It is defined as the minimizer of the (negative) local Whittle likelihood function in the frequency domain

$$Q(G_0, d) = \frac{1}{m}\sum_{j=1}^{m}[\log(G_0\lambda_j^{-2d}) + I_z(\lambda_j)/(G_0\lambda_j^{-2d})]$$

where $G_0 = G(0)$, $m = m(T)$ is the bandwidth which goes to infinity as $T \to \infty$ but at a slower rate than T, $\lambda_j = 2\pi j/T$ are the Fourier frequencies. Concentrating with respect to G_0, the estimator of d is $\hat{d}_{LW} = \arg\min_d[\log \hat{G}_0(d) - 2dm^{-1}\sum_{j=1}^{m}\log\lambda_j]$, where $\hat{G}_0(d) = m^{-1}\sum_{j=1}^{m}\lambda_j^{2d} I_z(\lambda_j)$. The types of processes considered for the low frequency contamination (LFC) component u_t are laid out in the following definition.

Definition 1. The low frequency contamination component u_t is generated by one of the following processes. (1) Random level shifts (RLS): $u_t = \Sigma_{t=1}^{T}\delta_{T,t}$ where $\delta_{T,t} = \pi_{T,t}\eta_t$ with $\eta_t \sim i.i.d.\ N(0, \sigma_\eta^2)$ and $\pi_{T,t} \sim i.i.d.\ Bernoulli(p/T, 1)$ for some $p \geq 0$. The components $\pi_{T,t}$, η_t are mutually independent. (2) Deterministic level shifts: $u_t = \Sigma_{i=1}^{B} c_i\chi(T_{i-1} < t \leq T_i)$ where B is the (fixed) number of regimes ($B - 1$ is the number of breaks), $0 < |c_i| < \infty$, $\chi(\cdot)$ is the indicator function, $0 = T_0 < T_1 < \cdots < T_{B-1} < T_B = T$ and $T_i/T \to \tau_i$ with $0 < \tau_1 < \cdots < \tau_{B-1} < 1$. (3) Deterministic trends: $u_t = h(t/T)$ where $h(\cdot)$ is a deterministic nonconstant function on $[0, 1]$ that is either Lipschitz continuous or monotone with $h(1) = 0$.[1] (4) Fractional trends: $u_t = O((t+1)^{\phi-1/2})$, $u_0 = 0$, $|u_{t+1} - u_t| = O(|u_t|/t)$ where $\phi \in (-1/2, 1/2)$.

[1] This includes all cases for which $h(\cdot)$ is monotonic and bounded because we can simply subtract $h(1)$ from $h(\cdot)$ and add $h(1)$ to c in (2) to have the same DGP.

Note that the probability of a level shift in the RLS model is sample size dependent. If this were not the case, u_t would have properties similar to that of a random walk. A defining characteristic of the RLS model is that the average number of level shifts p remains constant as the sample size grows. Note that p can be zero so that the assumption nests the no level shift or no contamination case as well. Perron and Qu (2010) considered the asymptotic properties of the periodogram of this type of process contaminating a short memory process and showed that, for any $k = 1, \ldots, [T/2]$, $(k^2/T)E(I_{u,k}) \to (p\sigma_\eta^2)/(4\pi^3)$ as $T \to \infty$. Mikosch and Stărică (2004) considered the asymptotic properties of the periodogram for a deterministic level shift component when $B = 2$ (one level shift), with the addition of a short-memory component, and showed that $E(I_{u,k}) = O(T/k^2)$. Kunsch (1987, Lemma 2) considered the asymptotic properties of the periodogram of a short-memory process contaminated by a bounded monotone trend. Qu (2011, lemma1) extended Kunsch's results to the Lipschitz continuous case and showed that $E(I_{u,k}) = O(T/k^2)$. Iacone (2010) discussed the order of the periodogram in the case of a fractional trend and showed that $E(I_{u,k}) = O_p(T/k^2)$.

The common feature of these contaminating processes is that the mean of their periodogram near frequency zero is of order $O(T/k^2)$, or equivalently of order $O(T^{-1}\lambda_k^{-2})$ since $\lambda_k = 2\pi k/T$ (note that the O term could be o since it is possible that $E[(I_{u,k})/(T/k^2)] \to 0$, a case we shall discuss further later). Processes with such LFC as additive components are non-stationary so they do not have the traditionally defined spectral density function. Following common practice in such cases, we define their spectral density function to be the expectation of their periodogram. Since the spectral density function of a long memory process near frequency zero is of order $O(\lambda_k^{-2d})$, in general the spectral density function of such contaminating components dominates that of a long memory process at low frequencies and vice-versa at high frequencies. Note that in the representation (1), when the process is contaminated by such LFC, we have $G_u \equiv G_u(0) = \lim_{T\to\infty}(k^2/T)E(I_{u,k})$.

Remark 1. Definition 1 could be replaced by the condition $E(I_{u,k}) = G_u(k)(T/k^2)(1 + O(1))$ with $G_u(k) \leq B$ where B is a fixed bounded positive constant. Hence, $E(I_{u,k})/(T/k^2)$ need not converge to a constant, it only needs to be bounded as $T \to \infty$. All LFC in Definition 1 satisfy this property and all results to be presented remain valid under this general condition.

Unlike short memory dynamics or contaminating noise, which cause only finite sample biases to the memory parameter estimator, the bias caused by LFC usually remains asymptotically. To see when this applies, let $A_k = (k^2/T)E(I_{u,k})$, then one can show that $\lambda_k^{2d}I_k = \lambda_k^{2d}I_{y,k} + A_k O_p(T^{1-2d}/k^{2-2d})$. So the bias introduced by LFC is of order $O_p(m^{-1}T^{1-2d}\sum_{k=1}^{m}(A_k/k^{2-2d}))$. The following definition will be useful.

Definition 2. A LFC is said to be non-degenerate if $\lim_{T\to\infty}\{(k^2/T)E(I_{u,k})\} > 0$ for every k. Otherwise it is said to be degenerate.

An example of a non-degenerate LFC is a RLS model, in which case $\lim_{T\to\infty}(k^2/T)E(I_{u,k}) = (p\sigma_\eta^2)/(4\pi^3)$. An example of a degenerate LFC is a monotone deterministic trend. The bias caused by a non-degenerate LFC remains asymptotically while the bias caused by a degenerate LFC can either remain or vanish asymptotically, with the degree of the (potentially asymptotic or finite sample) bias depending on d and the bandwidth m.

3. The modified local Whittle estimator

Let the Fourier transform of the process z_t be $h_z(\lambda_j) = (2\pi T)^{-1/2}(\sum_{t=1}^{T} z_j e^{-it\lambda_j})$ so that $f_z(\lambda_k) = E(I_z(\lambda_k)) = E(h_z(\lambda_k) h_z(\lambda_k)^*)$, where "$*$" denotes the complex conjugate value. One may then define the frequency domain pseudo Quasi Maximum Likelihood Function (QMLF) for $h_z(\lambda_k)$ as $\varphi_k = \log(f_z(\lambda_k)) + I_z(\lambda_k)/f_z(\lambda_k)$. When there is no contamination in the data, $f_z(\lambda_k)$ reduces to $f_y(\lambda_k)$ and the standard LW estimator is the minimizer of the pseudo-QMLF. With low frequency contamination given by u_t, a problem is how to construct a useful approximation to $f_z(\lambda_k)$ in such cases. Because the periodogram of u_t is of order $O_p(T^{-1}\lambda_k^{-2})$, a sensible strategy is to add a term $(G_u/T)\lambda_k^{-2}$ to the spectral density function of y_t to control for the low frequency contamination. Accordingly, we consider the pseudo spectral density function $f_k \triangleq f_z(\lambda_k) = G_0\lambda_k^{-2d} + G_u\lambda_k^{-2}/T$. Let $\theta = (G_u/G_0)$ be the signal to noise ratio, the pseudo spectral density function of the observed process is then:

$$\begin{aligned} f_k \triangleq f_z(\lambda_k) &= G_0\lambda_k^{-2d} + G_u\lambda_k^{-2}/T = G_0(\lambda_k^{-2d} + (G_u/G_0)\lambda_k^{-2}/T) \\ &= G_0(\lambda_k^{-2d} + \theta\lambda_k^{-2}/T) = G_0 g_k \end{aligned}$$

where $g_k = (\lambda_k^{-2d} + \theta\lambda_k^{-2}/T)$.

Remark 2. f_k is the "pseudo spectral density function" in the sense that it is not the true spectral density function of the data, but an artificial construct aimed at providing a good approximation to the behavior of the generalized spectral density function (i.e., the expectation of the periodogram) and an extended LW type estimator with desirable properties.

This pseudo spectral density function can then be used to approximate $E(I_{z,k})$ and the pseudo frequency domain QMLF is $\varphi(G, d, \theta) = m^{-1}\sum_{k=1}^{m}\varphi_k(G, d, \theta)$. Using the same technique as in Robinson (1995a), we can concentrate G out of the QMLF using $\hat{G} = m^{-1}\sum_{k=1}^{m}(I_k/g_k)$. Hence, the local Whittle (frequency domain QMLE) estimator applicable under LFC, denoted as the LWLFC estimator, is $(\hat{d}_m, \hat{\theta}_m) = \arg\min_{(d,\theta)} J_m(d, \theta)$, where

$$J_m(d, \theta) = \log\left(m^{-1}\sum_{k=1}^{m}(I_k/g_k) + m^{-1}\sum_{k=1}^{m}\log(g_k)\right).$$

Remark 3. The component θ is an "auxiliary variable" in the sense that it is not a parameter of primary interest but is introduced as a tool used to control the influence of the contaminations at low frequencies. Intuitively, θ is the appropriate signal to noise ratio to use as it measures the average of the relative magnitude of the contaminations across all frequencies. For the case of RLS contamination, we have an expression for θ in terms of the parameters of the model, given by $\theta = (G_u/G_0) \sim (2\pi\sigma_\eta^2/\sigma_\varepsilon^2)$; see Perron and Qu (2010).

Remark 4. The method can be extended to the case with a parametric specification for the long-memory process. For example, if y_t is assumed to follow the *ARFIMA*(p, d, q) process $(1-L)^d y_t = \tilde{y}_t$, where $A(L)\tilde{y}_t = B(L)\varepsilon_t$ and $\varepsilon_t \sim i.i.d.\ N(0, \sigma_\varepsilon^2)$, then we simply replace $G_0\lambda_k^{-2d}$ by $\sigma_\varepsilon^2(|B(e^{-i\lambda})|^2/|A(e^{-i\lambda})|^2)[2\pi|1 - e^{-i\lambda}|^{2d}]^{-1}$ in the objective function $J_m(d, \theta)$.

4. Asymptotic properties

We start by introducing the assumptions required to obtain the consistency result for the LWLFC estimator. Many are the same as in Robinson (1995a), but some are added or modified to accommodate the LFC components. Henceforth, we shall denote the true value of the long-memory parameter by d_0 and the true value of the signal-to-noise ratio by θ_0.

- Assumption A1. As $\lambda \to 0_+, f_y(\lambda) \sim G_0\lambda^{-2d_0}$ where $G_0 \in (0, \infty)$ and $d_0 \in [0, 1/2)$.
- Assumption A2. For λ in a neighborhood of 0, $f_y(\lambda)$ is differentiable and $d\log(f_y(\lambda))/d\lambda = O(\lambda^{-1})$.
- Assumption A3. y_t is stationary and admits an infinite MA representation: $y_t - E(y_t) = \sum_{j=0}^{\infty}\alpha_j\varepsilon_{t-j}$ with $\sum_{j=0}^{\infty}\alpha_j^2 < \infty$ where $\{\varepsilon_t\}$ is a martingale difference sequence with $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_\varepsilon^2$, $E(\varepsilon_t^3|\mathcal{F}_{t-1}) = \mu_3$, and $E(\varepsilon_t^4) = \mu_4$ where $\mathcal{F}_t$ is the σ-field generated by $\{\varepsilon_s; s \le t\}$. Also, there exists a random variable ε such that $E(\varepsilon^2) < \infty$ and for all $\eta > 0$ and some $K > 0$, $P(|\varepsilon_t| > \eta) \le KP(|\varepsilon| > \eta)$.

Remark 5. We require ε_t to have finite fourth moment even to establish consistency to invoke a strong law of large numbers for $m^{-1}\sum_{k=1}^{m}(I_k/g_k(d, \theta))$ and show that the convergence of the memory parameter estimate does not depend on the signal to noise ratio.

- Assumption A4. As $T \to \infty$, $T^{(1-(d_0^2-3d_0+9/4)^{-1})\curlyvee(1/2)}/m + m/T \to 0$.

Remark 6. The requirement on the bandwidth to establish consistency departs from Robinson (1995a) who only requires that $(1/m)+(m/T) \to 0$. This is due to the need to suppress the impact of (I_k/g_k) at low frequencies, $k < T^{[(1-2d_0)/(2-d_0)]}$, in which case the periodogram of the LFC dominates that of the long memory process. With the addition of the term $(\theta/T)\lambda_k^{-2}$ in the QMLF, we can then bound $|I_k/g_k|$. However, to control the effect of $\{I_k/g_k\}$ at high frequencies where the periodogram of the long memory process dominates that of the LFC, we need a larger bandwidth to suppress the cumulative impact from the low frequencies. The closer is d_0 to 0, the higher is the required bandwidth because the contamination will then dominate at higher frequencies. The quantity $(1-(d_0^2-3d_0+9/4)^{-1}) \curlyvee (1/2)$ achieves its maximum value 5/9 when $d_0 = 0$. Hence, in practice with an unknown memory parameter d_0, we need to choose a bandwidth of order greater than $T^{5/9}$.

- Assumption A5. u_t is one of the LFC as stated in Definition 1.

It will be useful to first establish a limit result pertaining to the estimate $\hat{\theta}_m$ of the signal to noise ratio. This will be used in the proof of the consistency of $\hat{d}_m$.

Lemma 1. *Under* A1–A5*, if a non-degenerate LFC is present, $\hat{\theta}_m$ is bounded above by zero.*

We now consider the consistency result and a preliminary bound on the convergence rate that will be used to establish the limit distribution of our estimator.

Theorem 1. *Under* A1–A5*: (a)$\hat{d}_m \xrightarrow{p} d_0$ as $T \to \infty$; (b)$|\hat{d}_m - d_0| = o_p((\log(m))^{-3})$.*

Note that this result does not require $\hat{\theta}_m$ to be a consistent estimate, all that is required is that if LFC components are present the probability limit of the estimate of $\hat{\theta}_m$ is bounded above by zero, which is guaranteed by Lemma 1. This implies that with probability arbitrarily close to one, $\hat{\theta}_m$ will be in a set $(0, \infty)$ and we can consider analyzing the limit of $\hat{d}_m$ for any value or sequences of θ_m in the set $(0, \infty)$.

Before proceeding further, we need to discuss a property of the estimate of the signal-to-noise ratio $\hat{\theta}_m$ when there is no LFC present. This, in conjunction with Lemma 1, will allow us to derive the limit distribution of $\hat{d}_m$ for both cases with and without LFC. The required result is stated in the next lemma, which is of independent interest.

Lemma 2. *Suppose no LFC is present and that* A1–A4 *hold, then, as* $T \to \infty$: $\hat{\theta}_m = O_p(T^{-(1-2d_0)/(2-2d_0)}) \to 0$.

To prove the asymptotic normality of $\hat{d}_m$, further assumptions are needed, some of which are strengthened versions of Assumptions A1–A3.

• Assumption A6. For some $\tau \in (0, 2]$, $f_y(\lambda) \sim G_0\lambda^{-2d_0}(1+O(\lambda^{\tau}))$ as $\lambda \to 0_+$, where $G_0 \in (0, \infty)$ and $d_0 \in [0, 1/2)$.

• Assumption A7. In a neighborhood of the origin, $f_y(\lambda)$ is differentiable and $df_y(\lambda)/d\lambda = O(f_y(\lambda)/\lambda)$ as $\lambda \to 0_+$.

• Assumption A8. As $T \to \infty$, $m^{-1} + T^{-2\tau}m^{1+2\tau}(\log m)^2 \to 0$.

The following theorem states the asymptotic distribution of the estimate $\hat{d}_m$.

Theorem 2. *Under* A1–A8: $m^{1/2}(\hat{d}_m - d_0) \xrightarrow{d} N(0, 1/4)$ *as* $T \to \infty$.

Note that the asymptotic variance of our estimator is the same as that of the standard LW estimator of Robinson (1995a) applicable with no LFC. The intuitive reason is that, asymptotically, the additional term $G_u(\lambda_k^{-2}/T)$ controls the effect of LFC on the spectral density function well enough so that no efficiency loss ensues.

When the magnitude of the LFC is weak, the asymptotic distribution of Theorem 2 provides a good approximation to the finite sample distribution. However, when the magnitude of the LFC is substantial, $2m^{1/2}(\hat{d}_m - d_0)$ does converge to a normal distribution rapidly as T increases (even with T as small as 512) but the approach to a standard normal may be slow, i.e., the mean and variance of $2m^{1/2}(\hat{d}_m - d_0)$ may converge slowly to 0 and 1, respectively. Some approximate formulas to compute the finite sample bias and variance of $2m^{1/2}(\hat{d}_m - d_0)$ have been found in unreported simulations and they provide good approximations. Unfortunately, they all depend on θ_0, the signal to noise ratio which cannot be identified when it is greater than zero, rendering the corrections not applicable in practice. An important avenue of further research is to obtain a finite-sample scaling factor, say S, to replace m in order to obtain good finite sample coverage rates for the LWLFC estimate. A conjecture is that S should be a decreasing function of θ_0 to reflect the impact of LFC on the variance of the memory parameter estimate. But since $\hat{\theta}_m$ is not a consistent estimator of θ_0, it is unlikely that one can find a good applicable formula. This problem about the coverage rate is not unique to our method, and applies to all existing methods to estimate the memory parameter under some contamination. Alternative scaling factors have been proposed. For the log-periodogram estimator, Geweke and Porter-Hudak (1983) suggested using the scaling factor $S(l, m)^{1/2}$, where $S(l, m) = \sum_{j=1}^{m}(\log j - (m - l + 1)^{-1}\sum_{\tau=l}^{m}\log\tau)^2$ for some lower trimming l, and its use was also discussed by Deo and Hurvich (2001). For local Whittle-type estimators, it was used by Hurvich et al. (2005) and Iacone (2010).

5. Extension to the case of additive noise and short memory dynamics

An advantage of LW-type estimators is that, since they use the QMLF in the frequency domain, they can easily be modified to accommodate more types of structures in the DGP, without the need to trim some of the low frequencies. We consider two extensions to account for additive noise and short-memory dynamics. These elements do not cause an asymptotic bias and, hence, the modifications are aimed solely at improving the finite sample performance. Consider first the case where both LFC and additive noise are to be accounted for. To be precise, instead of (2), the DGP is now $z_t = c + y_t + u_t + w_t$, where, following Assumption H2 in Hurvich et al. (2005), the additive noise w_t is a zero mean white noise with variance σ_w^2, such that for each $s \neq t$, $E[w_s\varepsilon_t] = 0$ and for each t, $E[w_t\varepsilon_t] = \rho_w\sigma_w$, where ε_t is as defined in A3 and ρ_w is the correlation between w_t and ε_t, assumed to be constant. Also, w_t is independent of the LFC u_t. Following Hurvich et al. (2005), we add a constant term into the spectral density function, so that the modified pseudo spectral density function is:

$$\begin{aligned} f_k &\triangleq f_z(\lambda_k) = G_0\lambda_k^{-2d} + G_w + G_u(\lambda_k^{-2}/T) \\ &= G_0(\lambda_k^{-2d} + (G_w/G_0) + (G_u/G_0)(\lambda_k^{-2}/T)) \\ &= G_0(\lambda_k^{-2d} + \theta_w + (\theta_u/T)\lambda_k^{-2}) = G_0 g_k \end{aligned} \quad (3)$$

where, with a slight abuse of notation relabeling $\theta_u = G_u/G_0$, $g_k = (\lambda_k^{-2d} + \theta_w + (\theta_u/T)\lambda_k^{-2})$ and the (approximate) frequency domain QMLF is $\varphi(G, d, \theta) = m^{-1}\sum_{k=1}^{m}\varphi_k(G, d, \theta)$ with $\theta = (\theta_w, \theta_u)'$. Concentrating G out of the QMLF, the estimate of G is $\hat{G} = m^{-1}\sum_{k=1}^{m}(I_k/g_k)$ and the local Whittle QMLE estimator under noise perturbations and low frequency contaminations, denoted as the LWPLFC estimator, is $(\hat{d}_m, \hat{\theta}_m) = \arg\min_{(d,\theta)} J_m(d, \theta)$, where

$$J_m(d, \theta) = \log\left(\frac{1}{m}\sum_{k=1}^{m}\frac{I_k}{g_k}\right) + \frac{1}{m}\sum_{k=1}^{m}\log(g_k).$$

For reasons discussed by Hurvich et al. (2005), the LWPLFC approach is expected to work when d_0 is not too close to zero. When $d_0 = 0$, the process is short-memory. We then have a combination of two additive short-memory processes which cannot be identified separately.

For the case of short memory dynamics plus LFC, we could follow the approach of Andrews and Sun (2004) who add a polynomial structure into G_0, i.e., replace G_0 in (3) by $G_0\exp(-p_r(\lambda_j, \theta))$ where $p_r(\lambda_j, \theta) = \Sigma_{s=1}^{r}\theta_s\lambda_j^{2s}$ and $\theta = (\theta_1, \ldots, \theta_r)$. However, unreported simulations with $r = 1$, showed that doing so did not offer any gain in performance over our LWLFC estimator with a smaller value of the bandwidth (see the simulations in Section 6). This feature can be explained as follows. From simulations to be reported in the next section, under strong short memory dynamics and RLS, the LWLFC estimator constructed with a large bandwidth has substantial bias but very small variance, so that the overall MSE is almost entirely due to the bias. When a polynomial component is added, the upward bias is reduced but the variance is increased considerably so that the overall MSE is almost the same or larger than that of the LWLFC estimator. With no RLS, the increased variance is smaller so that the MSE is indeed reduced as reported by Andrews and Sun (2004). At the root of the issue is the fact that both RLS and short memory dynamics cause upward biases in the estimate of the memory parameter. Hence, there is a confounding effect so that the QMLF is flat with respect to the correction factors for short memory dynamics and LFC. In unreported simulations with both RLS and short memory dynamics, it was often found that either the coefficient to correct for short memory dynamics or the coefficient to account for LFC was very close to zero, despite having the true value of both coefficients greater than zero. As will be reported in the simulations, the best way to account for short memory dynamics and RLS is to use the LWLFC estimator with a small bandwidth.

When both additive noise and short-memory dynamics are to be accounted for, three approaches are possible. One is to use the LWLFC estimator with a small bandwidth, another is to use the LWPLFC with a large bandwidth, or we could follow the approach of Frederiksen et al. (2012) who add polynomials and a constant as additive terms in the QMLF. One drawback of the latter approach

Table 1
Bias and RMSE for a short memory process $ARFIMA(\alpha, d = 0, 0)$ with RLS.

$T \setminus \beta$	$p = 0$			$p = 5$			$p = 10$			$p = 20$		
	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8
(a) Bias	$\alpha = 0$											
256	−0.087	−0.047	−0.021	−0.021	−0.038	0.004	0.004	−0.017	−0.002	−0.011	−0.059	0.023
512	−0.052	−0.025	−0.008	−0.016	−0.007	0.005	−0.014	−0.028	−0.012	−0.086	−0.055	0.011
1024	−0.037	−0.016	−0.006	−0.029	−0.001	0.003	−0.018	−0.001	−0.004	0.008	−0.026	0.001
2048	−0.013	−0.009	−0.006	−0.012	0.001	0.001	0.016	−0.006	0.005	0.013	−0.011	−0.005
4096	−0.007	−0.004	−0.004	−0.009	−0.006	−0.001	−0.032	−0.009	0.003	0.003	−0.009	0.000
	$\alpha = 0.3$											
256	−0.022	0.070	0.168	0.009	0.122	0.246	0.215	0.413	0.524	−0.012	0.131	0.270
512	−0.029	0.041	0.134	0.015	0.116	0.194	0.166	0.354	0.491	0.059	0.142	0.242
1024	−0.012	0.025	0.118	−0.009	0.048	0.137	0.123	0.267	0.436	0.023	0.053	0.160
2048	−0.005	0.027	0.092	0.004	0.033	0.107	0.072	0.206	0.383	−0.012	0.040	0.013
4096	−0.007	0.014	0.073	0.018	0.015	0.085	0.044	0.145	0.329	−0.011	0.024	0.100
	$\alpha = 0.6$											
256	0.128	0.299	0.432	0.176	0.387	0.498	0.213	0.414	0.524	0.221	0.449	0.561
512	0.093	0.223	0.392	0.135	0.307	0.459	0.178	0.344	0.492	0.174	0.388	0.517
1024	0.052	0.170	0.347	0.109	0.250	0.406	0.069	0.202	0.380	0.108	0.298	0.463
2048	0.020	0.125	0.307	0.066	0.186	0.361	0.063	0.143	0.330	0.085	0.234	0.409
4096	0.013	0.084	0.267	0.036	0.131	0.311	0.014	0.101	0.281	0.035	0.161	0.349
(b) RMSE	$\alpha = 0$											
256	0.191	0.111	0.078	0.423	0.291	0.140	0.546	0.327	0.162	0.656	0.421	0.231
512	0.132	0.075	0.046	0.353	0.169	0.082	0.402	0.197	0.121	0.575	0.362	0.104
1024	0.095	0.057	0.037	0.262	0.130	0.068	0.280	0.147	0.071	0.376	0.213	0.078
2048	0.068	0.041	0.026	0.215	0.089	0.044	0.274	0.108	0.046	0.316	0.131	0.052
4096	0.062	0.029	0.019	0.147	0.069	0.031	0.218	0.075	0.031	0.264	0.101	0.041
	$\alpha = 0.3$											
256	0.172	0.118	0.180	0.371	0.201	0.265	0.467	0.269	0.277	0.524	0.337	0.311
512	0.132	0.086	0.142	0.282	0.179	0.207	0.354	0.185	0.235	0.432	0.241	0.261
1024	0.082	0.058	0.122	0.157	0.087	0.142	0.203	0.096	0.154	0.276	0.117	0.172
2048	0.068	0.049	0.096	0.128	0.060	0.111	0.140	0.066	0.120	0.221	0.083	0.137
4096	0.049	0.035	0.076	0.102	0.041	0.088	0.115	0.047	0.094	0.150	0.060	0.102
	$\alpha = 0.6$											
256	0.194	0.312	0.437	0.357	0.407	0.504	0.385	0.442	0.530	0.482	0.493	0.569
512	0.145	0.233	0.395	0.237	0.322	0.463	0.300	0.370	0.497	0.340	0.414	0.522
1024	0.099	0.177	0.349	0.190	0.260	0.409	0.229	0.278	0.438	0.279	0.313	0.466
2048	0.066	0.132	0.308	0.136	0.194	0.362	0.167	0.216	0.385	0.203	0.245	0.411
4096	0.053	0.098	0.268	0.092	0.138	0.312	0.116	0.153	0.330	0.146	0.170	0.351

is that the increase in the number of parameters can induce an important increase in variance resulting in an increased mean-squared error.

6. Finite sample properties

The Data Generating Process (DGP) used for the simulations is $z_t = y_t + u_t + w_t$, where y_t is an $ARFIMA(1, d, 0)$ process given by $(1 - \alpha L)(1 - L)^d y_t = e_t$ with $e_t \sim i.i.d.\ N(0, 1)$, u_t is a RLS process as described in Definition 1 with $\sigma_\eta^2 = 1$, and $w_t \sim i.i.d.\ N(0, \sigma_w^2)$ is the additive noise component. The values used are: $d = 0, 0.2, 0.45$; $\alpha = 0.0, 0.3, 0.6$ and $p = 0, 5, 10, 20$. The sample sizes are $T = 256, 512, 1024, 2048$ and 4096 in order to use the fast Fourier transform algorithm with the whole data set. The estimate $\hat{d}_m$ is allowed to take values in the set $[-0.99, 0.99]$ when evaluating the maximizers of the objective function. The value of the bandwidth is set to $m = T^\beta$ for $\beta = 0.6, 0.7, 0.8$, the choice being dictated by the fact that β must be larger than 5/9. Throughout, 500 replications are used. These specifications were also used by McCloskey and Perron (2013) so that we can make direct comparisons of the relative performance of our estimators with theirs (the sample sizes they used are 1000 and 2000 but the minor differences in T should not be of concern given the rather large differences in performance). The trimmed LP estimator of McCloskey and Perron (2013) depends on a lower trimming and upper bandwidth, while ours depend on a bandwidth. We evaluate bias and Root Mean Squared Errors (RMSEs). When making comparisons, we do so using the values of the bandwidth (and trimming for the LP estimator) that gives the best RMSE for each of the statistics. We focus on random level shifts as the contaminating component as this is arguably the most relevant in practice. The results are presented in Tables 1–3 for the cases with only RLS and RLS plus short-memory dynamics, for which we focus on the LWLFC estimator. Table 4 presents the results for the case of RLS plus additive noise, while Table 5 presents results when all three types of contaminations are present, in which cases we consider both the LWLFC and LWPLFC estimators. We do not make a direct comparison with the trimmed LW estimator of Iacone (2010). McCloskey and Perron (2013) performed a comparison between the trimmed LP and LW estimators. They concluded that the trimmed LP has generally smaller bias and the trimmed LW generally lower variance and concluded that the overall performance in the presence of RLS was comparable.

6.1. The case with only RLS

The results for the case with only RLS are presented in the first panels of Tables 1–3 corresponding to the case $\alpha = 0$. Note first that the best results in terms of RMSE are obtained with a large bandwidth using $\beta = 0.8$, though biases are slightly smaller with

Table 2
Bias and RMSE for a long memory process $ARFIMA(\alpha, d = 0.2, 0)$ with RLS.

$T \setminus \beta$	$p = 0$			$p = 5$			$p = 10$			$p = 20$		
	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8
(a) Bias	$\alpha = 0$											
256	−0.106	−0.054	−0.033	−0.103	−0.050	−0.030	−0.093	−0.057	−0.053	−0.159	−0.094	−0.047
512	−0.050	−0.025	−0.020	−0.054	−0.011	−0.026	−0.107	−0.027	−0.025	−0.064	−0.052	−0.001
1024	−0.033	−0.019	−0.015	−0.025	−0.012	−0.036	−0.012	−0.014	−0.042	−0.054	−0.029	−0.039
2048	−0.016	−0.011	−0.012	−0.027	−0.008	−0.027	−0.016	−0.010	−0.031	−0.009	−0.013	−0.030
4096	−0.015	−0.008	−0.006	0.004	−0.009	0.020	−0.011	−0.006	−0.022	−0.006	−0.009	−0.022
	$\alpha = 0.3$											
256	−0.047	0.063	0.149	−0.080	0.101	0.203	0.021	0.114	0.214	−0.030	0.118	0.237
512	−0.017	0.034	0.126	−0.023	0.088	0.167	0.001	0.098	0.185	−0.019	0.102	0.220
1024	−0.014	0.019	0.102	−0.006	0.054	0.141	0.003	0.003	0.162	−0.035	0.073	0.172
2048	−0.022	0.023	0.088	−0.018	0.035	0.117	−0.020	0.040	0.128	−0.011	0.048	0.140
4096	−0.014	0.013	0.074	0.010	0.025	0.092	0.004	0.033	0.101	0.005	0.042	0.112
	$\alpha = 0.6$											
256	0.128	0.302	0.424	0.160	0.336	0.442	0.192	0.362	0.474	0.149	0.383	0.483
512	0.080	0.223	0.384	0.098	0.268	0.415	0.138	0.297	0.427	0.129	0.339	0.461
1024	0.045	0.170	0.348	0.078	0.209	0.365	0.091	0.227	0.380	0.118	0.251	0.412
2048	0.038	0.120	0.306	0.050	0.153	0.322	0.067	0.170	0.335	0.056	0.194	0.361
4096	0.015	0.086	0.267	0.031	0.114	0.277	0.039	0.125	0.292	0.026	0.137	0.309
(b) RMSE	$\alpha = 0$											
256	0.229	0.126	0.082	0.409	0.213	0.139	0.508	0.298	0.191	0.635	0.398	0.191
512	0.132	0.083	0.058	0.302	0.136	0.091	0.414	0.198	0.103	0.514	0.273	0.116
1024	0.097	0.057	0.040	0.205	0.076	0.059	0.244	0.068	0.057	0.370	0.088	0.059
2048	0.065	0.041	0.028	0.163	0.051	0.044	0.194	0.043	0.039	0.265	0.057	0.039
4096	0.053	0.033	0.021	0.109	0.037	0.032	0.148	0.029	0.028	0.205	0.039	0.029
	$\alpha = 0.3$											
256	0.174	0.114	0.161	0.330	0.201	0.221	0.379	0.226	0.243	0.485	0.316	0.274
512	0.128	0.085	0.137	0.256	0.146	0.179	0.292	0.173	0.198	0.409	0.203	0.235
1024	0.085	0.056	0.107	0.200	0.099	0.147	0.256	0.115	0.169	0.329	0.140	0.180
2048	0.078	0.045	0.091	0.114	0.070	0.121	0.170	0.079	0.132	0.202	0.097	0.145
4096	0.056	0.033	0.077	0.096	0.048	0.095	0.118	0.066	0.104	0.143	0.070	0.115
	$\alpha = 0.6$											
256	0.216	0.314	0.429	0.264	0.350	0.447	0.323	0.380	0.479	0.382	0.434	0.503
512	0.137	0.233	0.387	0.218	0.279	0.418	0.231	0.311	0.430	0.265	0.352	0.465
1024	0.098	0.180	0.350	0.135	0.217	0.367	0.171	0.238	0.382	0.139	0.262	0.414
2048	0.075	0.126	0.307	0.106	0.160	0.323	0.122	0.177	0.337	0.139	0.203	0.363
4096	0.051	0.090	0.268	0.077	0.120	0.278	0.093	0.132	0.293	0.099	0.143	0.310

a smaller bandwidth. Second, the results show that our estimator performs better than McCloskey and Perron's (2013) trimmed LP estimator. When $d_0 = 0$, there is a 30%–60% reduction in RMSE, when $d_0 = 0.2$ the reduction is in the range 30%–40% while when $d_0 = 0.45$ it is in the range 5%–20%. Hence, overall, the LWLFC estimator with a large bandwidth $\beta = 0.8$, shows smaller bias and RMSE than alternative estimators. When the process is uncontaminated ($p = 0$), the bias and RMSE of our estimator is small and close to that of the original LW estimator, so that very little efficiency loss is incurred when no contamination is present.

6.2. The case with RLS and short-run dynamics

We now consider the case with both RLS and short-run dynamics (presented in Tables 1–3 for non-zero values of α). In this case the best results for the LWLFC estimator are obtained with a small bandwidth, using $\beta = 0.6$, and more so as the magnitude of α increases. Compared to the trimmed LP estimator, the reduction in RMSE is very substantial especially for larger values of α. For example, with no RLS the reduction is around 65% when $d = 0$ and $\alpha = 0.6$, while it is around 40% when $d = 0.45$ and $\alpha = 0.6$. The LWLFC is able to reduce bias and variance when both RLS and short-run dynamics are present, even though it is designed to account only for LFC contamination. As discussed in Section 5, the approach of Andrews and Sun (2004) which adds a polynomial structure into G_0 does not offer additional improvement. As stated in the above discussion, the results show that the LWLFC estimator has indeed very small variance when both RLS and short-run dynamics are present.

6.3. The case with RLS and additive noise

The results for the case with RLS and additive noise are presented in Table 4 for the LWLFC (which accounts only for LFC) and LWPLFC estimators (which accounts for both). The variance of the noise is set to a large value $\sigma_w^2 = 4$. The results show that the LWPLFC estimator has very small biases irrespective of the choice of the bandwidth. The biases are indeed orders of magnitude smaller than those of the trimmed LP estimator which is severely affected by noise. The superiority of our estimator also holds when judged by the relative RMSE. According to the RMSE, the estimator performs best with a high bandwidth ($\beta = 0.8$). The LWLFC estimator shows higher bias (though still much smaller than that of the trimmed LP) but its variance is smaller. In three out of the four cases analyzed (the exception being $d = 0.2$ and $p = 20$) the reduction in variance is not big enough so that the LWPLFC estimator has overall a smaller RMSE when using a large bandwidth. As expected, the performance of the LWPLFC improves as d increases, for reasons explained in Section 5.

Table 3
Bias and RMSE for a long memory process *ARFIMA*(α, $d = 0.45$, 0) with RLS.

$T \setminus \beta$	$p = 0$			$p = 5$			$p = 10$			$p = 20$		
	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8
(a) Bias	$\alpha = 0$											
256	−0.148	−0.056	−0.065	−0.142	−0.082	−0.080	−0.144	−0.096	−0.062	−0.195	−0.090	−0.099
512	−0.073	−0.038	−0.036	−0.075	−0.047	−0.045	−0.081	−0.041	−0.050	−0.097	−0.050	−0.063
1024	−0.049	−0.025	−0.026	−0.041	−0.039	−0.033	−0.036	−0.027	−0.081	−0.036	−0.027	−0.081
2048	−0.027	−0.018	−0.017	−0.060	−0.040	−0.030	−0.023	−0.014	−0.062	−0.023	−0.014	−0.062
4096	−0.016	−0.009	−0.010	−0.049	−0.035	−0.022	0.004	−0.013	−0.047	0.004	−0.013	−0.047
	$\alpha = 0.3$											
256	−0.071	0.069	0.142	−0.091	0.034	0.162	−0.060	0.079	0.170	−0.053	0.101	0.200
512	−0.056	0.043	0.125	−0.061	0.048	0.143	−0.054	0.078	0.144	−0.046	0.074	0.074
1024	−0.034	0.027	0.106	−0.014	0.044	0.118	−0.034	0.050	0.126	−0.005	0.054	0.140
2048	−0.001	0.015	0.091	−0.008	0.024	0.094	−0.014	0.033	0.105	−0.014	0.038	0.112
4096	−0.007	0.015	0.070	−0.006	0.019	0.078	0.002	0.020	0.081	−0.005	0.027	0.088
	$\alpha = 0.6$											
256	0.107	0.289	0.407	0.069	0.302	0.409	0.129	0.291	0.417	0.135	0.304	0.414
512	0.067	0.223	0.376	0.067	0.234	0.380	0.079	0.246	0.388	0.095	0.266	0.389
1024	0.042	0.169	0.341	0.049	0.169	0.342	0.048	0.185	0.344	0.061	0.201	0.355
2048	0.032	0.129	0.300	0.028	0.130	0.284	0.028	0.141	0.305	0.040	0.154	0.310
4096	0.019	0.089	0.263	0.014	0.071	0.227	0.022	0.100	0.234	0.009	0.107	0.269
(b) RMSE	$\alpha = 0$											
256	0.324	0.160	0.129	0.405	0.236	0.172	0.456	0.340	0.151	0.561	0.391	0.258
512	0.238	0.105	0.069	0.288	0.169	0.090	0.327	0.167	0.107	0.404	0.253	0.143
1024	0.187	0.077	0.052	0.191	0.101	0.084	0.210	0.099	0.082	0.288	0.133	0.084
2048	0.147	0.064	0.039	0.130	0.078	0.064	0.142	0.055	0.059	0.158	0.089	0.047
4096	0.080	0.043	0.030	0.092	0.059	0.050	0.097	0.040	0.042	0.116	0.061	0.028
	$\alpha = 0.3$											
256	0.329	0.124	0.159	0.356	0.206	0.199	0.347	0.193	0.201	0.451	0.256	0.229
512	0.213	0.093	0.135	0.257	0.161	0.154	0.270	0.137	0.155	0.399	0.181	0.183
1024	0.133	0.066	0.111	0.144	0.090	0.124	0.215	0.094	0.133	0.227	0.114	0.146
2048	0.085	0.044	0.094	0.094	0.056	0.097	0.117	0.066	0.109	0.146	0.081	0.116
4096	0.060	0.033	0.072	0.063	0.042	0.081	0.082	0.048	0.083	0.097	0.055	0.091
	$\alpha = 0.6$											
256	0.244	0.307	0.413	0.324	0.320	0.414	0.250	0.325	0.422	0.289	0.385	0.421
512	0.195	0.234	0.380	0.196	0.243	0.383	0.215	0.257	0.391	0.239	0.278	0.392
1024	0.105	0.177	0.343	0.121	0.186	0.346	0.130	0.195	0.350	0.147	0.208	0.357
2048	0.073	0.135	0.301	0.088	0.135	0.284	0.091	0.148	0.310	0.101	0.160	0.311
4096	0.060	0.094	0.264	0.045	0.074	0.228	0.062	0.114	0.236	0.059	0.110	0.270

6.4. *The case with all three types of contaminations*

Table 5 presents results with all three types of contaminations. We consider strong short-memory dynamics ($\alpha = 0.6$) and a medium value for the average number of level shifts ($p = 10$). For the additive noise, we use $\sigma_w^2 = 1, 4$, and we set $d = 0.2, 0.45$. The results show that both the LWLFC and LWPLFC perform well. In general, the LWPLFC has better performance when a large bandwidth is used, while the LWLFC is better with a small bandwidth. For a large value of d_0 (0.45), the LWPLFC performs slightly better than the LWLFC under the optimal bandwidth applicable to each. When d_0 is small ($d_0 = 0.2$) the LWLFC has slightly better performance. This accords with Hurvich et al. (2005) who showed that the asymptotic variance of the LW estimator increases as d_0 decreases. Overall, the results show an advantage of using the LWPLFC with a large bandwidth. From unreported simulations, the performance of the LWLFC and LWPLFC deteriorates as α approaches 1 or with a moving-average parameter close to −1, with or without noise. This is a problem common to most, if not all, versions of LW or LP estimators, trimmed or not.

6.5. *Overall summary and recommendations*

The results showed that our estimators have good finite sample properties and offer improved methods of inference compared to what is available in the literature. As with all existing semiparametric estimators of this type, the results can be sensitive to the choice of the bandwidth. In our case, a large bandwidth (e.g., $\beta = 0.8$) is preferable in most cases. One exception is when there is a strongly positively correlated short-memory component, in which case a smaller bandwidth ($\beta = 0.6$) is desirable. As of yet, there is no fully developed method to choose the bandwidth. But some approaches are possible for the practitioner to assess what is the best bandwidth to use. One is to estimate a preliminary parametric LFC model with an AR component for the noise. Upon obtaining a large estimate of the AR coefficient a smaller bandwidth is dictated and vice versa if the coefficient is small. While somewhat ad hoc, it should provide a useful guide.

7. Conclusions

We proposed a local-Whittle estimator of the memory parameter of a long memory time series process which has good properties under an almost complete collection of contamination processes that have been discussed in the literature. The estimator has many advantages: no assumption of Gaussianity is required unlike the trimmed log-periodogram estimator; there is no trimming involved so that all information from the low frequency components are retained; when there is no LFC, its performance is comparable to that of the standard LW estimator so that no asymptotic efficiency loss is incurred, with the loss of efficiency in finite sample being small as revealed by the simulations; with a proper

Table 4
Bias and RMSE for a long memory process with additive noise ($\sigma_w^2 = 4$) and RLS; LWLFC and LWPLFC estimators.

	$p = 0$						$p = 20$					
	$d = 0.2$			$d = 0.45$			$d = 0.2$			$d = 0.45$		
$T \setminus \beta$	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8
(a) Bias of LWLFC												
256	−0.218	−0.180	−0.172	−0.349	−0.328	−0.342	−0.248	−0.177	−0.164	−0.403	−0.366	−0.368
512	−0.182	−0.165	−0.164	−0.284	−0.296	−0.324	−0.228	−0.164	−0.164	−0.332	−0.314	−0.344
1024	−0.161	−0.146	−0.152	−0.224	−0.254	−0.299	−0.138	−0.141	−0.151	−0.257	−0.282	−0.322
2048	−0.133	−0.138	−0.145	−0.183	−0.223	−0.279	−0.096	−0.123	−0.123	−0.163	−0.182	−0.232
4096	−0.124	−0.128	−0.137	−0.142	−0.197	−0.261	−0.102	−0.110	−0.121	−0.114	−0.152	−0.215
(b) RMSE of LWLFC												
256	0.278	0.213	0.186	0.436	0.355	0.355	0.462	0.285	0.212	0.601	0.445	0.398
512	0.225	0.178	0.172	0.335	0.320	0.332	0.420	0.229	0.182	0.490	0.352	0.355
1024	0.191	0.156	0.156	0.261	0.266	0.303	0.276	0.181	0.162	0.358	0.307	0.328
2048	0.150	0.144	0.147	0.208	0.230	0.281	0.190	0.132	0.125	0.256	0.205	0.238
4096	0.135	0.131	0.139	0.158	0.202	0.263	0.157	0.109	0.119	0.157	0.163	0.218
(c) Bias of LWPLFC												
256	−0.420	−0.296	−0.158	−0.310	−0.241	−0.228	0.236	0.248	0.195	0.028	−0.007	0.060
512	−0.450	−0.246	−0.171	−0.219	−0.144	−0.115	0.258	0.137	0.116	0.096	0.068	0.066
1024	−0.368	−0.184	−0.107	−0.094	−0.074	−0.033	0.279	0.155	0.104	0.134	0.082	−0.006
2048	−0.232	−0.083	−0.038	−0.074	−0.050	−0.044	0.190	0.163	0.131	0.092	0.051	−0.006
4096	−0.188	−0.048	−0.030	−0.036	−0.030	−0.026	0.136	0.136	0.051	0.081	0.009	0.009
(d) RMSE of LWPLFC												
256	0.645	0.551	0.370	0.524	0.433	0.361	0.559	0.582	0.522	0.456	0.454	0.392
512	0.675	0.485	0.387	0.436	0.284	0.298	0.587	0.538	0.505	0.358	0.381	0.327
1024	0.613	0.431	0.258	0.288	0.184	0.163	0.563	0.564	0.496	0.338	0.272	0.252
2048	0.490	0.309	0.131	0.195	0.144	0.123	0.544	0.470	0.457	0.234	0.208	0.151
4096	0.447	0.231	0.105	0.160	0.113	0.090	0.531	0.419	0.341	0.145	0.153	0.117

Table 5
Bias and RMSE for a long memory process *ARFIMA*(0.6, d, 0) with RLS ($p = 10$) and additive noise.

	LWLFC						LWPLFC					
	$\sigma_w^2 = 1$			$\sigma_w^2 = 4$			$\sigma_w^2 = 1$			$\sigma_w^2 = 4$		
$T \setminus \beta$	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8
(a) Bias: $d = 0.2$												
256	0.131	0.252	0.246	−0.015	0.060	0.038	0.271	0.293	0.310	0.045	0.117	0.183
512	0.106	0.212	0.251	0.013	0.080	0.054	0.231	0.239	0.260	0.060	0.136	0.140
1024	0.061	0.178	0.242	0.004	0.061	0.076	0.179	0.185	0.250	0.061	0.106	0.121
2048	0.045	0.128	0.226	0.008	0.050	0.070	0.135	0.139	0.227	0.103	0.076	0.092
4096	0.021	0.092	0.202	−0.011	0.029	0.069	0.113	0.107	0.204	0.127	0.044	0.071
8192	0.012	0.066	0.176	−0.009	0.013	0.062	0.112	0.076	0.182	0.101	0.031	0.046
(b) Bias: $d = 0.45$												
256	0.057	0.167	0.145	−0.027	−0.064	−0.139	0.131	0.218	0.230	−0.014	0.072	0.115
512	0.032	0.171	0.184	−0.013	0.025	−0.051	0.112	0.185	0.215	0.050	0.082	0.113
1024	0.030	0.139	0.193	−0.023	0.033	−0.004	0.094	0.142	0.195	0.046	0.064	0.096
2048	0.015	0.112	0.186	−0.016	0.030	0.021	0.062	0.113	0.183	0.030	0.050	0.068
4096	0.010	0.083	0.174	−0.012	0.028	0.034	0.059	0.085	0.173	0.044	0.040	0.052
8192	0.005	0.059	0.156	−0.009	0.015	0.039	0.050	0.063	0.161	0.030	0.035	0.030
(c) RMSE: $d = 0.2$												
256	0.290	0.287	0.280	0.305	0.171	0.112	0.380	0.325	0.331	0.396	0.321	0.293
512	0.202	0.231	0.258	0.236	0.119	0.087	0.345	0.263	0.268	0.322	0.197	0.190
1024	0.144	0.189	0.246	0.142	0.089	0.088	0.291	0.199	0.254	0.282	0.149	0.147
2048	0.111	0.137	0.228	0.102	0.075	0.077	0.229	0.153	0.229	0.225	0.110	0.103
4096	0.088	0.100	0.204	0.077	0.050	0.072	0.206	0.121	0.205	0.231	0.078	0.076
8192	0.067	0.072	0.177	0.063	0.033	0.067	0.205	0.086	0.183	0.212	0.061	0.056
(d) RMSE: $d = 0.45$												
256	0.319	0.251	0.219	0.273	0.258	0.216	0.303	0.245	0.255	0.362	0.272	0.212
512	0.232	0.188	0.198	0.204	0.122	0.097	0.186	0.198	0.222	0.200	0.145	0.158
1024	0.142	0.150	0.196	0.160	0.075	0.058	0.145	0.157	0.199	0.162	0.100	0.121
2048	0.081	0.120	0.190	0.097	0.059	0.040	0.111	0.121	0.185	0.120	0.072	0.089
4096	0.061	0.089	0.175	0.061	0.043	0.041	0.099	0.091	0.175	0.094	0.060	0.060
8192	0.050	0.064	0.157	0.047	0.029	0.042	0.076	0.066	0.162	0.067	0.049	0.040

choice of the bandwidth, the extended estimator has good finite sample properties with short-run dynamics and/or additive noise; it is semi-parametric so that there is no need for a full specification of the underlying short-memory structure, though it can also be extended to cover a fully specified parametric structure for the long-memory component such as an ARFIMA process.

It does, nevertheless, have some drawbacks. First, the performance of the estimator is sensitive to the choice of the bandwidth. An adaptive, data-dependent method to select the bandwidth is an important avenue for future research. Note, however, that all current semiparametric estimators exhibit sensitivity to the bandwidth choice. Also, when the estimator is extended to account for noise, as in Hurvich et al. (2005), the RMSE is proportional to $(1/d_0)$ so that when the true parameter d_0 is close to zero the reduction in bias is offset by an increase in variance and a possible increase in the overall RMSE.

Appendix

We first introduce three lemmas which show that to some extent the pseudo spectral density function controls the periodogram of the process well, in the sense that the ratio $|I_k/f_k|$ is bounded and the average of $(I_k/f_k - 1)$ is $o_p(1)$.

Lemma A.1. *Let $A_k = (2\pi T)^{-1/2}\sum_{t=1}^{T} z_t\cos(\lambda_k t)$, $B_k = (2\pi T)^{-1/2}\sum_{t=1}^{T} z_t\sin(\lambda_k t)$, so that $I_k = (A_k)^2 + (B_k)^2$, and define the vector $\gamma = ((f_k)^{-1/2}A_k, (f_k)^{-1/2}B_k, (f_j)^{-1/2}A_j, (f_j)^{-1/2}B_j)'$. Let $\kappa(X_1, X_2, X_3, X_4)$ denote the joint cumulant of the random variables X_1, X_2, X_3, X_4 with n_1, n_2, n_3, n_4 nonnegative integers that sum to n. Then under Assumptions* A1–A5, *for $\theta_0 > 0$ and letting $M_0 = \theta_0/(2\pi)^{2-2d_0}$, for any sequences of positive integers k and j such that $k > j$ and $k/T \to 0$, the following result holds for $n > 2$:*

$$\kappa(\gamma_1^{n_1}, \gamma_2^{n_2}, \gamma_3^{n_3}, \gamma_4^{n_4}) = O\left(\left(\frac{T^{n/2-nd}}{k^{(n_1+n_3)(1-d_0)}j^{(n_2+n_4)(1-d_0)}}\right)\Bigg/\left(1+M_0\frac{T^{1-2d_0}}{k^{2-2d_0}}\right)^{(n_1+n_3)} \times\left(1+M_0\frac{T^{1-2d_0}}{j^{2-2d_0}}\right)^{(n_2+n_4)}\right)^{1/2}$$

which is $O(1)$ if $j \le T^{(1-2d_0)/(2-2d_0)}$ and $o(1)$ if $j > T^{(1-2d_0)/(2-2d_0)}$. Similarly, for $n > 2$, the nth cumulant of $\tilde{\gamma} = (A_k/(f_k)^{1/2}, B_k/(f_k)^{1/2})'$ is $O((T^{n/2-nd_0}/k^{n(1-d_0)})/(1+M_0(T^{1-2d_0}/k^{2-2d_0})^{n/2}))$. When $\theta_0 = 0$, $M_0 = 0$ and the result reduces to

$$\kappa(\gamma_1^{n_1}, \gamma_2^{n_2}, \gamma_3^{n_3}, \gamma_4^{n_4}) = O(T^{n/(2-nd_0)}/[k^{(n_1+n_3)(1-d_0)}j^{(n_2+n_4)(1-d_0)}]).$$

Proof. This lemma is a direct consequence of Lemma A.3 in McCloskey and Perron (2013), henceforth MP, and the definition of the pseudo spectral density function f_k. The difference in the results is simply due to the fact that we use $f_k = \lambda_k^{-2d_0} + (\theta_0/T)\lambda_k^{-2}$, while MP use $f_k = \lambda_k^{-2d_0}$. Hence, a different expression is obtained when $\theta_0 > 0$.

Lemma A.2. *Under* A1–A5, *with $I_k = \omega_k\omega_k^*$ and $M_0 = \theta_0/(2\pi)^{2-2d_0}$, for $1 \le j < k \le m$:*

(i) $E(I_k/f_k) = 1 + [O(k^{-1}\log k) + O(k/T)^{1+2d_0}]/[1 + M_0(T^{1-2d_0}/k^{2-2d_0})]$

(ii) $E((\omega_k)^2/f_k) = O(k^{-1}\log k) + O(T^{1-2d_0}/k^{2-2d_0})/[1 + M_0(T^{1-2d_0}/k^{2-2d_0})]$

(iii) $$E\left(\frac{\omega_k\omega_j^*}{\sqrt{f_kf_j}}\right) = O(k^{-1}\log j) + \frac{O(T^{1-2d_0}/(k^{1-d_0}j^{1-d_0}))}{\sqrt{(1+M_0(T^{1-2d_0}/k^{2-2d_0}))(1+M_0(T^{1-2d_0}/j^{2-2d_0}))}}$$

(iv) $$E\left(\frac{\omega_k\omega_j}{\sqrt{f_kf_j}}\right) = O(k^{-1}\log j) + \frac{O(T^{1-2d_0}/(k^{1-d_0}j^{1-d_0}))}{\sqrt{(1+M_0(T^{1-2d_0}/k^{2-2d_0}))(1+M_0(T^{1-2d_0}/j^{2-2d_0}))}}.$$

Proof. For part (i), we have $E(I_{u,k}/(T^{-1}\lambda_k^{-2})) = O_p(1)$. Hence, from Theorem 1 in MP,

$$\begin{aligned}
E\left(\frac{I_k}{f_k}\right) &= E\left(\frac{I_k}{f_{y,k}}\frac{f_{y,k}}{f_k}\right) = \frac{f_{y,k}}{f_k}E\left(\frac{I_k}{f_{y,k}}\right)\\
&= \frac{f_{y,k}}{f_k}E\left(\frac{I_{y,k}}{f_{y,k}} + \frac{I_{u,k}}{f_{y,k}} + \frac{2I_{yu,k}}{f_{y,k}}\right)\\
&= \frac{\lambda_k^{-2d_0}}{\lambda_k^{-2d_0} + (\theta_0/T)\lambda_k^{-2}}\left(1 + O\left(\frac{\log k}{k} + \left(\frac{k}{T}\right)^2\right) + M_0\frac{T^{1-2d_0}}{k^{2-2d_0}} + O\left(\frac{k^3}{T^2}\frac{T^{1-2d_0}}{k^{2-2d_0}}\right)\right)\\
&= \frac{1}{1+M_0(T^{1-2d_0}/k^{2-2d_0})}\left(1 + M_0\frac{T^{1-2d_0}}{k^{2-2d_0}} + O\left(\frac{\log k}{k} + \left(\frac{k}{T}\right)^2\right) + O\left(\frac{k}{T}\right)^{1+2d_0}\right)\\
&= 1 + \frac{O(k^{-1}\log k) + O(k/T)^{1+2d_0}}{1+M_0(T^{1-2d_0}/k^{2-2d_0})}.
\end{aligned}$$

For part (ii),

$$\begin{aligned}
E\left(\frac{(\omega_k)^2}{f_k}\right) &= E\left(\frac{(\omega_k)^2}{f_{y,k}}\frac{f_{y,k}}{f_k}\right) = \frac{f_{y,k}}{f_k}E\left(\frac{(\omega_k)^2}{f_{y,k}}\right)\\
&= \frac{1}{1+M_0(T^{1-2d_0}/k^{2-2d_0})}O\left(\frac{\log k}{k} + \frac{T^{1-2d_0}}{k^{2-2d_0}}\right)\\
&= O\left(\frac{\log k}{k}\right) + \frac{O(T^{1-2d_0}/k^{2-2d_0})}{1+M_0(T^{1-2d_0}/k^{2-2d_0})}.
\end{aligned}$$

For part (iii),

$$\begin{aligned}
E&\left(\frac{\omega_k\omega_j^*}{\sqrt{f_kf_j}}\right)\\
&= E\left(\frac{\omega_k\omega_j^*}{\sqrt{f_{y,k}f_{y,j}}}\frac{\sqrt{f_{y,k}f_{y,j}}}{\sqrt{f_kf_j}}\right) = \frac{\sqrt{f_{y,k}f_{y,j}}}{\sqrt{f_kf_j}}E\left(\frac{\omega_k\omega_j^*}{\sqrt{f_{y,k}f_{y,j}}}\right)\\
&= (1+M_0(T^{1-2d_0}/k^{2-2d_0}))(1+M_0(T^{1-2d_0}/k^{2-2d_0}))^{-1/2} \times O\left(\frac{\log j}{k} + \frac{T^{1-2d_0}}{k^{1-d_0}j^{1-d_0}}\right)\\
&= O\left(\frac{\log j}{k}\right) + \frac{O\left(T^{1-2d_0}/k^{1-d_0}j^{1-d_0}\right)}{[(1+M_0(T^{1-2d_0}/k^{2-2d_0}))(1+M_0(T^{1-2d_0}/j^{2-2d_0}))]^{1/2}}
\end{aligned}$$

and the proof is entirely analogous for part (iv).

Lemma A.3. *Under* A1–A5*: if* (a) *$\theta = \theta_m$ is bounded away from zero or* (b) *there is no LFC in data, then:* (1) *$|I_k/f_k|$ is bounded, and* (2) *$m^{-1}\sum_{k=1}^{m}(I_k/f_k - 1) = o_p(1)$.*

Proof. First,

$$\frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_k}{f_k} - 1\right) = \frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_k}{f_k} - \frac{I_{y,k}}{f_{y,k}}\right) + \frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_{y,k}}{f_{y,k}} - 1\right).$$

For the first term, we have:

$$\frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}\right)=\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}\left(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}\right)+\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}\right)$$

whose first component is such that,

$$\begin{aligned}\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}\left(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}\right)&=\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}\left(\frac{I_{y,k}}{f_{z,k}}-\frac{I_{y,k}}{f_{y,k}}+\frac{I_{u,k}}{f_{z,k}}+2\frac{I_{yu,k}}{f_{z,k}}\right)\\&=\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}\left(\frac{I_{y,k}}{f_{y,k}}\left(-\frac{f_{u,k}}{f_{z,k}}\right)+\frac{I_{u,k}}{f_{z,k}}+2\frac{I_{yu,k}}{f_{z,k}}\right)\\&=\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}\left(\frac{I_{u,k}-f_{u,k}}{f_{z,k}}-\left(\frac{I_{y,k}}{f_{y,k}}-1\right)\left(\frac{f_{u,k}}{f_{z,k}}\right)+2\frac{I_{yu,k}}{f_{z,k}}\right).\end{aligned}$$

Note that

$$E\left|\frac{I_{u,k}-f_{u,k}}{f_{z,k}}\right|=E\left|\left(\frac{I_{u,k}}{f_{u,k}}-1\right)\Big/\left(\frac{f_{z,k}}{f_{u,k}}\right)\right|=\frac{f_{u,k}}{f_{z,k}}E\left|\frac{I_{u,k}}{f_{u,k}}-1\right|.$$

From MP (Lemma A.3) with $n_1=n_2=n_3=n_4=1$ and $\gamma_1=\gamma_2=\gamma_3=\gamma_4=I_{u,k}/f_{y,k}$:

$$E\left|\frac{I_{u,k}}{f_{u,k}}-1\right|\le E\left|\frac{I_{u,k}}{f_{u,k}}\right|+1\le\left[E\left(\left|\frac{I_{u,k}}{f_{u,k}}\right|^2\right)\right]^{1/2}+1\le C_1.$$

So

$$\begin{aligned}E\left|\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}\left(\frac{I_{u,k}-f_{u,k}}{f_{z,k}}\right)\right|&\le\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}\left|\frac{f_{u,k}}{f_{z,k}}\right|E\left|\frac{I_{u,k}}{f_{u,k}}-1\right|\\&\le\frac{\sqrt{T}}{m}C_1\to 0\end{aligned}$$

if $\sqrt{T}/m\to 0$. We also have $E|m^{-1}\sum_{k=1}^{\sqrt{T}-1}(I_{y,k}/f_{y,k}-1)(f_{u,k}/f_{z,k})|\to 0$, since $|f_{u,k}/f_{z,k}|<1$. From MP, Perron and Qu (2010) and Qu (2011): $I_{yu}(\lambda_k)=O_p(T^{-1/2}\lambda_k^{-(1+d_0)})$ and $f_k=f_{z,k}=f_{y,k}+f_{u,k}=G\lambda_k^{-2d_0}+G_uT^{-1}\lambda_k^{-2}$. Hence,

$$\begin{aligned}\left|\frac{I_{yu,k}}{f_{z,k}}\right|&\sim\frac{O_p(T^{-1/2}\lambda_k^{-(1+d_0)})}{O_p(\lambda_k^{-2d_0})+O_p(T^{-1}\lambda_k^{-2})}\\&\sim\frac{1}{O_p(T^{1/2}\lambda_k^{1-d_0})+O_p(T^{-1/2}\lambda_k^{d_0-1})}<O_p(1)\end{aligned}$$

and

$$\begin{aligned}E\left|\frac{2}{m}\sum_{k=1}^{\sqrt{T}-1}\frac{I_{yu,k}}{f_{z,k}}\right|&\le\frac{2}{m}\sum_{k=1}^{\sqrt{T}-1}E\left|\frac{I_{yu,k}}{f_{z,k}}\right|<\frac{2}{m}\sqrt{T}O_p(1)\\&=O_p\left(\frac{\sqrt{T}}{m}\right)\to 0\end{aligned}$$

if $\sqrt{T}/m\to 0$. Hence,

$$\begin{aligned}&E\left|\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}\left(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}\right)\right|\\&=E\left|\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}\left(\frac{I_{u,k}-f_{u,k}}{f_{z,k}}-\left(\frac{I_{y,k}}{f_{y,k}}-1\right)\left(\frac{f_{u,k}}{f_{z,k}}\right)+2\frac{I_{yu,k}}{f_{z,k}}\right)\right|\\&\le E\left|\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}\left(\frac{I_{u,k}-f_{u,k}}{f_{z,k}}\right)\right|+E\left|\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}\left(\frac{I_{y,k}}{f_{y,k}}-1\right)\left(\frac{f_{u,k}}{f_{z,k}}\right)\right|\\&\quad+E\left|\frac{2}{m}\sum_{k=1}^{\sqrt{T}-1}\frac{I_{yu,k}}{f_{z,k}}\right|\to 0\end{aligned}$$

if $\sqrt{T}/m\to 0$. It is easy to show that $E|m^{-1}\sum_{k=\sqrt{T}}^{m}(I_k/f_k-I_{y,k}/f_{y,k})|\to 0$, and the fact that $E|m^{-1}\sum_{k=1}^{m}(I_{y,k}/f_{y,k}-1)|\to 0$ follows from Hurvich et al. (2005). So

$$\begin{aligned}E\left|\frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_k}{f_k}-1\right)\right|&\le E\left|\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}\left(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}\right)\right|\\&\quad+E\left|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}\right)\right|\\&\quad+E\left|\frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_{y,k}}{f_{y,k}}-1\right)\right|\to 0\end{aligned}$$

if $\sqrt{T}/m\to 0$. Note that during the proof we also showed that $|I_k/f_k|\le|I_k/f_k-I_{y,k}/f_{y,k}|+|I_{y,k}/f_{y,k}-1|+1$ is bounded.

Proof of Lemma 1. Let $M_m=\hat{\theta}_m/(2\pi)^{2-2\hat{d}_m}$ and $M_0=\theta_0/(2\pi)^{2-2d_0}$. We analyze the partial derivative of the objective function with respect to θ:

$$\begin{aligned}\frac{\partial}{\partial\theta}J_m(\hat{d}_m,\hat{\theta}_m)&=\frac{1}{mT}\left[\sum_{k=1}^{m}\frac{1}{g_k(\hat{d}_m,\hat{\theta}_m)}\lambda_k^{-2}\right.\\&\quad\left.-\left(\frac{1}{m}\sum_{k=1}^{m}\frac{I_k}{g_k(\hat{d}_m,\hat{\theta}_m)}\right)^{-1}\sum_{k=1}^{m}\frac{I_k}{(g_k(\hat{d}_m,\hat{\theta}_m))^2}\lambda_k^{-2}\right]\\&=\frac{1}{mT}\left[\sum_{k=1}^{m}\left(1-\frac{I_k}{G_0g_k(\hat{d}_m,\hat{\theta}_m)}\frac{G_0}{m^{-1}\sum_{j=1}^{m}(I_j/g_j(\hat{d}_m,\hat{\theta}_m))}\right)\right.\\&\quad\left.\times\frac{\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}\right]\\&=\frac{1}{mT}\left[\sum_{k=1}^{m}\left(1-\frac{I_k}{f_k}\frac{G_0}{m^{-1}\sum_{j=1}^{m}(I_j/g_j(\hat{d}_m,\hat{\theta}_m))}\frac{g_k(d_0,\theta_0)}{g_k(\hat{d}_m,\hat{\theta}_m)}\right)\right.\\&\quad\left.\times\frac{\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}\right]\\&=\frac{1}{mT}\left\{\sum_{k=1}^{m}\left(\frac{\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}\right)\right.\end{aligned}$$

$$\times\left[1-\left(m\frac{I_k}{f_k}\left(\frac{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}\right)\right.\right.$$
$$\left.\left.\left.\Big\backslash\sum_{j=1}^{m}\frac{I_j}{f_j}\left(\frac{\lambda_j^{-2d_0}+(\theta_0/T)\lambda_j^{-2}}{\lambda_j^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_j^{-2}}\right)\right)\right]\right\}. \tag{A.1}$$

Using summation by parts, (A.1) becomes:

$$\left\{\sum_{k=1}^{m}\left(\frac{\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}\right)\left[1-\left(m\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}\right.\right.\right.$$
$$\left.\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2d}/k^{2-2d})}\right)\right)$$
$$\left.\left.\Big\backslash\left(\sum_{j=1}^{m}\frac{I_j}{f_j}\lambda_j^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\right)\right]\right\}$$
$$=\left(\frac{\lambda_m^{-2}}{\lambda_m^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_m^{-2}}\right)\sum_{k=1}^{m}\left[1-\left(m\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}\right.\right.$$
$$\left.\times\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\right)$$
$$\left.\Big\backslash\left(\sum_{j=1}^{m}\frac{I_j}{f_j}\lambda_j^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\right)\right]$$
$$+\sum_{j=1}^{m-1}\left[\left(\frac{\lambda_j^{-2}}{\lambda_j^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_j^{-2}}\right)-\left(\frac{\lambda_{j+1}^{-2}}{\lambda_{j+1}^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_{j+1}^{-2}}\right)\right]$$
$$\left\{\sum_{k=1}^{j}\left(1-\left(m\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\right)\right.\right.$$
$$\left.\left.\Big\backslash\left(\sum_{k=1}^{m}\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\right)\right]\right\}$$
$$=\left(\frac{\lambda_m^{-2}}{\lambda_m^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_m^{-2}}\right)\left[m\right.$$
$$-m\sum_{k=1}^{m}\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)$$
$$\left.\Big\backslash\left(\sum_{j=1}^{m}\frac{I_j}{f_j}\lambda_j^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\right)\right]$$
$$+\sum_{j=1}^{m-1}\left[\left(\frac{\lambda_j^{-2}}{\lambda_j^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_j^{-2}}\right)-\left(\frac{\lambda_{j+1}^{-2}}{\lambda_{j+1}^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_{j+1}^{-2}}\right)\right]$$
$$\times\left[j-m\sum_{k=1}^{j}\left(\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\right)\right.$$
$$\left.\Big\backslash\left(\sum_{k=1}^{m}\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\right)\right].$$

Now, suppose $\hat{\theta}_m \to 0$ with $(\partial/\partial\theta)J_m(\hat{d}_m, \hat{\theta}_m) = 0$. We define $h_k \equiv [1 + M_0(T^{1-2d_0}/k^{2-2d_0})]/[1 + M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})]$. We consider two cases. In the first, suppose $M_m \to 0$ at a slow rate such that for some small k, we still have $M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m}) \to \infty$. Let $\tau_m = \inf_k\{M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m}) \to 0\}$, then

$$h_k \sim \begin{cases}\dfrac{M_0}{M_m}\left(\dfrac{T}{k}\right)^{2(\hat{d}_m-d_0)}+\dfrac{1}{M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})} & \text{when } k \le \tau_m\\ 1+M_0(T^{1-2d_0}/k^{2-2d_0}) \quad \text{when } k > \tau_m.\end{cases}$$

Note that we must have either $(M_0/M_m)(T/k)^{2(\hat{d}_m-d_0)}$ or $M_0(T^{1-2d_0}/k^{2-2d_0})$ go to infinity for some small k. Also,

$$\begin{aligned}M_0(T^{1-2d_0}/k^{2-2d_0}) &= \frac{M_0}{M_m}\left(\frac{T}{k}\right)^{2(\hat{d}_m-d_0)}\\ &\quad\times\left(M_m\left(\frac{T}{k}\right)^{-2(\hat{d}_m-d_0)}\frac{T^{1-2d_0}}{k^{2-2d_0}}\right)\\ &= \frac{M_0}{M_m}\left(\frac{T}{k}\right)^{2(\hat{d}_m-d_0)}\left(M_m\frac{T^{1-2\hat{d}_m}}{k^{2-2\hat{d}_m}}\right)\\ &= o_p\left(\frac{M_0}{M_m}\left(\frac{T}{k}\right)^{2(\hat{d}_m-d_0)}\right)\end{aligned}$$

when $k > \tau_m$. Hence,

$$\lambda_k^{2\hat{d}-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\sim\begin{cases}(M_0/M_m) & \text{when } k\le\tau_m\\ o_p(M_0/M_m) & \text{when } k>\tau_m.\end{cases}$$

Let

$$a_j=\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)$$

then we know that $a_j = O_p(M_0/M_m)$ when $k \le \tau_m$ and $a_j = o_p(M_0/M_m)$ when $k > \tau_m$. So, $\{a_j\}$ is a positive sequence whose first few terms have higher order than the rest. So we have

$$\begin{aligned}&(j/m)-\sum_{k=1}^{j}\left(\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\right)\\ &\quad\Big\backslash\left(\sum_{k=1}^{m}\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\right)\\ &\quad\le C_j<0\end{aligned}\tag{A.2}$$

where C_j is some constant. Under the second case, $M_m \to 0$ fast enough so that, for any k, $M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m}) \le O_p(1)$. For this case, $h_k \sim 1 + M_0(T^{1-2d_0}/k^{2-2d_0})$ and

$$\begin{aligned}&\lambda_k^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\\ &\quad\sim\lambda_k^{2\hat{d}_m-2d_0}(1+M_0(T^{1-2d_0}/k^{2-2d_0}))\\ &\quad\sim(T/k)^{2d_0-2\hat{d}_m}+M_0(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})\end{aligned}$$

If $d_0 \ge \hat{d}_m$, the last expression is decreasing in k for all $k = 1, \ldots, m$; if $d_0 < \hat{d}_m$, the first is increasing in k, but always smaller than 1, and the second is decreasing in k and goes to infinity when k is small. Hence, (A.2) still holds. Since for T large enough,

$$\begin{aligned}&\lambda_j^{-2}(\lambda_j^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_j^{-2})^{-1}-\lambda_{j+1}^{-2}(\lambda_{j+1}^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_{j+1}^{-2})^{-1}\\ &\quad\ge D_j>0\end{aligned}$$

where D_j is some constant, we have shown that $(\partial/\partial\theta)J_m(\hat{d}_m, \hat{\theta}_m) < 0$ if $\hat{\theta}_m \to 0$, which is a contradiction. So $\hat{\theta}_m$ has to be bounded from zero when $\theta_0 > 0$.

Proof of Theorem 1. First, we consider the case when LFC indeed exists in the true DGP. The proof for the case with no LFC will follow with trivial modifications. Note that if LFC components are present, the probability limit of the estimate $\hat{\theta}_m$ is bounded above zero, by Lemma 1. This implies that with probability arbitrarily close to one, $\hat{\theta}_m$ will be in the set $(0, \infty)$ and, without loss of generality, we can consider analyzing the limit of $\hat{d}_m$ for any sequence or values of θ_m in the set $(0, \infty)$. Accordingly, we want to show that, with probability arbitrarily close to one for large T and m, if $\{\theta_m\}$ is a sequence bounded above from zero and if $\{\hat{d}_m\}$ minimizes $J_m(d, \theta_m)$ given $\{\theta_m\}$, then for $\hat{d}_m$ such that $|\hat{d}_m - d_0| \geq \delta$ for any $\delta > 0$, we have $J_m(\hat{d}_m, \theta_m) - J_m(d_0, \theta_m) > 0$, which delivers a contradiction showing that in the limit the minimizer of $J_m(d, \theta_m)$ must converge to d_0. Let $G(d, \theta_m) = m^{-1} \sum_{k=1}^m I_k/g_k$, where $g_k = (\lambda_k^{-2d} + (\theta_m/T)\lambda_k^{-2})$. We first have:

$$\begin{aligned}
&J_m(\hat{d}_m, \theta_m) - J_m(d_0, \theta_m) \\
&= \left[\log G(\hat{d}_m, \theta_m) + \frac{1}{m}\sum_{k=1}^m \log\left(\lambda_k^{-2\hat{d}_m}\left(1 + \frac{\theta_m}{T}\lambda_k^{-2+2\hat{d}_m}\right)\right)\right] \\
&\quad - \left[\log G(d_0, \theta_m) + \frac{1}{m}\sum_{k=1}^m \log\left(\lambda_k^{-2\hat{d}_m}\right.\right. \\
&\quad \times \left.\left.\left(1 + \frac{\theta_m}{T}\lambda_k^{-2+2d_0}\right)\right)\right] \\
&= \log G(\hat{d}_m, \theta_m) - \log G(d_0, \theta_m) \\
&\quad + \frac{1}{m}\sum_{k=1}^m \log\left(\lambda_k^{-2(\hat{d}_m - d_0)}\left(\frac{1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m}}{1 + (\theta_m/T)\lambda_k^{-2+2d_0}}\right)\right) \\
&= \log \frac{G(\hat{d}_m, \theta_m)}{G_0\left(m^{-1}\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)}\right)} - \log \frac{G(d_0, \theta_m)}{G_0} \\
&\quad + \log\left(\frac{1}{m}\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)}\right) \\
&\quad - \frac{2(\hat{d}_m - d_0)}{m}\sum_{k=1}^m \lambda_k + \frac{1}{m}\sum_{k=1}^m \log\left(\frac{1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m}}{1 + (\theta_m/T)\lambda_k^{-2+2d_0}}\right) \\
&= \log \frac{G(\hat{d}_m, \theta_m)}{G_0\left(m^{-1}\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)}\right)} - \log \frac{G(d_0, \theta_m)}{G_0} \\
&\quad + \log(\lambda_k^{2(\hat{d}_m - d_0)}(2(\hat{d}_m - d_0) + 1)) - \log(2(\hat{d}_m - d_0) + 1) \\
&\quad - 2(\hat{d}_m - d_0)\left[\frac{1}{m}\sum_{k=1}^m (\log k - \log m)\right] \\
&\quad + \frac{1}{m}\sum_{k=1}^m \log\left(\frac{1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m}}{1 + (\theta_m/T)\lambda_k^{-2+2d_0}}\right) \\
&= \log \frac{G(\hat{d}_m, \theta_m)}{G_0\left(m^{-1}\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)}\right)} - \log \frac{G(d_0, \theta_m)}{G_0} \\
&\quad + \log\left(\frac{2(\hat{d}_m - d_0) + 1}{m}\sum_{k=1}^m \left(\frac{k}{m}\right)^{2(\hat{d}_m - d_0)}\right) \\
&\quad - 2(\hat{d}_m - d_0)\left[\frac{1}{m}\sum_{k=1}^m \log k - (\log m - 1)\right] \\
&\quad + \frac{1}{m}\sum_{k=1}^m \log\left(\frac{1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m}}{1 + (\theta_m/T)\lambda_k^{-2+2d_0}}\right) \\
&\quad - \log(1 + 2(\hat{d}_m - d_0)) + 2(\hat{d}_m - d_0). \qquad \text{(A.3)}
\end{aligned}$$

Note that for the last term of (A.3), we have $-\log(1+2(\hat{d}_m-d_0)) + 2(\hat{d}_m - d_0) \geq (1/6)(\hat{d}_m - d_0)^2 \geq (1/6)\delta^2$. Hence, if we can show that the other five terms are $o_p(1)$, we can derive a contradiction. The third and fourth are $o_p(1)$ from Robinson (1995a).

Proof that the first term of (A.3) is $o_p(1)$. To show that, it is equivalent to prove that $G(\hat{d}_m, \theta_m)/[G_0(m^{-1}\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)})] - 1 = o_p(1)$. To that effect,

$$\begin{aligned}
&\frac{G(\hat{d}_m, \theta_m)}{G_0\left(m^{-1}\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)}\right)} - 1 \\
&= \left(G_0\left(m^{-1}\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)}\right)\right)^{-1} \\
&\quad \times \left\{\frac{1}{m}\sum_{k=1}^m \frac{I_k}{\lambda_k^{-2\hat{d}_m}(1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m})}\right. \\
&\quad \left. - \frac{1}{m}G_0\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)}\right\} \\
&= \left(\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)}\right)^{-1}\left\{\sum_{k=1}^m\left[\frac{I_k}{G_0\lambda_k^{-2d_0}(1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m})}\right.\right. \\
&\quad \left.\left. \times \lambda_k^{2(\hat{d}_m - d_0)} - \lambda_k^{2(\hat{d}_m - d_0)}\right]\right\} \\
&= \left(\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)}\right)^{-1}\left\{\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)}\right. \\
&\quad \left. \times \left[\frac{I_k}{f_k}\frac{1 + (\theta_0/T)\lambda_k^{-2+2d_0}}{1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m}} - 1\right]\right\} \\
&= \left(\frac{1}{m}\sum_{k=1}^m \left(\frac{k}{m}\right)^{2(\hat{d}_m - d_0)}\right)^{-1}\left\{\frac{1}{m}\sum_{k=1}^m \left(\frac{k}{m}\right)^{2(\hat{d}_m - d_0)}\right. \\
&\quad \left. \times \left(\frac{I_k}{f_k}\frac{1 + (\theta_0/T)\lambda_k^{-2+2d_0}}{1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m}} - 1\right)\right\}.
\end{aligned}$$

From Robinson (1995a), $m^{-1}\sum_{k=1}^m (k/m)^{2(\hat{d}_m - d_0)} = o_p(1)$ if $\hat{d}_m - d_0 \neq -1/2$. When $k \geq \sqrt{T}$,

$$\frac{I_k}{f_k}\frac{1 + (\theta_m/T)\lambda_k^{-2+2d_0}}{1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m}} - 1 = \left(\frac{I_k}{f_k} - 1\right) + \frac{I_k}{f_k}O\left(\left(\frac{k}{T}\right)^{2d_0}\right).$$

From Lemma A.2, $m^{-1}\sum_{k=1}^m (I_k/f_k - 1) = o_p(1)$, hence $m^{-1}\sum_{k=1}^m I_k/f_k = 1 + o_p(1)$. Also,

$$\frac{1}{m}\sum_{k=1}^{\sqrt{T}} \frac{I_k}{f_k} = \frac{1}{m}\sum_{k=1}^{\sqrt{T}}\left(\frac{I_k}{f_k} - \frac{I_{y,k}}{f_{y,k}}\right) + \frac{1}{m}\sum_{k=1}^{\sqrt{T}}\frac{I_{y,k}}{f_{y,k}}.$$

The first term is $o_p(1)$ from Lemma A.3 and the second is $o_p(m^{-1}\sqrt{T}) = o_p(1)$ from Hurvich et al. (2005). Combining these results, we have $m^{-1}\sum_{k=\sqrt{T}}^m I_k/f_k = 1 + o_p(1)$. Now,

$$\begin{aligned}
&\frac{1}{m}\sum_{k=\sqrt{T}}^m \left(\frac{k}{m}\right)^{2(\hat{d}_m - d_0)}\left(\frac{I_k}{f_k}\frac{1 + (\theta_0/T)\lambda_k^{-2+2d_0}}{1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m}} - 1\right) \\
&= \frac{1}{m}\sum_{k=\sqrt{T}}^m \left(\frac{k}{m}\right)^{2(\hat{d}_m - d_0)}\left(\frac{I_k}{f_k} - 1\right) \\
&\quad + \frac{1}{m}\sum_{k=\sqrt{T}}^m \left(\frac{k}{m}\right)^{2(\hat{d}_m - d_0)} O\left(\frac{T}{k^2}\left(\frac{k}{T}\right)^{2d_0}\right)\frac{I_k}{f_k}
\end{aligned}$$

$$= \frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\left(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}\right) + \frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\left(\frac{I_{y,k}}{f_{y,k}}-1\right) + \frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{k}{m}\right)^{2\hat{d}_m} O\left(\frac{T}{k^2}\left(\frac{m}{T}\right)^{2d_0}\right)\left(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}\right) + \frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{k}{m}\right)^{2\hat{d}_m} O\left(\frac{T}{k^2}\left(\frac{m}{T}\right)^{2d_0}\right)\frac{I_{y,k}}{f_{y,k}}. \quad (A.4)$$

We will show that all four terms of (A.4) are $o_p(1)$ by showing that the expectations of their absolute values are $o_p(1)$. For the first term, we have from Lemma A.3,

$$E\left(\left|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\left(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}\right)\right|\right) \le \frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)} E\left|\left(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}\right)\right|.$$

From Lemma A.1, we know that $E|(I_k/f_k - I_{y,k}/f_{y,k})| \sim C(k/T)^{d_0} \le C(m/T)^{d_0}$, where C is some constant not depending on T and m. We also have

$$\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)} = \frac{\sqrt{T}}{m}\frac{1}{\sqrt{T}}\sum_{k=1}^{\sqrt{T}}\left(\frac{k}{\sqrt{T}}\right)^{2(\hat{d}_m-d_0)}\left(\frac{\sqrt{T}}{m}\right)^{2(\hat{d}_m-d_0)} = \left(\frac{\sqrt{T}}{m}\right)^{1+2(\hat{d}_m-d_0)}\frac{1}{\sqrt{T}}\sum_{k=1}^{\sqrt{T}}\left(\frac{k}{\sqrt{T}}\right)^{2(\hat{d}_m-d_0)} = O_p\left(\frac{\sqrt{T}}{m}\right) \to 0$$

where the last equality is from Robinson (1995a, Eq. (3.7)). Hence, $m^{-1}\sum_{k=\sqrt{T}}^{m}(k/m)^{2(\hat{d}_m-d_0)} \sim 1/(1+2(\hat{d}_m-d_0))$, which shows that the first term is $o_p(1)$. For the second term,

$$E\left|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\left(\frac{I_{y,k}}{f_{y,k}}-1\right)\right| = E\left|\frac{1}{m}\sum_{k=1}^{m}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\left(\frac{I_{y,k}}{f_{y,k}}-1\right) - \frac{1}{m}\sum_{k=1}^{\sqrt{T}}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\left(\frac{I_{y,k}}{f_{y,k}}-1\right)\right|$$
$$\le E\left|\frac{1}{m}\sum_{k=1}^{m}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\left(\frac{I_{y,k}}{f_{y,k}}-1\right)\right| + \left(\frac{\sqrt{T}}{m}\right)^{1+2(\hat{d}_m-d_0)} \times E\left|\frac{1}{\sqrt{T}}\sum_{k=1}^{\sqrt{T}}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\left(\frac{I_{y,k}}{f_{y,k}}-1\right)\right| = o_p(1).$$

For the third term,

$$E\left(\left|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{k}{m}\right)^{2\hat{d}_m} O\left(\frac{T}{k^2}\left(\frac{m}{T}\right)^{2d_0}\right)\left(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}\right)\right|\right) = E\left|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{k}{m}\right)^{2\hat{d}_m} O\left(\frac{m}{T}\right)^{2d_0}\left(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}\right)\right|$$
$$\le E\left(\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\frac{T}{k^2}O\left(\frac{m}{T}\right)^{2d_0} C\left(\frac{k}{m}\right)^{2d_0}\right) \le O_p\left(\frac{T}{m^2}\right) = o_p(1).$$

For the fourth term,

$$E\left(\left|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{k}{m}\right)^{2\hat{d}_m} O\left(\frac{T}{k^2}\left(\frac{m}{T}\right)^{2d_0}\right)\frac{I_{y,k}}{f_{y,k}}\right|\right) = E\left|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{k}{m}\right)^{2\hat{d}_m}\frac{T}{k^2}O\left(\frac{m}{T}\right)^{2d_0}\frac{I_{y,k}}{f_{y,k}}\right|$$
$$\le \frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{k}{m}\right)^{2\hat{d}_m}\frac{T}{k^2}O\left(\frac{m}{T}\right)^{2d_0} E\left|\frac{I_{y,k}}{f_{y,k}}\right| \le O_p\left(\frac{T}{m^2}\right) = o_p(1)$$

according to Eq. (3.15) in Robinson (1995a). Hence,

$$\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\left(\frac{I_k}{f_k}\frac{1+(\theta_0/T)\lambda_k^{-2+2d_0}}{1+(\theta_m/T)\lambda_k^{-2+2d}} - 1\right) = o_p(1).$$

Proof that the second term of (A.3) is $o_p(1)$. Note that

$$(\theta_m/T)\lambda_k^{-2+2\hat{d}_m} = \theta_m(2\pi)^{-2+2\hat{d}_m}(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})$$

and let $M_0 = \theta_0(2\pi)^{-2+2d_0}$, $M_m = \theta_m(2\pi)^{-2+2\hat{d}_m}$ and $\tilde{M} = \inf_{m\ge1}\{\theta_m\}(2\pi)^{-2+2\hat{d}_m}$. Then

$$\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\left(\frac{I_k}{f_k}\frac{1+(\theta_0/T)\lambda_k^{-2+2d_0}}{1+(\theta_m/T)\lambda_k^{-2+2\hat{d}_m}} - 1\right) = \frac{1}{m}\sum_{k=1}^{\sqrt{T}}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\left(\frac{I_k}{f_k}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right) + o_p(1).$$

Suppose first that $\hat{d}_m \in [0, d_0-\delta)$, then $(1-2d_0)/(2-2d_0) < (1-2\hat{d}_m)/(2-2\hat{d}_m)$. When $M_0 > 0$ and $M > 0$, we have

$$\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})} \begin{cases} \in [(1+\tilde{M})^{-1}, 1+M_0], & \text{if } k \ge T^{\frac{1-2\hat{d}_m}{2-2\hat{d}_m}} \\ = O_p(k^{2-2\hat{d}_m}/T^{1-2\hat{d}_m}) = o_p(1), & \text{if } k \in (T^{\frac{1-2d_0}{2-2d_0}}, T^{\frac{1-2\hat{d}_m}{2-2\hat{d}_m}}) \\ \le 2(M_0/\tilde{M})(k/T)^{2(d_0-\hat{d}_m)}, & \text{if } k \le T^{\frac{1-2d_0}{2-2d_0}} \end{cases}.$$

Hence,

$$E\left|\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\left(\frac{I_k}{f_k}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\right|$$
$$\le \frac{1}{m}\sum_{k=1}^{\sqrt{T}}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}E\left|\frac{I_k}{f_k}\right|\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)$$
$$\le \frac{C}{m}\sum_{k=1}^{T^{(1-2d_0)/(2-2d_0)}}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}2\frac{M_0}{\tilde{M}}\left(\frac{k}{T}\right)^{2(d_0-\hat{d}_m)}$$
$$+\frac{1+M_0}{m}C\sum_{k=T^{(1-2d_0)/(2-2d_0)}}^{\sqrt{T}}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}=o_p(1).$$

Second, suppose $\hat{d}_m \in (d_0+\delta, 1/2)$, then $(1-2\hat{d}_m)/(2-2\hat{d}_m) < (1-2d_0)/(2-2d_0)$, and

$$\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}$$
$$\left\{\begin{array}{l}\in[(1+\tilde{M})^{-1},1+M_0],\quad \text{if } k\ge T^{\frac{1-2d_0}{2-2d_0}}\\ =O_p(k^{2-2\hat{d}_m}/T^{1-2\hat{d}_m})=o_p(1),\quad \text{if } k\in(T^{\frac{1-2\hat{d}_m}{2-2\hat{d}_m}},T^{\frac{1-2d_0}{2-2d_0}})\\ \le 2(M_0/\tilde{M})(k/T)^{2(d_0-\hat{d}_m)},\quad \text{if } k\le T^{\frac{1-2\hat{d}_m}{2-2\hat{d}_m}}.\end{array}\right|$$

Hence,

$$E\left|\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\left(\frac{I_k}{f_k}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\right|$$
$$\le \frac{1}{m}\sum_{k=1}^{\sqrt{T}}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}E\left|\frac{I_k}{f_k}\right|\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)$$
$$\le 2\left(\frac{C}{m}\right)\sum_{k=1}^{T^{(1-2\hat{d}_m)/(2-2\hat{d}_m)}}\frac{M_0}{\tilde{M}}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\left(\frac{k}{T}\right)^{2(d_0-\hat{d}_m)} \tag{A.5}$$
$$+2\left(\frac{C}{m}M_0\right)\sum_{k=T^{(1-2\hat{d}_m)/(2-2\hat{d}_m)}}^{T^{(1-2d_0)/(2-2d_0)}}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}(T^{1-2d_0}/k^{2-2d_0}) \tag{A.6}$$
$$+\left((1+M_0)\frac{C}{m}\right)\sum_{k=T^{(1-2d_0)/(2-2d_0)}}^{\sqrt{T}}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}. \tag{A.7}$$

Note that (A.5) is of order $T^{(1-2\hat{d}_m)/(2-2\hat{d}_m)+2(d-d_0)}/m^{1+2(\hat{d}_m-d_0)} = o_p(1)$ and (A.6) is of order $T^{1-2d_0-(1-2\hat{d}_m)^2/(2-2\hat{d}_m)}/m^{1+2(\hat{d}_m-d_0)} = o_p(1)$ if $m/T^{[1-2d_0-(1-2\hat{d}_m)^2/(2-2\hat{d}_m)]/[1+2(\hat{d}_m-d_0)]}\to\infty$. Let

$$\beta_1(\hat{d}_m,d_0)$$
$$=[(1-2\hat{d}_m)/(2-2\hat{d}_m)+2(\hat{d}_m-d_0)]/[1+2(\hat{d}_m-d_0)],$$
$$\beta_2(\hat{d}_m,d_0)$$
$$=[1-2d_0-(1-2\hat{d}_m)^2/(2-2\hat{d}_m)]/[1+2(\hat{d}_m-d_0)].$$

Note that $(1-2\hat{d}_m)/(2-2\hat{d}_m)+2(\hat{d}_m-d_0)=(1-2d_0)-(1-2\hat{d}_m)^2/(2-2\hat{d}_m)$, so that $\beta_1(\hat{d}_m,d_0)=\beta_2(\hat{d}_m,d_0)\triangleq\beta(\hat{d}_m,d_0)=1-(2(1-\hat{d}_m)(1-2d_0+2\hat{d}_m))^{-1}$. Tedious algebra shows that if $0\le d_0<\hat{d}_m<1/2$ (which holds since we are considering the case $\hat{d}_m\in(d_0+\delta,1/2)$), then for a given d_0, the maximized value of $\beta(\hat{d}_m,d_0)$ is $1-(d_0^2-3d_0+9/4)^{-1}$. So if $T^{1-(d_0^2-3d_0+9/4)^{-1}}/m\to 0$, which holds under Assumption A4, then (A.6) is $o_p(1)$. The arguments to show that (A.7) is $o_p(1)$ are similar but applied to the case $\hat{d}_m\in[0,d_0-\delta)$.

Proof that the fifth term of (A.3) is $o_p(1)$. We have:

$$\frac{1}{m}\sum_{k=1}^{m}\log\left(\frac{1+(\theta_m/T)\lambda_k^{-2+2\hat{d}_m}}{1+(\theta_m/T)\lambda_k^{-2+2d_0}}\right)$$
$$=\frac{1}{m}\sum_{k=1}^{m}\log\left(\frac{1+(\theta_mT/(4\pi^2k^2))\lambda_k^{2\hat{d}_m}}{1+(\theta_mT/(4\pi^2k^2))\lambda_k^{2d_0}}\right)$$
$$=\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\log\left(\frac{1+(\theta_mT/(4\pi^2k^2))\lambda_k^{2\hat{d}_m}}{1+(\theta_mT/(4\pi^2k^2))\lambda_k^{2d_0}}\right)$$
$$+\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\log\left(\frac{1+(\theta_mT/(4\pi^2k^2))\lambda_k^{2\hat{d}_m}}{1+(\theta_mT/(4\pi^2k^2))\lambda_k^{2d_0}}\right). \tag{A.8}$$

It is easy to show that the second term of (A.8) is $o_p(1)$. For the first term,

$$\left|\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\log\left(\frac{1+(\theta_mT/(4\pi^2k^2))\lambda_k^{2\hat{d}_m}}{1+(\theta_mT/(4\pi^2k^2))\lambda_k^{2d_0}}\right)\right|$$
$$=\left|\frac{1}{m}\sum_{k=1}^{\sqrt{T}}[\log(1+(\theta_mT/(4\pi^2k^2))\lambda_k^{2\hat{d}_m})\right.$$
$$\left.-\log(1+(\theta_mT/(4\pi^2k^2))\lambda_k^{2d_0})]\right|$$
$$\le\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\log(1+\tilde{M}(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m}))$$
$$+\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\log(1+M_0(T^{1-2d_0}/k^{2-2d_0}))$$
$$\le\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\log(1+\tilde{M}T^{1-2\hat{d}_m})$$
$$+\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\log(1+M_0T^{1-2d_0})$$
$$\sim O_p\left(\frac{\sqrt{T}}{m}\log(T^{1-2\hat{d}_m})\right)+O_p\left(\frac{\sqrt{T}}{m}\log(T^{1-2d_0})\right)$$
$$\sim O_p\left(\frac{\sqrt{T}\log T}{m}\right)+O_p\left(\frac{\sqrt{T}\log T}{m}\right)=o_p(1).$$

This completes the proof of Theorem 1(a). For part (b), note that $J_m(\hat{d}_m,\theta_m)-J_m(d_0,\theta_m)=O_p(m^{-1}T^{(1/2)(d_0^2-3d_0+9/4)\curlyvee(1/2)})$. So if $m\ge O_p(T^\beta)$ with $\beta>(1/2)(d_0^2-3d_0+9/4)\curlyvee(1/2)$:

$$0\ge O_p(T^{(1/2)(d_0^2-3d_0+9/4)\curlyvee(1/2)-\beta})$$
$$-(1/2)\log(1+2(\hat{d}_m-d_0))+2(\hat{d}_m-d_0)$$
$$\ge O_p(T^{(1-(d_0^2-3d_0+9/4)^{-1})\curlyvee(1/2)-\beta})+(1/6)(\hat{d}_m-d_0)^2.$$

Hence $(1/6)(\hat{d}_m-d_0)^2\le O_p(T^{(1/2)(d_0^2-3d_0+9/4)\curlyvee(1/2)-\beta})$, so that $|\hat{d}_m-d_0|=o_p((\log m)^{-3})$ if $(T^{(1-(d_0^2-3d_0+9/4)^{-1})\curlyvee(1/2)-\beta})=o_p((\log m)^{-3})$ which is guaranteed if $\beta>(1-(d_0^2-3d_0+9/4)^{-1})\curlyvee(1/2)$, by Assumption A4. This completes the proof of Theorem 1 for the case in which LFC are present. If no LFC is present $\theta_0=0$, in which case all proofs go through with no requirement on θ_m, since the ratio $(1+M_0(T^{1-2d_0}/k^{2-2d_0}))/(1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m}))=1/(1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m}))$ is bounded for any choice of $M_m\ge 0$.

Proof of Lemma 2. Let $\bar{\theta} = \limsup \hat{\theta}_m$, $\bar{M} = (2\pi)^{2\hat{d}_m-2}\bar{\theta}$, $M_m = (2\pi)^{2\hat{d}_m-2}\hat{\theta}_m$, $T_{M_m} = \sup_k\{k|k^{2-2d_0}/T^{1-2d_0} \leq M_m\}$ and $T_{\bar{M}} = \sup_k\{k|k^{2-2d_0}/T^{1-2d_0} \leq \bar{M}\} = O_p(\bar{M}T^{(1-2d_0)/(2-2d_0)})$. Note that from Theorem 1, $\hat{d}_m \to d_0$. Then, (A.1) becomes

$$
\begin{aligned}
0 &= \left\{ \sum_{k=1}^{m} \left(\frac{\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_k^{-2}} \right) \right. \\
&\quad \times \left[1 - \left(m \frac{I_k}{f_k} \left(\frac{\lambda_k^{-2d_0}}{\lambda_k^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_k^{-2}} \right) \right) \right. \\
&\quad \left.\left. \Big\backslash \left(\sum_{j=1}^{m} \frac{I_j}{f_j} \left(\frac{\lambda_j^{-2d_0}}{\lambda_j^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_j^{-2}} \right) \right) \right] \right\} \\
&= \left\{ \sum_{k=1}^{m} \left(\frac{\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_k^{-2}} \right) \left[1 - \left(m \frac{I_k}{f_k} \lambda_k^{2\hat{d}_m - 2d_0} \right.\right.\right. \\
&\quad \left. \times \left(\frac{1}{1 + M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})} \right) \right) \\
&\quad \left.\left. \Big\backslash \left(\sum_{j=1}^{m} \frac{I_j}{f_j} \lambda_j^{2\hat{d}_m - 2d_0} \left(\frac{1}{1 + M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})} \right) \right) \right] \right\} \\
&\sim (2\pi)^{2-2\hat{d}_m} \left\{ \sum_{k=1}^{m} \left(\frac{T}{k} \right)^{2-2\hat{d}_m} \left[1 - \frac{I_k}{f_k} \frac{(k^{2-2\hat{d}_m}/T^{1-2\hat{d}_m})}{(k^{2-2\hat{d}_m}/T^{1-2\hat{d}_m}) + M_m} \right] \right\} \\
&\geq \sum_{k=1}^{m} \left(\frac{T}{k} \right)^{2-2\hat{d}_m} - \frac{1}{\bar{M}} \sum_{k=1}^{T_{M_m}} \left(\frac{T}{k} \right)^{2-2\hat{d}_m} \left| \frac{I_k}{f_k} \right| \frac{k^{2-2\hat{d}_m}}{T^{1-2\hat{d}_m}} \\
&\quad - \sum_{k=T_{M_m}+1}^{m} \left(\frac{T}{k} \right) 2 - 2\hat{d}_m \frac{I_k}{f_k} \\
&\geq \sum_{k=1}^{T_M} \left(\frac{T}{k} \right)^{2-2\hat{d}_m} - \frac{1}{\bar{M}} \sum_{k=1}^{T_{M_m}} T \left| \frac{I_k}{f_k} \right| \\
&\quad - \sum_{k=T_{M_m}+1}^{m} \left(\frac{T}{k} \right)^{2-2\hat{d}_m} \left| \frac{I_k}{f_k} - 1 \right| \\
&= T^{2-2d_0}(1 - T_{\bar{M}}^{2d_0-1}) - O_p(T^{1+(1-2d_0)/(2-2d_0)}) \\
&\quad - \sum_{k=T_{\bar{M}}+1}^{m} \frac{T^{2-2d_0} \log k}{k^{3-2d_0}} + o_p(1).
\end{aligned}
$$

If $\bar{\theta} > 0$, then $\bar{M} > 0$, and

$$
\begin{aligned}
&T^{2-2d_0}(1 - T_{\bar{M}}^{2d_0-1}) - O_p(T^{1+(1-2d_0)/(2-2d_0)}) \\
&\quad - \sum_{k=T_{\bar{M}}+1}^{m} \frac{T^{2-2d_0} \log k}{k^{3-2d_0}} \\
&> O_p(T^{1+(1-2d_0)}) - O_p(T^{1+(1-2d_0)/(2-2d_0)}) \\
&\quad - O_p(T^{2-2d_0} \log m(T_M^{2d_0-2} - m^{2d_0-2})) \to \infty.
\end{aligned}
$$

So the partial derivative with respect of θ will be always greater than zero and the objective function cannot be minimized at $\hat{\theta}_m$, which is a contradiction. Hence, $\hat{\theta}_m \xrightarrow{p} 0$ when there is no LFC in the data. To complete the proof, note that:

$$
\begin{aligned}
&T^{2-2d_0}(1 - T_{\bar{M}}^{2d_0-1}) - O_p(T^{1+(1-2d_0)/(2-2d_0)}) \\
&\quad - \sum_{k=T_{\bar{M}}+1}^{m} \frac{T^{2-2d_0} \log k}{k^{3-2d_0}} < 0
\end{aligned}
$$

so that $T_{\bar{M}}^{2d_0-1} \geq O_p(1)$. Hence,

$$
\begin{aligned}
T_{\bar{M}}^{1-2d_0} &= O_p(\bar{M}^{1-2d_0} T^{(1-2d_0)^2/(2-2d_0)}) \\
&= O_p(\bar{\theta}^{1-2d_0} T^{(1-2d_0)^2/(2-2d_0)}) \leq O_p(1)
\end{aligned}
$$

which implies that $\bar{\theta} = \limsup \hat{\theta}_m \leq O_p(T^{-(1-2d_0)/(2-2d_0)})$ and proves the result.

Proof of Theorem 2. The proof follows Robinson (1995a) with appropriate modifications to accommodate the extra term. Note that given Theorem 1(b), we can restrict the analysis to values of $\hat{d}_m$ in the set $C_m(d) = \{\hat{d}_m : |\hat{d}_m - d_0| < \log(m)^{-3}\}$ and $\hat{\theta}_m$ in the set $(0, \infty)$ by Lemma 1 when LFC are present. We can write the objective function as

$$
\begin{aligned}
J_m(d, \theta) &= \log \left(\frac{1}{m} \sum_{k=1}^{m} \frac{I_k}{\lambda_k^{-2d} + (\theta/T)\lambda_k^{-2}} \right) \\
&\quad + \frac{1}{m} \sum_{k=1}^{m} \log(\lambda_k^{-2d} + (\theta/T)\lambda_k^{-2}).
\end{aligned}
$$

Since $\hat{G}(d, \theta) = m^{-1} \sum_{k=1}^{m} (I_k/(\lambda_k^{-2d} + (\theta/T)\lambda_k^{-2}))$, we have

$$
\begin{aligned}
J'_m(d, \theta) &= \frac{\partial}{\partial d} J_m(d, \theta) \\
&= \frac{1}{\hat{G}(d, \theta)} \frac{2}{m} \sum_{k=1}^{m} \frac{I_k}{g_k} \log(\lambda_k) \frac{g_{yk}}{g_k} - \frac{2}{m} \sum_{k=1}^{m} \log(\lambda_k) \frac{g_{yk}}{g_k}
\end{aligned}
$$

and the second order derivative is

$$
\begin{aligned}
J''_m(d, \theta) &= \frac{\partial^2}{\partial d^2} J_m(d, \theta) \\
&= -\frac{4}{m^2} \frac{1}{\hat{G}(d, \theta)^2} \left(\sum_{k=1}^{m} \frac{I_k}{g_k} \frac{\lambda_k^{-2d}}{g_k} \log(\lambda_k) \right)^2 \\
&\quad + \frac{4}{\hat{G}(d, \theta)} \frac{1}{m} \sum_{k=1}^{m} \left(\frac{(\log(\lambda_k))^2}{(g_k)^3} I_k \lambda_k^{-2d} g_{yk} \right) \\
&\quad - \frac{4}{\hat{G}(d, \theta)} \frac{1}{m} \sum_{k=1}^{m} \left(\frac{(\log(\lambda_k))^2}{(g_k)^3} I_k \lambda_k^{-2d} g_{uk} \right) \\
&\quad + \frac{4}{m} \sum_{k=1}^{m} \left((\log(\lambda_k))^2 \frac{\lambda_k^{-2d} g_{uk}}{(g_k)^2} \right).
\end{aligned} \tag{A.9}
$$

We first show that when evaluated at $\hat{d}_m$ and $\hat{\theta}_m$ both terms of (A.9) are $o_p(1)$. For the first:

$$
\begin{aligned}
&-\frac{4}{\hat{G}(\hat{d}_m, \hat{\theta}_m)} \frac{1}{m} \sum_{k=1}^{m} \left(\frac{(\log(\lambda_k))^2}{(g_k)^3} I_k \lambda_k^{-2\hat{d}_m} g_{uk} \right) \\
&\sim \frac{1}{m} \sum_{k=1}^{m} \left(\log(\lambda_k)^2 \left(\frac{I_k}{g_k^0} \right) \left(\frac{g_k^0}{g_k} \right) \frac{\lambda_k^{-2\hat{d}_m} g_{uk}}{(g_k)^2} \right) \\
&\sim \frac{1}{m} \sum_{k=1}^{m} \left(\log(\lambda_k)^2 \left(\frac{I_k}{g_k^0} \right) \lambda_k^{2(2\hat{d}_m - d_0)} \right).
\end{aligned}
$$

For $\hat{d}_m$ in $C_m(d)$, we have for T and m large enough, $2\hat{d}_m - d_0 \geq (1/2)d_0$, so that the first term is $o_p(1)$. It is trivial to show that the second is $o_p(1)$. Hence, the second derivative of the objective function evaluated at $(\hat{d}_m, \hat{\theta}_m)$ is such that:

$$J_m''(\hat{d}_m, \hat{\theta}_m) = -\frac{4}{m^2}\frac{1}{\hat{G}(\hat{d}_m, \hat{\theta}_m)^2}\left(\sum_{k=1}^{m}\frac{I_k}{g_k}\frac{\lambda_k^{-2\hat{d}_m}}{g_k}\log(\lambda_k)\right)^2 + \frac{4}{\hat{G}(\hat{d}_m, \hat{\theta}_m)}\frac{1}{m}\sum_{k=1}^{m}\left(\frac{(\log(\lambda_k))^2}{(g_k)^3}I_k\lambda_k^{-2\hat{d}_m}g_{yk}\right) + o_p(1).$$

Let $\hat{G}_l(\hat{d}_m, \hat{\theta}_m) = m^{-1}\sum_{k=1}^{m}(I_k/g_k)(\log(\lambda_k))^l(g_{yk}/g_k)^l$, then

$$J_m''(\hat{d}_m, \hat{\theta}_m) = \frac{4}{\hat{G}_0(\hat{d}_m, \hat{\theta}_m)}[\hat{G}_0(\hat{d}_m, \hat{\theta}_m)\hat{G}_2(\hat{d}_m, \hat{\theta}_m) - \hat{G}_1(\hat{d}_m, \hat{\theta}_m)^2] + o_p(1).$$

Defining $\tilde{G}_l(\hat{d}_m, \hat{\theta}_m) = m^{-1}\sum_{k=1}^{m}(I_k/g_k)(\log(\lambda_k))^l$, we will show that $\hat{G}_l(\hat{d}_m, \hat{\theta}_m) = \tilde{G}_l(\hat{d}_m, \hat{\theta}_m) + o_p(\tilde{G}_l(\hat{d}_m, \hat{\theta}_m))$, for $l = 0, 1, 2$. When $l = 1$,

$$\begin{aligned}\hat{G}_1(\hat{d}_m, \hat{\theta}_m) &= \frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_k}{g_k}\right)\log(\lambda_k)\left(\frac{g_{yk}}{g_k}\right)\\ &= \frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_k}{g_k}\right)\log(\lambda_k) - \frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_k}{g_k}\right)\log(\lambda_k)\left(\frac{g_{uk}}{g_k}\right)\\ &= \tilde{G}_1(\hat{d}_m, \hat{\theta}_m) + o_p(\tilde{G}_1(\hat{d}_m, \hat{\theta}_m))\end{aligned}$$

if $[m^{-1}\sum_{k=1}^{m}(I_k/g_k)\log(\lambda_k)(g_{uk}/g_k)]/[m^{-1}\sum_{k=1}^{m}(I_k/g_k)\log(\lambda_k)] \to 0$, which we now prove.

$$\begin{aligned}&\frac{m^{-1}\sum_{k=1}^{m}(I_k/g_k)\log(\lambda_k)(g_{uk}/g_k)}{m^{-1}\sum_{k=1}^{m}(I_k/g_k)\log(\lambda_k)}\\ &= \frac{m^{-1}\sum_{k=1}^{\sqrt{T}-1}(I_k/g_k)\log(\lambda_k)(g_{uk}/g_k)}{\tilde{G}_1(\hat{d}_m, \hat{\theta}_m)} + \frac{m^{-1}\sum_{k=\sqrt{T}}^{m}(I_k/g_k)\log(\lambda_k)(g_{uk}/g_k)}{\tilde{G}_1(\hat{d}_m, \hat{\theta}_m)}.\end{aligned} \tag{A.10}$$

For the first term,

$$\begin{aligned}&\left|\frac{m^{-1}\sum_{k=1}^{\sqrt{T}-1}(I_k/g_k)\log(\lambda_k)(g_{uk}/g_k)}{\tilde{G}_1(\hat{d}_m, \hat{\theta}_m)}\right|\\ &= \frac{\sum_{k=1}^{\sqrt{T}-1}(I_k/g_k^0)(g_k^0/g_k)(g_{uk}/g_k)\log(\lambda_k)}{\sum_{k=1}^{m}(I_k/g_k^0)(g_k^0/g_k)\log(\lambda_k)}\\ &= \frac{\sum_{k=1}^{\sqrt{T}-1}(I_k/g_k^0)(k/T)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\log(\lambda_k)(g_{uk}/g_k)}{\sum_{k=1}^{m}(I_k/g_k^0)(k/T)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\log(\lambda_k)}\\ &\le \frac{\sum_{k=1}^{\sqrt{T}-1}(I_k/g_k^0)(k/T)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\log(\lambda_k)}{\sum_{k=1}^{m}(I_k/g_k^0)(k/T)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\log(\lambda_k)}\end{aligned}$$

$$= O_p\left(\frac{\log(T)\sum_{k=1}^{\sqrt{T}-1}(k/m)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})} - \sum_{k=1}^{\sqrt{T}-1}(k/m)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\log(k)}{\log(T)\sum_{k=\sqrt{T}}^{m}(k/m)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})} - \sum_{k=\sqrt{T}}^{m}(k/m)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\log(k)}\right)$$

$$= O_p\left(\frac{\sum_{k=1}^{\sqrt{T}-1}(k/m)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}}{\sum_{k=\sqrt{T}}^{m}(k/m)^{2(\hat{d}_m-d_0)}}\right).$$

From Robinson (1995a, Eq. (3.7)) and a result in the proof of consistency, $\sum_{k=\sqrt{T}}^{m}(k/m)^{2(\hat{d}_m-d_0)} \sim \sum_{k=1}^{m}(k/m)^{2(\hat{d}_m-d_0)} = m(1+2(\hat{d}_m-d_0))^{-1} + o_p(m) \sim m$ for $\hat{d}$ in $C_m(d)$. Hence,

$$\frac{\sum_{k=1}^{\sqrt{T}-1}(k/m)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}}{\sum_{k=\sqrt{T}}^{m}(k/m)^{2(\hat{d}_m-d_0)}} \sim \frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}\left(\frac{k}{m}\right)^{2(d-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}$$

which is $o_p(1)$ under Assumption A4 from the proof of consistency. For the second term in (A.10), using similar arguments, we have:

$$\begin{aligned}&\frac{m^{-1}\sum_{k=\sqrt{T}}^{m}(I_k/g_k)\log(\lambda_k)(g_{uk}/g_k)}{\tilde{G}_1(\hat{d}_m, \hat{\theta}_m)}\\ &\sim \frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\times\frac{(\hat{\theta}_m/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}.\end{aligned} \tag{A.11}$$

If $d_0 > 0$, then (A.11) is asymptotically equivalent to $m^{-1}\sum_{k=\sqrt{T}}^{m}(k/m)^{2(\hat{d}_m-d_0)}(k/T)^{2\hat{d}_m} = (m/T)^{2\hat{d}_m}m^{-1}\sum_{k=\sqrt{T}}^{m}(k/m)^{2(2\hat{d}_m-d_0)} \to 0$, for $\hat{d}_m$ in $C_m(d)$. If $d_0 = 0$, then (A.11) is asymptotically equivalent to $(T/m)\sum_{k=\sqrt{T}}^{m}(k/m)^4(\hat{\theta}_m/k^2) \sim (T/m^5)\sum_{k=\sqrt{T}}^{m}k^2 \sim (T/m^5)m^3 = (T/m^2) \to 0$. Hence, both terms of (A.10) are $o_p(1)$ and we have $\hat{G}_1(\hat{d}_m, \hat{\theta}_m) = \tilde{G}_1(\hat{d}_m, \hat{\theta}_m) + o(\tilde{G}_1(\hat{d}_m, \hat{\theta}_m))$. Similarly, $\hat{G}_2(\hat{d}_m, \hat{\theta}_m) = \tilde{G}_2(\hat{d}_m, \hat{\theta}_m) + o(\tilde{G}_2(\hat{d}_m, \hat{\theta}_m))$. Accordingly,

$$\begin{aligned}&\frac{\partial^2}{\partial d^2}J_m(\hat{d}_m, \hat{\theta}_m)\\ &\sim \frac{4}{\tilde{G}_0(\hat{d}_m, \hat{\theta}_m)^2}[\tilde{G}_0(\hat{d}_m, \hat{\theta}_m)\tilde{G}_1(\hat{d}_m, \hat{\theta}_m) - \tilde{G}_1(\hat{d}_m, \hat{\theta}_m)^2]\\ &= \frac{4}{\left(m^{-1}\sum_{k=1}^{m}(I_k/g_k)\right)^2}\left[\left(\frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_k}{g_k}\right)\right)\right.\end{aligned}$$

$$\times\left(\frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_k}{g_k}\right)\log^2(\lambda_k)\right)-\left(\frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_k}{g_k}\right)\log(\lambda_k)\right)^2\Bigg]$$

$$=\frac{4}{\left(m^{-1}\sum_{k=1}^{m}(I_k/g_k)\right)^2}\left[\left(\left(\frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_k}{g_k}\right)\right)\right.\right.$$

$$\left.\times\left(\frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_k}{g_k}\right)\log^2(k)\right)-\left(\frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_k}{g_k}\right)\log(k)\right)^2\right].$$

Let $\hat{F}_l(\hat{d}_m,\hat{\theta}_m)=(m^{-1}\sum_{k=1}^{m}(I_k/g_k)\log(k))^l$, $h_k=h_k(\hat{d}_m,\hat{\theta}_m)=1+(\hat{\theta}_m/T)\lambda_k^{-2+2\hat{d}_m}=1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})$, and $h_k^0=h_k(d_0,\theta_0)=1+M_0(T^{1-2d_0}/k^{2-2d_0})$. Then

$$\frac{I_k}{g_k}=\frac{I_k}{1+(\hat{\theta}_m/T)\lambda_k^{-2+2\hat{d}_m}}\left[k^{2\hat{d}_m}\left(\frac{2\pi}{T}\right)^{2\hat{d}_m}\right]$$
$$=\frac{I_k}{h_k}\left[k^{2\hat{d}_m}\left(\frac{2\pi}{T}\right)^{2\hat{d}_m}\right].$$

For $\tau=0,1,2$,

$$|\hat{F}_\tau(\hat{d}_m,\hat{\theta}_m)-\hat{F}_\tau(d_0,\theta_0)|$$
$$=\left|\frac{1}{m}\sum_{k=1}^{m}I_k\log^\tau(k)\left[\frac{1}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}-\frac{1}{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}\right]\right|$$
$$=\left|\frac{1}{m}\sum_{k=1}^{m}\frac{I_k\log^\tau(k)}{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}\left(\frac{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}-1\right)\right|$$
$$\le\left|\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\frac{I_k\log^\tau(k)}{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}\left(\frac{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}-1\right)\right|$$
$$+\left|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\frac{I_k\log^\tau(k)}{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}\times\left(\frac{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}-1\right)\right|. \tag{A.12}$$

For the first term of (A.12), we have

$$\left|\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\frac{I_k\log^\tau(k)}{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}\left(\frac{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}-1\right)\right|$$
$$\le\left|\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\frac{I_k\log^\tau(k)}{g_k^0}\frac{g_k^0}{g_k}\right|+\left|\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\frac{I_k\log^\tau(k)}{g_k^0}\right|$$
$$\le\left|\frac{\log^\tau(m)}{m}\sum_{k=1}^{\sqrt{T}}\frac{I_k}{g_k^0}\lambda_k^{2(\hat{d}_m-d_0)}\frac{h_k(d_0,\theta_0)}{h_k(\hat{d}_m,\hat{\theta}_m)}\right|+\left|\frac{\log^\tau(m)}{m}\sum_{k=1}^{\sqrt{T}}\frac{I_k}{g_k^0}\right|$$
$$\sim\left|\frac{\log^\tau(m)}{m}\left(\frac{T}{m}\right)^{2(d_0-\hat{d}_m)}\sum_{k=1}^{\sqrt{T}}\left(\frac{k}{m}\right)^{2(\hat{d}_m-d_0)}\frac{h_k(d_0,\theta_0)}{h_k(\hat{d}_m,\hat{\theta}_m)}\right|+O_p\left(\frac{\log^\tau(m)}{m}\sqrt{T}\right).$$

Note that from results in the proof for consistency and the fact that $\hat{d}_m$ is in $C_m(d)$, this last term is $o_p(1)$ if A4 holds. For the second term of (A.12), we have

$$\left|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\frac{I_k\log^\tau(k)}{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}\left(\frac{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}-1\right)\right|$$
$$\sim\left|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}-1\right)\log^\tau(k)\right|$$

and the first derivative of the second term of (A.12) is

$$\left|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\frac{g_k^0}{g_k}\log(\lambda_k)\frac{\lambda_k^{-2\hat{d}_m}}{g_k}(-2)\log^\tau(k)\right|$$
$$=O_p\left(\left|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\frac{g_k^0}{g_k}(\log k-\log T)\log^\tau(k)\right|\right)$$
$$=O_p\left(\left|\frac{\log T\log^\tau(m)}{m}\sum_{k=\sqrt{T}}^{m}\lambda_k^{2(\hat{d}_m-d_0)}\frac{I_k}{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}\right|\right)$$
$$\le\left|\frac{2\log m\log^\tau(m)}{m}\sum_{k=\sqrt{T}}^{m}\left(\frac{T}{k}\right)^{2|\hat{d}_m-d_0|}(2\pi)^{2(\hat{d}_m-d_0)}\right|$$
$$=O_p\left(\frac{\log^{\tau+1}(m)}{m}mT^{|\hat{d}_m-d_0|}\right)\le\log^{\tau+1}(m)m^{2|\hat{d}_m-d_0|}\to 0$$

since $\hat{d}_m$ in $C_m(d)$. Also, under A4 and for $\hat{d}_m$ in $C_m(d)$: $\log^{\tau+1}(m)m^{2|\hat{d}_m-d_0|}|\hat{d}_m-d_0|=o_p(\log^{\tau-2}(m)m^{2|\hat{d}_m-d_0|})\le o_p((\log m)^{\tau-2}m^{\log m-1})\le o_p((\log m)^{\tau-2})=o_p(1)$. Hence, the second term of (A.12) converges to 0, so that $|\hat{F}_\tau(\hat{d}_m,\hat{\theta}_m)-\hat{F}_\tau(d_0,\theta_0)|\xrightarrow{p}0$, and

$$\frac{\partial^2}{\partial d^2}J_m(\hat{d}_m,\hat{\theta}_m)\xrightarrow{p}\frac{4}{\hat{F}_0(d_0,\theta_0)^2}[\hat{F}_2(d_0,\theta_0)\hat{F}_0(d_0,\theta_0)-\hat{F}_1^2(d_0,\theta_0)].$$

We now show that $\hat{F}_\tau(d_0,\theta_0)=G_0m^{-1}\sum_{k=1}^{m}(\log k)^\tau+o_p(1)$. Using summations by parts,

$$\left|\hat{F}_\tau(d_0,\theta_0)-G_0\frac{1}{m}\sum_{k=1}^{m}(\log k)^\tau\right|$$
$$=\left|\frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_k}{g_k^0}\right)\log(k)^\tau-G_0\frac{1}{m}\sum_{k=1}^{m}(\log k)^\tau\right|$$
$$\le\frac{G_0}{m}\sum_{r=1}^{m-1}\left(\left|(\log r)^k-(\log(r+1))^k\right|\left|\sum_{k=1}^{r}\left(\frac{I_k}{g_k^0}-1\right)\right|\right)+\frac{G_0}{m}(\log m)^\tau\left|\sum_{k=1}^{m}\left(\frac{I_k}{g_k^0}-1\right)\right|$$
$$\le\frac{G_0}{m}\sum_{r=1}^{m-1}\left(\left|(\log(r+1))^{\tau-1}\right|\left|\frac{1}{r}\sum_{k=1}^{r}\left(\frac{I_k}{g_k^0}-1\right)\right|\right)+o_p(1)$$
$$=\frac{G_0}{m}\sum_{r=1}^{T^{1/2+\varepsilon}}\left(\left|(\log(r+1))^{\tau-1}\right|\left|\frac{1}{r}\sum_{k=1}^{r}\left(\frac{I_k}{g_k^0}-1\right)\right|\right)$$

$$+\frac{G_0}{m}\sum_{r=T^{1/2+\varepsilon}+1}^{m-1}\left(\left|(\log(r+1))^{\tau-1}\right|\left|\frac{1}{r}\sum_{k=1}^{r}\left(\frac{I_k}{g_k^0}-1\right)\right|\right)$$
$$+o_p(1)$$
$$=O_p\left(\frac{G_0}{m}\sum_{r=1}^{T^{1/2+\varepsilon}}\left(\left|(\log(r+1))^{\tau-1}\right|\right)\right)$$
$$+\frac{G_0}{m}\sum_{r=T^{1/2+\varepsilon}+1}^{m-1}\left(\left|(\log(r+1))^{\tau-1}\right|\right)o_p$$
$$\times((\log(r+1))^{-2})+o_p(1)$$
$$=O_p\left(G_0\frac{T^{1/2+\varepsilon}}{m}(\log(m+1))^{\tau-1}\right)$$
$$+o_p(G_0(\log(T^{1/2+\varepsilon}+1))^{\tau-2})+o_p(1)=o_p(1).$$

So

$$\frac{\partial^2}{\partial d^2}J_m(\hat{d}_m,\hat{\theta}_m)=4\left[\frac{1}{m}\left(\sum_{k=1}^{m}(\log k)^2\right)-\left(\frac{1}{m}\sum_{k=1}^{m}(\log k)\right)^2\right]+o_p(1)\to 4.$$

Note that the derivations above are valid so long as the sequence $\{\hat{\theta}_m\}$ is bounded below from zero, which holds if LFC are present. Now because $\hat{G}(d_0,\hat{\theta}_m)=m^{-1}\sum_{k=1}^{m}I_k[\lambda_k^{-2d_0}+(\hat{\theta}_m/T)\lambda_k^{-2}]^{-1}\xrightarrow{p}G_0$, from the fact that the second term in (A.3) is $o_p(1)$ in the proof of Theorem 1, then using similar arguments as in Robinson (1995a), we have

$$m^{1/2}\frac{\partial}{\partial d}J_m(d_0,\hat{\theta}_m)=m^{-1/2}\sum_{k=1}^{m}\left(\frac{I_k}{f_k}\frac{g_{y,k}}{g_k}+\frac{g_{u,k}}{g_k}\right)\nu_k+o_p(1)$$
$$=m^{-1/2}\sum_{k=1}^{m}\left(\left(\frac{I_k}{f_k}-1\right)\frac{g_{y,k}}{g_k}\right)\nu_k+o_p(1)$$
$$=m^{-1/2}\sum_{k=1}^{m}\left(\left(\frac{I_{yk}}{f_{yk}}-1\right)\frac{g_{y,k}}{g_k}\right)\nu_k$$
$$+m^{-1/2}\sum_{k=1}^{m}\left(\left(\frac{I_k}{f_k}-\frac{I_{yk}}{f_{yk}}\right)\frac{g_{y,k}}{g_k}\right)\nu_k+o_p(1)$$

where $\nu_k=[\log k-(m^{-1}\sum_{j=1}^{m}\log j)]$. Using the same approach as in Robinson (1995a, pp. 1644–1653), the first part converges to a $N(0,4)$ (note that for the part involving the 4th cumulant $cum(\omega_j/f_j,\omega_k/f_k,\bar{\omega}_j/f_j,\bar{\omega}_k/f_k)$ we need to use the results of Lemmas A.1 and A.2 to get the corresponding results for the DGP with LFC). What remains to be shown is that the second part is $o_p(1)$. We have, where $\tilde{I}_{uk}\doteq I_k-I_{yk}$:

$$m^{-1/2}\sum_{k=1}^{m}\left(\left(\frac{I_k}{f_k}-\frac{I_{yk}}{f_{yk}}\right)\frac{g_{y,k}}{g_k}\right)\nu_k$$
$$=m^{-1/2}\sum_{k=1}^{m}\left(\left(1-\frac{I_{yk}}{f_{yk}}\right)\frac{g_{u,k}}{g_k}\frac{g_{y,k}}{g_k}\right)\nu_k$$
$$+m^{-1/2}\sum_{k=1}^{m}\left(\left(\frac{I_k-I_{yk}-f_{uk}}{f_{uk}}\right)\frac{g_{u,k}}{g_k}\frac{g_{y,k}}{g_k}\right)\nu_k$$
$$=m^{-1/2}\sum_{k=1}^{m}\left(\left(\frac{\tilde{I}_{uk}-f_{uk}}{f_{uk}}\right)\frac{g_{u,k}}{g_k}\frac{g_{y,k}}{g_k}\right)\nu_k+o_p(1)$$
$$=m^{-1/2}\sum_{k=1}^{m}\left(\left(\frac{\frac{k^2}{T}\tilde{I}_{uk}-M_mG_0}{M_mG_0}\right)\frac{g_{u,k}}{g_k}\frac{g_{y,k}}{g_k}\right)\nu_k+o_p(1)$$
$$=\frac{1}{TG_0}m^{-1/2}\sum_{k=1}^{m}\left[\left(\frac{k^2}{T}\tilde{I}_{uk}-M_mG_0\right)\left(\frac{\lambda_k^{-2}}{g_k}\right)^2(\lambda_k^{2-2d_0}\nu_k)\right]$$
$$+o_p(1)=o_p(1)$$

using summations by parts and $M_mG_0=[\sum_{k=1}^{m}(\lambda_k^{-2}/g_k)^2(k^2/T)\tilde{I}_{uk}/\sum_{k=1}^{m}(\lambda_k^{-2}/g_k)^2]$. Hence, a CLT can be applied so that $\sqrt{m}(\partial/\partial d)J_m(d_0,\hat{\theta}_m)\xrightarrow{d}N(0,4)$. Thus from

$$\frac{\partial}{\partial d}J_m(d_0,\hat{\theta}_m)=\frac{\partial}{\partial d}J_m(\hat{d}_m,\hat{\theta}_m)+\frac{\partial^2}{\partial d^2}J_m(\hat{d}_m,\hat{\theta}_m)(\hat{d}_m-d_0)$$

and the fact that $(\partial/\partial d)J_m(\hat{d}_m,\hat{\theta}_m)=0$, we have

$$\sqrt{m}(\hat{d}_m-d_0)\xrightarrow{d}\frac{\sqrt{m}\partial(J_m(d_0,\hat{\theta}_m))\partial d}{\partial^2(J_m(\hat{d}_m,\hat{\theta}_m))/\partial d^2}=N(0,1/4).$$

This completes the proof of Theorem 2 for the case with LFC present. To complete the proof for the case with no LFC, we need to show that

$$m^{-1/2}\sum_{k=1}^{m}\left(\left(\frac{I_k}{f_k}-\frac{I_{yk}}{f_{yk}}\right)\frac{g_{yk}}{g_k}\right)\upsilon_k=o_p(1). \tag{A.13}$$

Note that with no LFC, we have $I_k=I_{yk}$ and $\tilde{I}_{uk}=0$. Hence,

$$m^{-1/2}\sum_{k=1}^{m}\left(\left(\frac{I_k}{f_k}-\frac{I_{yk}}{f_{yk}}\right)\frac{g_{yk}}{g_k}\right)\upsilon_k$$
$$=-m^{-1/2}\sum_{k=1}^{m}\frac{g_{uk}}{g_k}\frac{g_{yk}}{g_k}\upsilon_k+o_p(1).$$

So we want to show that $m^{-1/2}\sum_{k=1}^{m}(g_{uk}/g_k)(g_{yk}/g_k)\upsilon_k=o_p(1)$. To that effect, it suffices to show that $m^{-1/2}\sum_{k=1}^{m}(g_{uk}/g_k)\upsilon_k=o_p(1)$. To prove this, note that

$$m^{-1/2}\sum_{k=1}^{m}\frac{g_{uk}}{g_k}\upsilon_k=m^{-1/2}\sum_{k=1}^{m}\frac{(\hat{\theta}_m/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}\upsilon_k$$
$$=m^{-1/2}\sum_{k=1}^{m}\frac{(\hat{\theta}_m/T)\lambda_k^{-2}}{\lambda_k^{-2d_0}+(\hat{\theta}_m/T)\lambda_k^{-2}}\upsilon_k+o_p(1)$$
$$=m^{-1/2}\sum_{k=1}^{m}\frac{M_m(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2d_0}/k^{2-2d_0})}\upsilon_k+o_p(1)$$

since $\hat{d}_m=d_0+O_p(m^{-1/2})$. Now, let $T_\theta=\sup_k\{k|M_m(k^{2-2d_0}/T^{1-2d_0})=O_p(1)\}$. We have

$$T_\theta=O_p(T^{(1-2d_0)/(2-2d_0)}\hat{\theta}_m^{1/(2-2d_0)})$$
$$\leq O_p(T^{(1-2d_0)/(2-2d_0)}T^{-[(1-2d_0)/(2-2d_0)]/(2-2d_0)})$$
$$=O_p(T^{((1-2d_0)/(2-2d_0))^2})$$

using Lemma 2. Then,

$$m^{-1/2}\sum_{k=1}^{m}\frac{M_m(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2d_0}/k^{2-2d_0})}\upsilon_k$$
$$=m^{-1/2}\sum_{k=1}^{T_\theta}\frac{M_m(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2d_0}/k^{2-2d_0})}\upsilon_k$$
$$+m^{-1/2}\sum_{k=T_\theta+1}^{m}\frac{M_m(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2d_0}/k^{2-2d_0})}\upsilon_k$$
$$\leq m^{-1/2}T_\theta+m^{-1/2}\hat{\theta}_mT^{1-2d_0}T_\theta^{-(1-2d_0)}$$
$$=O_p(m^{-1/2}T^{((1-2d_0)/(2-2d_0))^2})=o_p(1)$$

from Assumption A4. Hence, (A.13) holds when there is no LFC, which completes the proof of Theorem 2.

References

Andrews, D.W.K., Sun, Y., 2004. Adaptive local polynomial Whittle estimation of long-range dependence. Econometrica 72, 569–614.

Arteche, J., 2004. Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models. J. Econometrics 119, 131–154.

Arteche, J., 2006. Semiparametric estimation in perturbed long memory series. Comput. Statist. Data Anal. 51, 2118–2141.

Beran, J., 1995. Maximum likelihood estimation of the differencing parameter for invertible short and long-memory autoregressive integrated moving average models. J. Roy. Statist. Soc. 57, 659–672.

Dahlhaus, R., 1989. Efficient parameter estimation for self similar processes. Ann. Statist. 17, 1749–1766.

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

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

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

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

Frederiksen, P., Nielsen, F.S., Nielsen, M.O., 2012. Local polynomial Whittle estimation of perturbed fractional proceses. J. Econometrics 167, 426–447.

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

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

Granger, C.W.J., Joyeux, R., 1980. An introduction to long memory time series models and fractional differencing. J. Time Ser. Anal. 1, 15–29.

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

Hurst, H., 1951. Long-term storage capacity of reservoirs. Trans. Am. Soc. Civil Eng. 116, 770–799.

Hurvich, C.M., Moulines, E., Soulier, P., 2005. Estimating long memory in volatility. Econometrica 73, 1283–1328.

Hurvich, C.M., Ray, B.K., 2003. The local Whittle estimator of long-memory stochastic volatility. J. Financ. Econom 1, 445–470.

Iacone, F., 2010. Local Whittle estimation of the memory parameter in presence of deterministic components. J. Time Ser. Anal. 31, 37–49.

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

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

McCloskey, A., 2013. Estimation of the long-memory stochastic volatility model parameters that is robust to level shifts and deterministic trends. J. Time Ser. Anal. 34, 285–301.

McCloskey, A., Hill, J.B., 2013, Parameter estimation robust to low-frequency contamination. Unpublished Manuscript, Department of Economics, Brown University..

McCloskey, A., Perron, P., 2013. Memory parameter estimation in the presence of level shifts and deterministic trends. Econometric Theory 29, 1196–1237.

Mikosch, T., Stărică, C., 2004. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. Rev. Econ. Stat. 86, 378–390.

Ohanissian, A., Russell, J., Tsay, R., 2008. True or spurious long memory? A new test. J. Bus. Econom. Statist. 26, 161–175.

Perron, P., Qu, Z., 2010. Long-memory and level shifts in the volatility of stock market return indices. J. Bus. Econom. Statist. 28, 275–290.

Qu, Z., 2011. A test against spurious long memory. J. Bus. Econom. Statist. 29, 423–438.

Qu, Z., Perron, P., 2013. A stochastic volatility model with random level shifts and its application to S&P 500 and NASDAQ return indices. Econom. J. 16, 309–339.

Robinson, P.M., 1995a. Gaussian semiparametric estimation of long range dependence. Ann. Statist. 23, 1630–1661.

Robinson, P.M., 1995b. Log-periodogram regression of time series with long range dependence. Ann. Statist. 23, 1048–1072.

Stărică, C., Granger, C.W.J., 2005. Nonstationarities in stock returns. Rev. Econ. Stat. 87, 503–522.

Varneskov, R.T., Perron, P., 2013, Combining long memory and level shifts in modeling and forecasting the volatility of asset returns. Unpublished Manuscript, Department of Economics, Boston University.

Xu, J., Perron, P., 2014. Forecasting return volatility: level shifts with varying jump probability and mean reversion. Int. J. Forecast. 30, 449–463.