Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend*

Katsumi Shimotsu
Department of Economics
University of Essex

September 23, 2002

Abstract

Recently, Shimotsu and Phillips (2002a) developed a new semiparametric estimator, the exact local Whittle (ELW) estimator, of the memory parameter (d) in fractionally integrated processes. The ELW estimator has been shown to be consistent and have the same $N(0, \frac{1}{4})$ limit distribution for all values of d. With economic applications in mind, we extend the ELW estimator so that it accommodates an unknown mean and a linear time trend. We show that the resulting feasible ELW estimator is consistent for $d > -\frac{1}{2}$ and has a $N(0, \frac{1}{4})$ limit distribution for $d \in (-\frac{1}{2}, 2)$ ($d \in (-\frac{1}{2}, \frac{7}{4})$ when the data has a linear trend) except for a few negligible intervals. A simulation study shows that the feasible ELW estimator inherits the desirable properties of the ELW estimator even in a small sample.

AMS 1991 subject classification: 62M10; *JEL Classification*: C22

1 Introduction

Fractionally integrated ($I(d)$) processes have attracted growing attention among empirical researchers in economics and finance. In part, this is because $I(d)$ processes provide an extension to the classical dichotomy of $I(0)$ and $I(1)$ time series and equip us with more general alternatives in terms of long-range dependence. Empirical research continues to find evidence that $I(d)$ processes can provide a suitable description of certain long range characteristics of economic and financial data (for a survey, see Henry and Zaffaroni 2002). Because of its flexibility in modeling temporal dependence, $I(d)$ processes also help to reconcile implications from economic models with observed data and have provided solutions for empirical "puzzles" in many areas in economics and finance, e.g., consumption (Diebold and Rudebusch 1991, Haubrich

*The author thanks Peter C. B. Phillips and Morten Ø. Nielsen for helpful comments and the Cowles Foundation for hospitality during his stay from January to August 2002. This research was supported by ESRC under Grant R000223629. Simulations and empirical applications were performed in MATLAB.

1993), term structure (Backus and Zin 1993), international finance (Maynard and Phillips 2001), and economic growth (Michelacci and Zaffaroni 2000).

The memory parameter, d, plays a central role in the definition of fractional integration and is often the focus of empirical interest. Semiparametric estimation of d is appealing in empirical work because it is agnostic about the short-run dynamics of the process and hence is robust to the misspecification of short-run dynamics. Two common statistical procedures in this class are log periodogram regression and local Whittle estimation (Robinson 1995a, 1995b). Although these estimators are consistent for $d \in (\frac{1}{2}, 1]$ and asymptotically normally distributed for $d \in (\frac{1}{2}, \frac{3}{4})$, they are also known to exhibit nonstandard behavior when $d > \frac{3}{4}$. For instance, they have a nonnormal limit distribution for $d \in [\frac{3}{4}, 1]$, and they converge to unity in probability for $d > 1$ and are inconsistent (Kim and Phillips 1999, Phillips 1999b, Phillips and Shimotsu 2001). To avoid inconsistency and an unreliable basis for inference when d may be larger than $\frac{3}{4}$, a simple and commonly used procedure is to estimate d by taking first differences of the data, estimating $d-1$, and adding back one to the estimate $\widehat{d-1}$. However, if the data is trend stationary, i.e., $I(d)$ with $d \in [0, \frac{1}{2})$ around a linear time trend, taking a first difference of a time series reduces it to $I(d)$ with $d \in [-1, -\frac{1}{2})$. In this case, the local Whittle estimator converges either to the true parameter value or to 0 depending on the number of frequencies used in estimation (Shimotsu and Phillips 2002b).

These restrictions on the possible range of d pose problems for the analysis and interpretation of the estimates and put empirical researchers in a very awkward situation. For instance, consider a standard procedure of constructing a confidence interval by adding and subtracting a constant from an estimate. For an obvious reason, this gives a valid confidence interval only if the constructed interval falls within the range where the estimator has an asymptotic normal distribution. It is possible to avoid this problem by using both $\widehat{d}$ and $\widehat{d-1}$ and forming a conservative confidence interval by $[\widehat{d}-c, \widehat{d}+c] \cup [\widehat{d-1}+1-c, \widehat{d-1}+1+c]$. However, this procedure leads to an unduly wide confidence interval and blurred inference, especially in view of the possibility that these estimators are inconsistent.

Many economists and econometricians took part in the debate on whether economic time series are trend stationary or difference stationary, which remains rather inconclusive partly because of the low power and discontinuity in the data-generating model of the unit root tests. In the context of $I(d)$ processes, these questions are translated into whether $d \geq 1/2$ or $d < 1/2$, because $I(d)$ processes become nonstationary when $d \geq 1/2$. Therefore, testing whether $d \geq 1/2$ or $d < 1/2$ is of great interest, but neither using the raw data, nor differenced data, nor combining them can answer this question, because these procedures must assume either $d < 3/4$ or $d > 1/2$ prior to estimation.

Recently Shimotsu and Phillips (2002a) developed a new semiparametric estimator, the exact local Whittle (ELW) estimator, which seems to offer a good general purpose estimation procedure for the memory parameter that applies throughout the stationary and nonstationary regions of d. The ELW estimator is consistent, has the same $N(0, \frac{1}{4})$ limit distribution for all values of d, and provides a basis for constructing valid asymptotic confidence intervals for d that are valid regardless of the true value of the memory parameter.

Economic time series are often modeled with an unknown mean and a linear time trend. If the data have an unknown mean and a linear time trend, it might appear possible to estimate d by first regressing the data on a constant and a time trend and then applying the ELW estimator to the residual. However, this procedure results in an inconsistent estimate when d is large. A closer inspection reveals that the estimation error of the unknown mean (initial condition), rather than the coefficient of the trend, is the source of the inconsistency. An unknown mean needs to be estimated carefully in the ELW estimation.

The aim of this paper is to extend the ELW estimation so that it accommodates an unknown mean and a linear time trend. One approach, which we call *feasible* ELW estimation, appears promising. It combines two estimators of the unknown mean of the process, the sample mean and the first observation, depending on the value of d. Presence of a linear time trend is dealt with by prior detrending of the data. The feasible ELW estimator is shown to be consistent for $d > -\frac{1}{2}$ and have the same $N(0, \frac{1}{4})$ limit distribution for $d \in (-\frac{1}{2}, 2)$ ($d \in (-\frac{1}{2}, \frac{7}{4})$ when the data are detrended) excluding arbitrary small intervals around 0 and 1. The finite sample performance of the feasible ELW estimator inherits the desirable property of the ELW estimator, apart from a small increase in bias and variance when the data are detrended and d is close to 0.

The remainder of the paper is organized as follows. Section 2 briefly reviews ELW estimation. In Section 3, the problem of estimating d from regression residual and the importance of estimation of mean are discussed. In Section 4, two estimators for the unknown mean are compared, and the consistency and asymptotic normality of the feasible ELW estimator are demonstrated. Feasible ELW estimation for trending data is analyzed in Section 5. Section 6 reports some simulation results and gives an empirical application using the extended Nelson-Plosser data. Section 7 concludes the paper. Some technical results are collected in Appendix A in Section 8. Proofs are given in Appendix B in Section 9.

2 A model of fractional integration and ELW estimation

First we review the exact local Whittle estimation developed by Shimotsu and Phillips (2002a) because it serves as the basis for the following analysis. Consider the fractionally integrated process X_t generated by the model

$$\Delta^d X_t = (1-L)^d X_t = u_t I\{t \geq 1\}, \quad t = 1, 2, \ldots \tag{1}$$

where u_t is stationary with zero mean and spectral density $f_u(\lambda)$. Expanding the binomial in (1) gives the form

$$\sum_{k=0}^{t} \frac{(-d)_k}{k!} X_{t-k} = u_t I\{t \geq 1\}, \tag{2}$$

where

$$(d)_k = \frac{\Gamma(d+k)}{\Gamma(d)} = (d)(d+1)\ldots(d+k-1)$$

is Pochhammer's symbol for the forward factorial function and $\Gamma(\cdot)$ is the gamma function. When d is a positive integer, the series in (2) terminates, giving the usual formula for the model (1) in terms of the differences and higher order differences of X_t. An alternate form for X_t is obtained by inversion of (1), giving a valid representation for all values of d

$$X_t = \Delta^{-d} u_t I\{t \geq 1\} = (1-L)^{-d} u_t I\{t \geq 1\} = \sum_{k=0}^{t-1} \frac{(d)_k}{k!} u_{t-k}. \quad t = 1, 2, \ldots$$

Define the discrete Fourier transform (dft) and the periodogram of a time series a_t evaluated at the fundamental frequencies as

$$\begin{aligned} w_a(\lambda_s) &= \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} a_t e^{it\lambda_s}, \quad \lambda_s = \frac{2\pi s}{n}, s = 1, \ldots, n, \\ I_a(\lambda_s) &= w_a(\lambda_s) w_a(\lambda_s)^*. \end{aligned} \tag{3}$$

The (negative) Whittle likelihood of u_t based on frequencies up to λ_m and up to scale multiplication is

$$\sum_{j=1}^{m} \log f_u(\lambda_j) + \sum_{j=1}^{m} \frac{I_u(\lambda_j)}{f_u(\lambda_j)}, \tag{4}$$

where m is some integer less than n. When X_t is generated by (1) and $f_u(\lambda) \sim G$ for $\lambda \sim 0$, we can transform the likelihood function (4) to be data dependent yielding

$$Q_m^*(G, d) = \frac{1}{m} \sum_{j=1}^{m} \left[\log \left(G \lambda_j^{-2d} \right) + \frac{1}{G} I_{\Delta^d x}(\lambda_j) \right].$$

Concentrating $Q_m(G, d)$ with respect to G, the ELW estimator is defined as

$$d^* = \underset{d \in [\Delta_1, \Delta_2]}{\arg\min} \; R^*(d), \tag{5}$$

where Δ_1 and Δ_2 are the lower and upper bounds of the admissible values of d and

$$R^*(d) = \log G^*(d) - 2d \frac{1}{m} \sum_{1}^{m} \log \lambda_j, \quad G^*(d) = \frac{1}{m} \sum_{1}^{m} I_{\Delta^d x}(\lambda_j). \tag{6}$$

No restriction is imposed on Δ_1 and Δ_2 except that $-\infty < \Delta_1 < \Delta_2 < \infty$. In what follows, we distinguish the true values of the parameters by the notation $G_0 = f_u(0)$ and d_0. The ELW estimator has been shown to be consistent and asymptotically normally distributed for any $d_0 \in (\Delta_1, \Delta_2)$ under fairly mild assumptions on m and the stationary component u_t in (1):

Assumption 1

$$u_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j^{1/2} |c_j| < \infty, \quad C(1) \neq 0, \tag{7}$$

where $E(\varepsilon_t | F_{t-1}) = 0$, $E(\varepsilon_t^2 | F_{t-1}) = 1$ *a.s.*, $t = 0, \pm 1, \ldots,$ *in which* F_t *is the* σ*-field generated by* ε_s, $s \leq t$, *and there exists a random variable* ε *such that* $E\varepsilon^2 < \infty$ *and for all* $\eta > 0$ *and some* $K > 0$, $\Pr(|\varepsilon_t| > \eta) \leq K \Pr(|\varepsilon| > \eta)$.

Assumption 2 *As* $n \to \infty$,

$$\frac{1}{m} + \frac{m (\log n)^2 (\log m)^3}{n} + \frac{\log n}{m^\gamma} \to 0 \quad \text{for any } \gamma > 0.$$

Under (7), the spectral density of u_t is $f_u(\lambda) = \frac{1}{2\pi}|C(e^{i\lambda})|^2$ and clearly satisfies

$$f_u(\lambda) \sim f_u(0) \in (0, \infty) \quad \text{as } \lambda \to 0+. \tag{8}$$

See Shimotsu and Phillips (2002a) for comparison of the above assumptions with those in Robinson (1995b).

2.1 Lemma (Shimotsu and Phillips 2002a, Theorem 3.3)

Suppose X_t *is generated by (1) and Assumptions 1 and 2 hold. Then, for* $d_0 \in [\Delta_1, \Delta_2]$, $d^* \to_p d_0$ *as* $n \to \infty$.

Assumption 1′

(a) Assumption 1 holds and also

$$E(\varepsilon_t^3 | F_{t-1}) = \mu_3 \quad a.s., \quad E(\varepsilon_t^4 | F_{t-1}) = \mu_4, \quad t = 0, \pm 1, \ldots,$$

for finite constants μ_3 *and* μ_4.

(b) For some $\beta \in (0, 2]$,

$$f_u(\lambda) = f_u(0)(1 + O(\lambda^\beta)), \quad \text{as } \lambda \to 0+.$$

Assumption 2′ *As* $n \to \infty$,

$$\frac{1}{m} + \frac{m^{1+2\beta} (\log m)^2}{n^{2\beta}} + \frac{\log n}{m^\gamma} \to 0 \quad \text{for any } \gamma > 0.$$

2.2 Lemma (Shimotsu and Phillips 2002a, Theorem 3.5)

Suppose X_t *is generated by (1) and Assumptions 1′ and 2′ hold. Then, for* $d_0 \in (\Delta_1, \Delta_2)$,

$$m^{1/2}(d^* - d_0) \to_d N\left(0, \frac{1}{4}\right).$$

3 ELW estimation with unknown mean and time trend: the source of the problem

As shown above, the ELW estimator is consistent and asymptotically normally distributed if X_t is generated by (1). However, when a researcher models an economic time series, typically the mean is assumed to be unknown and it is often accompanied by a linear time trend, giving the following dgp:

$$X_t = X_0 + \mu t + X_t^0; \quad X_t^0 = (1 - L)^{-d_0} u_t I\{t \geq 1\}. \tag{9}$$

Now the model has two nuisance parameters, X_0 and μ. One possible way of dealing with them is to regress the data on a constant and a linear trend and apply the ELW estimator to the residuals $\widehat{X}_t$;

$$\widehat{X}_t = X_t - \widetilde{X}_0 - \widetilde{\mu}t; \quad \widetilde{\mu} = \frac{\sum_{t=1}^n (t-\bar{t})X_t}{\sum_{t=1}^n (t-\bar{t})^2}, \quad \widetilde{X}_0 = \overline{X} - \widetilde{\mu}\bar{t},$$

where $\overline{a}$ denote the sample average of a time series a_t. However, this approach does not necessarily yield a consistent estimate of d for large d_0. To clarify the problem, write down $\widehat{X}_t$ as a deviation from X_t :

$$\widehat{X}_t = X_t^0 + (X_0 - \widetilde{X}_0) + (\mu - \widetilde{\mu})t.$$

For $d \in [d_0 - 1/2,\, d_0 + 1/2]$, and $d \geq 1$, it can be shown that

$$\begin{aligned} \lambda_j^{d_0-d} w_{\Delta^d x^0}(\lambda_j) &= O_p(1), \\ \lambda_j^{d_0-d} w_{\Delta^d(x_0-\widehat{x}_0)}(\lambda_j) &= \xi_1 O(j^{d_0-1}), \\ \lambda_j^{d_0-d} w_{\Delta^d(\mu-\widetilde{\mu})t}(\lambda_j) &= \xi_2 O(j^{d_0-2}), \end{aligned}$$

where ξ_1, ξ_2 are $O_p(1)$ random variables. Therefore, when d_0 is larger than 1, the error from $w_{\Delta^d(x_0-\widehat{x}_0)}(\lambda_j)$ becomes so large that it distorts the signal from $w_{\Delta^d x^0}(\lambda_j)$ and makes the estimator inconsistent. The error from $w_{\Delta^d(\mu-\widetilde{\mu})t}(\lambda_j)$ creates less of a problem. Indeed, if the data are generated by (9) with $X_0 = 0$ and we apply the ELW estimator to the residuals from regressing X_t on t

$$\dot{X}_t = X_t - \dot{\mu}t; \quad \dot{\mu} = (\textstyle\sum_1^n t^2)^{-1} \sum_1^n tX_t,$$

then the estimator is consistent for $d_0 < 2$. Table 1 illustrates the above discussion by a simulation example. We generate the data according to (9) with $u_t \sim iidN(0,1)$, $X_0 = 0, 10$, and $\mu = 5$. Δ_1 and Δ_2 are set to -2 and 4. Sample size and m are chosen to be $n = 200$ and $m = n^{0.65} = 31$, and 1,000 replications are used. The first row reports the bias of the ELW estimator applied to $\widehat{X}_t$ when $X_0 = 10$. The second row reports the bias in estimating d by $\dot{X}_t$ when $X_0 = 0$.

Table 1. Monte Carlo simulation bias: $n = 200$, $m = n^{0.65} = 31$

d	-0.4	0.0	0.4	0.8	1.2	1.6	2.0
$X_0 = 10$	-0.0094	-0.0293	-0.0246	-0.0203	-0.0558	-0.2659	-0.6967
$X_0 = 0$	-0.0124	-0.0062	0.0072	0.0040	-0.0039	0.0033	-0.0031

Therefore, the ELW estimator can become inconsistent if the error in estimating X_0 is not controlled properly. Hence, to focus our attention, we analyze the estimation of d when the data are generated by (9) with $\mu = 0$, i.e., when the data have an unknown mean.

4 ELW estimation with unknown X_0

4.1 Two choices of $\widehat{X}_0$: $\overline{X}$ and X_1

We consider estimating d when the data X_t are generated by

$$X_t = X_0 + X_t^0; \quad X_t^0 = (1-L)^{-d_0} u_t I\{t \geq 1\}, \tag{10}$$

where X_0 is a random variable with a certain fixed distribution. Because u_t is mean zero, X_0 is both the mean and the initial condition of the process X_t. X_0 is unobservable and needs to be estimated by $\widehat{X}_0$. One candidate for $\widehat{X}_0$ is the sample average $\overline{X}$. For $d_0 > -\frac{1}{2}$, the error in estimating X_0 by $\overline{X}$ is

$$\overline{X} - X_0 = n^{-1} (1-L)^{-d_0-1} u_n I\{t \geq 1\} = O_p(n^{d_0-1/2}). \tag{11}$$

The magnitude of the error increases as d_0 increases, which leads to inconsistency for large d_0. Another candidate for $\widehat{X}_0$ is X_1, the first observation. In this case, the error in estimating X_0 is

$$X_1 - X_0 = (1-L)^{-d_0} u_1 I\{t \geq 1\} = u_1 = O_p(1).$$

Although X_1 is not a consistent estimator of X_0 regardless of the value of d_0, when $d_0 \geq \frac{1}{2}$ we have $\text{var}(X_n) \to \infty$ as $n \to \infty$ and X_n dominates u_1. Therefore, X_1 serves as an acceptable estimator of X_0 for large d_0 and complements $\overline{X}$.

We state the results more formally. When we estimate X_0 by $\widehat{X}_0$, the resulting estimator is defined as

$$\widehat{d} = \arg\min_{d \in \Theta} R(d), \tag{12}$$

where Θ is the space of the admissible values of d and

$$R(d) = \log \widehat{G}(d) - 2d \frac{1}{m} \sum_1^m \log \lambda_j, \quad \widehat{G}(d) = \frac{1}{m} \sum_1^m I_{\Delta^d(x-\widehat{x}_0)}(\lambda_j),$$

and $I_{\Delta^d(x-\widehat{x}_0)}(\lambda_j)$ is the periodogram of $X_t - \widehat{X}_0$. The ELW estimator with $\widehat{X}_0 = \overline{X}$ is consistent for $d_0 < 1$, and the ELW estimator with $\widehat{X}_0 = X_1$ is consistent for $d_0 \geq \frac{1}{2}$. The following theorems establish it.

Assumption 3a

$$\Theta = [\Delta_1, \Delta_2] \quad \textit{with} \; -1/2 < \Delta_1 < \Delta_2 < 1.$$

4.2 Theorem

Suppose X_t is generated by (10), Assumptions 1, 2, and 3a hold, and $\widehat{X}_0 = \overline{X}$. Then, for $d_0 \in [\Delta_1, \Delta_2]$, $\widehat{d} \to_p d_0$ as $n \to \infty$.

Assumption 3b

$$\Theta = [\Delta_1, \Delta_2] \quad \textit{with} \; 1/2 \leq \Delta_1 < \Delta_2 < \infty.$$

4.3 Theorem

Suppose X_t is generated by (10), Assumptions 1, 2, and 3b hold, and $\widehat{X}_0 = X_1$. Then, for $d_0 \in [\Delta_1, \Delta_2]$, $\widehat{d} \to_p d_0$ as $n \to \infty$.

4.4 Remarks

We assume $\Delta_1 > -1/2$, because the order of $\overline{X} - X_0$ is not given by (11) (indeed, it becomes $O_p(n^{-1}\log n)$) if $d_0 \leq -1/2$. For practical applications, this assumption is innocuous, because the ELW estimation does not require prior differencing of the data and the cases with $d_0 < 0$ do not occur in practice.

Intriguingly, when $d_0 \in [\frac{1}{2}, 1)$, $\widehat{d}$ with $\widehat{X}_0 = \overline{X}$ is still consistent, although $\overline{X}$ is an inconsistent estimate of X_0. Table 2 shows the finite sample performance of the above two estimators. The same simulation design is used as in Table 1, except that $\mu = 0$ and $X_0 = 10$. The ELW estimator with $\widehat{X}_0 = \overline{X}$ becomes negatively biased for large d, whereas the estimator with $\widehat{X}_0 = X_1$ appears to be inconsistent when d is negative.

Table 2. Monte Carlo simulation bias: $n = 200$, $m = n^{0.65} = 31$

d	-0.4	0.0	0.4	0.8	1.2	1.6	2.0
$\widehat{X}_0 = \overline{X}$	-0.0010	0.0033	0.0072	0.0121	-0.0644	-0.3771	-0.7862
$\widehat{X}_0 = X_1$	0.2981	0.0061	0.0007	-0.0019	-0.0053	0.0002	-0.0042

4.5 Feasible ELW estimation

The above results indicate that

1. $\overline{X}$ is an acceptable estimator of X_0 for small d_0;
2. X_1 is an acceptable estimator of X_0 for large d_0;
3. for $d_0 \in [\frac{1}{2}, 1)$, both $\overline{X}$ and X_1 are acceptable estimators of X_0.

Therefore, one promising approach for estimating d consistently for a wide range of d is to estimate X_0 with a certain combination of $\overline{X}$ and X_1. Specifically, estimating X_0 by

$$\widehat{X}_0(d) = \overline{X} \cdot I\{d < c\} + X_1 \cdot I\{d \geq c\}; \quad c = \tfrac{1}{2} + \Delta, \quad \Delta \in (0, \tfrac{1}{8}], \tag{13}$$

for a fixed Δ provides a desirable estimator. We call the ELW estimator with $\widehat{X}_0(d)$ the *feasible* ELW estimator. The feasible ELW estimator is consistent for $d_0 > -\frac{1}{2}$, although we need to exclude a small interval around 0 and 1 for technical reasons.

Assumption 3c *For arbitrary small* $\varepsilon > 0$,

$$\Theta = [\Delta_1, \Delta_2] \backslash ((-\varepsilon, \varepsilon) \cup (1-\varepsilon, 1+\varepsilon)) \quad \textit{with} \; -1/2 < \Delta_1 < \Delta_2 < \infty.$$

4.6 Theorem

Suppose X_t *is generated by (10), Assumptions 1, 2, and 3c hold, and* $\widehat{X}_0 = \widehat{X}_0(d)$. *Then, for* $d_0 \in \Theta$, $\widehat{d} \to_p d_0$ *as* $n \to \infty$.

The exclusion of $(-\varepsilon, \varepsilon)$ and $(1-\varepsilon, 1+\varepsilon)$ is necessary because a precise approximation of $w_{\Delta^{d}v}(\lambda_s)$ becomes rather difficult for d in the vicinity of 0 and 1. From the inspection of the proof, the consistency still holds if we let ε tend to zero slowly (e.g., $\varepsilon = (\log\log m)^{-1}$), hence this restriction poses no problems in practice.

4.7 Feasible ELW estimation: asymptotic distribution

To derive the asymptotic distribution of the feasible ELW estimator, we need to strengthen Assumption 3c to $\Delta_2 < 2$. This restriction is due to the asymptotic behavior of the derivatives of $w_{\Delta^{d_v}}(\lambda_s)$. Furthermore, we exclude the point $d_0 = c$ because $\widehat{X}_0(d)$ has a jump at $d = c$. Of course, we can easily modify $\widehat{X}_0(d)$ to be C^2 in d, so that consistency and asymptotic normality hold at $d_0 = c$. We keep $\widehat{X}_0(d)$ as defined above, however, to keep the proof simple.

Assumption 3c′ *For arbitrary small* $\varepsilon > 0$,

$$\Theta = [\Delta_1, \Delta_2] \backslash ((-\varepsilon, \varepsilon) \cup (1-\varepsilon, 1+\varepsilon) \cup \{c\}) \qquad \textit{with } -1/2 < \Delta_1 < \Delta_2 < 2.$$

Because $d_0 < 2$ for most, if not all, economic data, the feasible ELW estimator is asymptotically normally distributed for any value of d_0 encountered in practice. The following theorem establishes this.

4.8 Theorem

Suppose X_t *is generated by (10), Assumptions 1′, 2′, and 3c′ hold, and* $\widehat{X}_0 = \widehat{X}_0(d)$. *Then, for* $d_0 \in \mathrm{Int}(\Theta)$,

$$m^{1/2}\left(\widehat{d} - d_0\right) \to_d N\left(0, \frac{1}{4}\right).$$

5 ELW estimation with unknown X_0 and linear time trend

In this section we return to the estimation of d when the data are generated by (9). We propose to estimate d by first regressing X_t on a constant and a linear trend and then applying the feasible ELW estimation to the residuals $\widehat{X}_t$. Recall that the dgp is

$$X_t = X_0 + \mu t + X_t^0; \quad X_t^0 = (1-L)^{-d_0} u_t I\{t \geq 1\}, \tag{14}$$

and the residuals from the regression are

$$\widehat{X}_t = X_t - \widetilde{X}_0 - \widetilde{\mu} t; \quad \widetilde{\mu} = \frac{\sum_{t=1}^n (t-\bar{t}) X_t}{\sum_{t=1}^n (t-\bar{t})^2}, \quad \widetilde{X}_0 = \overline{X} - \widetilde{\mu}\bar{t}.$$

Define

$$\zeta_n(d_0) = \widetilde{\mu} - \mu = \frac{\sum_{t=1}^n (t-\bar{t})(X_0 + \mu t + X_t^0)}{\sum_{t=1}^n (t-\bar{t})^2} - \mu = \frac{\sum_{t=1}^n (t-\bar{t}) X_t^0}{\sum_{t=1}^n (t-\bar{t})^2},$$

then $\widehat{X}_t$ can be expressed as

$$\widehat{X}_t = X_t^0 - \overline{X^0} + (\mu - \widetilde{\mu})(t - \bar{t}) = X_t^0 - \left[\overline{X^0} - \zeta_n(d_0)\bar{t}\right] - \zeta_n(d_0) t.$$

When we apply the feasible ELW estimator to the residuals, the estimate of X_0 takes the form

$$\varphi(d) = \overline{\widehat{X}} \cdot I\{d < c\} + \widehat{X}_1 \cdot I\{d \geq c\} = \widehat{X}_1 \cdot I\{d \geq c\},$$

where the second equality follows from the fact that $\overline{\widehat{X}} = 0$. It follows that

$$\widehat{X}_t - \varphi(d) = \begin{cases} X_t^0 - \left[\overline{X^0} - \zeta_n(d_0)\bar{t}\right] - \zeta_n(d_0)t, & d < c, \\ X_t^0 - \left[X_1^0 - \zeta_n(d_0)\right] - \zeta_n(d_0)t, & d \geq c. \end{cases} \tag{15}$$

Now the dft of $\Delta^d(\widehat{X}_t - \varphi(d))$ has a term $\zeta_n(d_0)w_{\Delta^d t}(\lambda_j)$, but indeed it can be handled in a similar manner as the dft of $\Delta^d(X_t - \widehat{X}_0)$. The consistency of the feasible ELW estimator is not affected by prior detrending, but asymptotic normality requires d_0 to be smaller than 7/4. The following theorem establishes the asymptotics of the estimator.

5.1 Theorem

Suppose X_t is generated by (14) and $\widehat{X}_t - \varphi(d)$ is used in place of $X_t - \widehat{X}_0$ in defining $R(d)$ in (12). Then,
(a) If Assumptions 1, 2, and 3c hold, then, for $d_0 \in \Theta$, $\widehat{d} \to_p d_0$ as $n \to \infty$.
(b) If Assumptions 1′, 2′, and 3c′ hold, then, for $d_0 \in \{\text{Int}(\Theta) \cap (\Delta_1, \frac{4}{7})\}$, $m^{1/2}(\widehat{d} - d_0) \to_d N(0, \frac{1}{4})$.

6 Simulations and an empirical application

This section reports some simulations that were conducted to examine the finite sample performance of the feasible ELW estimator. Although estimation of the mean and trend coefficients does not affect the asymptotic property of the estimator, it is of interest to investigate its effect on the small sample performance of the estimator. Therefore, we design the simulation so that X_t is generated by

$$\begin{array}{ll} (1-L)^{-d} u_t I\{t \geq 1\}, & \text{for the ordinary ELW estimator,} \\ 10 + (1-L)^{-d} u_t I\{t \geq 1\}, & \text{for the feasible ELW estimator ,} \\ 10 + 5t + (1-L)^{-d} u_t I\{t \geq 1\}, & \text{for the feasible ELW estimator with detrending,} \end{array}$$

although this comparison is not favorable to the feasible ELW estimator. Δ is set to 0.1. u_t is generated as $iidN(0,1)$. Δ_1 and Δ_2 are set to -2 and 4. We use 10,000 replications, and n and m were chosen to be $n = 100$, 500 and $m = n^{0.65}$. We also compare the three ELW estimators with the local Whittle estimator with tapering studied by Hurvich and Chen (2000). Data tapering extends the range of consistent estimation of d, but at the cost of inflated bias.

Tables 3 and 4 show the simulation results. The estimation of the mean has little effect on the bias and standard deviation of the estimators, and the MSE of the ELW estimator and the feasible ELW estimator are virtually the same for $n = 500$. If the data are detrended prior to estimation, the feasible ELW suffers from a mild increase in standard deviation and a small negative bias for $d = 0.0$ and 0.4. Overall, the small sample performance of both the feasible ELW estimator and feasible ELW estimator with detrending is very close to that of the ELW estimator except for a few cases. On the other hand, the tapered estimator has substantially larger standard deviations and MSE compared with the ELW estimator for all values of d. In sum, the simulation evidence shows that the feasible ELW estimator's performance is comparable to the ELW estimator's.

Table 3. Simulation results: $n = 100$, $m = n^{0.65} = 19$

	ELW			**Feasible ELW**		
d	bias	s.d.	MSE	bias	s.d.	MSE
-0.4	0.0052	0.1602	0.0257	0.0012	0.1609	0.0259
0.0	0.0006	0.1599	0.0256	0.0019	0.1594	0.0254
0.4	0.0002	0.1606	0.0258	0.0124	0.1633	0.0268
0.8	-0.0010	0.1586	0.0252	0.0040	0.1492	0.0223
1.0	0.0009	0.1616	0.0261	0.0020	0.1583	0.0251
1.2	0.0003	0.1622	0.0263	0.0011	0.1608	0.0259
1.6	-0.0010	0.1605	0.0258	-0.0003	0.1599	0.0256
2.0	0.0008	0.1593	0.0254	0.0003	0.1595	0.0254
	FELW with detrending			**Tapered estimator**		
d	bias	s.d.	MSE	bias	s.d.	MSE
-0.4	-0.0104	0.1557	0.0243	0.0494	0.2028	0.0436
0.0	-0.0514	0.1703	0.0317	0.0185	0.2041	0.0420
0.4	-0.0461	0.1846	0.0362	-0.0050	0.2047	0.0419
0.8	-0.0176	0.1720	0.0299	-0.0207	0.2027	0.0415
1.0	-0.0049	0.1670	0.0279	-0.0203	0.2073	0.0434
1.2	0.0005	0.1627	0.0265	-0.0205	0.2044	0.0422
1.6	0.0109	0.1549	0.0241	-0.0136	0.1993	0.0399
2.0	0.0109	0.1400	0.0197	0.0265	0.1936	0.0382

Table 4. Simulation results: $n = 500$, $m = n^{0.65} = 56$

	ELW			**Feasible ELW**		
d	bias	s.d.	MSE	bias	s.d.	MSE
-0.4	-0.0017	0.0763	0.0058	-0.0036	0.0765	0.0059
0.0	-0.0028	0.0787	0.0062	-0.0027	0.0787	0.0062
0.4	-0.0021	0.0782	0.0061	0.0001	0.0786	0.0062
0.8	-0.0021	0.0780	0.0061	-0.0020	0.0777	0.0060
1.0	-0.0018	0.0772	0.0060	-0.0018	0.0773	0.0060
1.2	-0.0007	0.0783	0.0061	-0.0007	0.0782	0.0061
1.6	-0.0005	0.0774	0.0060	-0.0004	0.0774	0.0060
2.0	-0.0023	0.0777	0.0060	-0.0023	0.0777	0.0060
	FELW with detrending			**Tapered estimator**		
d	bias	s.d.	MSE	bias	s.d.	MSE
-0.4	-0.0079	0.0764	0.0059	0.0150	0.0967	0.0096
0.0	-0.0221	0.0824	0.0073	0.0033	0.0986	0.0097
0.4	-0.0188	0.0841	0.0074	-0.0038	0.0998	0.0100
0.8	-0.0064	0.0801	0.0065	-0.0081	0.0997	0.0100
1.0	-0.0029	0.0781	0.0061	-0.0084	0.0985	0.0098
1.2	-0.0002	0.0781	0.0061	-0.0063	0.0987	0.0098
1.6	0.0062	0.0771	0.0060	-0.0003	0.0972	0.0094
2.0	0.0004	0.0716	0.0051	0.0165	0.0957	0.0094

As an empirical illustration, the feasible ELW estimator with detrending was applied to the historical economic times series considered in Nelson and Plosser (1982) and extended by Schotman and van Dijk (1991). For comparison, we also estimate d by first taking the difference of the data, estimating $d-1$ by the local Whittle estimator, and adding unity to the estimate $\widehat{d-1}$. This procedure is invariant to the linear trend. For the feasible ELW estimates, 95% asymptotic confidence intervals are constructed by adding and subtracting $1.96 \times 1/\sqrt{4m}$ to the estimates. Table 4 shows the results based on $m = n^{0.7}$. The upper end of the confidence interval is larger than 7/4 in a few series, in which case the confidence intervals are only for reference. The feasible ELW estimate and the local Whittle estimate from the differenced data are fairly close to each other. For real measures such as real GNP, real per capita GNP, and employment, the estimates are close to 1. For price variables such as the GNP deflator, CPI, and nominal wage, the estimates are substantially larger than 1. This confirms previous empirical results (Hassler and Wolters, 1995) that inflations are $I(d)$ with $d \in (0, 1)$. Interestingly, the null of trend stationarity $\mathrm{H}_0 : d = 0$ is accepted in none of the series. Crato and Rothman (1994) obtained a similar result using the ARFIMA model, therefore it appears that the case for trend stationarity is weaker than has been suggested from the KPSS test by Kwiatkowski et al. (1992).

Table 5: Estimates of d for US Economic Data: $m = n^{0.7}$

	n	LW	FELW	95% asy. CI
Real GNP	80	1.077	1.126	[0.706, 1.545]
Nominal GNP	80	1.273	1.303	[0.884, 1.722]
Real per capita GNP	80	1.077	1.127	[0.708, 1.546]
Industrial production	129	0.821	0.850	[0.500, 1.201]
Employment	99	0.968	1.000	[0.608, 1.392]
Unemployment rate	129	0.951	0.980	[0.630, 1.331]
GNP deflator	100	1.374	1.398	[1.014, 1.782]
CPI	129	1.273	1.287	[0.937, 1.638]
Nominal wage	89	1.300	1.351	[0.951, 1.752]
Real wage	89	1.047	1.089	[0.688, 1.489]
Money stock	100	1.460	1.501	[1.117, 1.885]
Velocity of money	120	0.953	0.993	[0.630, 1.356]
Bond yield	89	1.091	1.108	[0.707, 1.508]
Stock prices	118	0.900	0.958	[0.595, 1.321]

7 Conclusion

By tailoring the ELW estimator developed by Shimotsu and Phillips (2002a) to accommodate an unknown mean and a linear time trend, this paper develops a very general purpose tool for estimation and inference of the memory parameter of typical economic time series. The new estimator, the feasible ELW estimator, covers a range of values of d that is commonly encountered in applied work with economic data and makes it possible to construct valid confidence intervals in a standard and simple way. Both in asymptotics and in small samples, the feasible ELW estimator inherits the desirable properties of the ELW estimator. A more extensive application of the feasible ELW estimator with other economic data is currently being undertaken by the author.

The restrictions on d ($d < 7/4$ for asymptotic normality and small intervals around 0 and 1) are somewhat bothersome. However, other semiparametric estimators (Robinson 1995a, 1995b, Velasco 1999, Hurvich and Chen 2000) are also liable to restrictions, and this estimator covers a wider range of d with the smallest variance for the same m. A possibly bound-free estimator is obtained by including the initial condition and time trend into the objective function and estimating d by

$$(\overline{d},\overline{\beta},\overline{X}_0)=\underset{d,\beta,X_0\in\Theta}{\arg\min}\ \overline{R}\left(d,\beta,X_0\right),$$

where

$$\begin{aligned}\overline{R}\left(d,\beta,X_0\right) &= \log\overline{G}\left(d,\beta,X_0\right)-2d\frac{1}{m}\sum_1^m\log\lambda_j,\\ \overline{G}\left(d,\beta,X_0\right) &= \frac{1}{m}\sum_1^m\left|w_{\Delta^d x}(\lambda_j)-w_{\Delta^d x_0}(\lambda_j)-\beta w_{\Delta^d t}(\lambda_j)\right|^2.\end{aligned}$$

The feasible ELW estimator is much simpler when compared with $\overline{d}$ and should not be outperformed by $\overline{d}$ substantially in small sample, because the finite sample properties of the feasible ELW estimator is already close to those of the ELW estimator.

8 Appendix A: Technical Lemmas

Lemma 8.1 gives an exact expression that we use for the model in frequency domain form. Some results from Phillips and Shimotsu (2001) and Shimotsu and Phillips (2002a) that are relevant to this paper are collected as Lemmas 8.2 - 8.7. We refer the reader to these papers for those proofs. In the following, C and ε denote generic constants such that $C\in(1,\infty)$ and $\varepsilon\in(0,1)$ unless specified otherwise.

8.1 Lemma (Phillips 1999, Theorem 2.2)

(a) If X_t follows (1), then

$$w_u\left(\lambda\right)=D_n\left(e^{i\lambda};d\right)w_x\left(\lambda\right)-\frac{1}{\sqrt{2\pi n}}e^{in\lambda}\widetilde{X}_{\lambda n}\left(d\right), \tag{16}$$

where $D_n(e^{i\lambda};d)=\sum_{k=0}^n\frac{(-d)_k}{k!}e^{ik\lambda}$ and

$$\widetilde{X}_{\lambda n}\left(d\right)=\widetilde{D}_{n\lambda}\left(e^{-i\lambda}L;d\right)X_n=\sum_{p=0}^{n-1}\widetilde{d}_{\lambda p}e^{-ip\lambda}X_{n-p},\qquad \widetilde{d}_{\lambda p}=\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda}. \tag{17}$$

(b) If X_t follows (1) with $d=1$, then

$$w_x\left(\lambda\right)\left(1-e^{i\lambda}\right)=w_u\left(\lambda\right)-\frac{e^{i\lambda}}{\sqrt{2\pi n}}e^{in\lambda}X_n. \tag{18}$$

8.2 Lemma (Shimotsu and Phillips 2002a, Lemmas 5.1 and 5.2)

Uniformly in $\theta\in[-1+\varepsilon,C]$ and in $s=1,2,\ldots,m$ with $m=o\left(n\right)$,

$$\begin{aligned}D_n\left(e^{i\lambda_s};\theta\right) &= \left(1-e^{i\lambda_s}\right)^\theta+O\left(n^{-\theta}s^{-1}\right), \qquad (19)\\ \lambda_s^{-\theta}\left(1-e^{i\lambda_s}\right)^\theta &= e^{-\frac{\pi}{2}\theta i}+O\left(\lambda_s\right),\\ \lambda_s^{-\theta}D_n\left(e^{i\lambda_s};\theta\right) &= e^{-\frac{\pi}{2}\theta i}+O\left(\lambda_s\right)+O\left(s^{-1-\theta}\right),\\ \lambda_s^{-2\theta}D_n\left|\left(e^{i\lambda_s};\theta\right)\right|^2 &= 1+O\left(\lambda_s^2\right)+O\left(s^{-1-\theta}\right).\end{aligned}$$

8.3 Lemma (Phillips and Shimotsu 2001, Lemma 8.11)

$$E\left|X_n - X_0\right|^2 = O(n^{2d-1}) \quad \textit{for } d \in (1/2, C].$$

8.4 Lemma (Shimotsu and Phillips 2002a, Lemma A)

Let $\widetilde{U}_{\lambda_j n}(d) = \sum_{p=0}^{n-1} \widetilde{d}_{\lambda p} e^{-ip\lambda} u_{n-p}$. *Then, for* $d \in [-1/2, 1/2]$ *and* $m = o(n)$,

$$\lambda_j^{-d} (2\pi n)^{-1/2} \widetilde{U}_{\lambda_j n}(d) = A_{nj}(d) + B_{nj}(d), \qquad s = 1, 2, \ldots, m,$$

where

$$\begin{cases} E\sup_d \left|A_{nj}(d) + B_{nj}(d)\right|^2 = O(j^{-1}(\log n)^6), & d \geq 0, \\ E\sup_d \left|A_{nj}(d)\right|^2 = O(j^{-1/2}(\log n)^4),\ E\sup_d \left|B_{nj}(d)\right|^2 = O(jn^{-1}(\log n)^2), & d \leq 0. \end{cases}$$

8.5 Lemma (Shimotsu and Phillips 2002a, Lemma 5.8)

Define $J_n(L) = \sum_1^n \frac{1}{k} L^k$. *Then*

$$J_n(L) = J_n(e^{i\lambda}) + \widetilde{J}_{n\lambda}(e^{-i\lambda}L)(e^{-i\lambda}L - 1),$$

where $\widetilde{J}_{n\lambda}(e^{-i\lambda}L) = \sum_{p=0}^{n-1} \widetilde{j}_{\lambda p} e^{-ip\lambda} L^p$ *and* $\widetilde{j}_{\lambda p} = \sum_{p+1}^{n} \frac{1}{k} e^{ik\lambda}$.

8.6 Lemma (Shimotsu and Phillips 2002a, Lemma 5.10)

Uniformly in $p = 1, \ldots, n$ *and* $s = 1, \ldots, m$ *with* $m = o(n)$,

$$(a) \qquad J_n(e^{i\lambda_s}) = -\log\lambda_s + \frac{i}{2}(\pi - \lambda_s) + O\left(\lambda_s^2\right) + O\left(s^{-1}\right),$$

$$(b) \qquad \widetilde{j}_{\lambda_s p} = O\left(|p|_+^{-1} n s^{-1}\right), \quad (c) \quad \widetilde{j}_{\lambda_s p} = O(\log n).$$

8.7 Lemma (from Shimotsu and Phillips 2002a, Lemma 5.12)

Suppose $Y_t = (1-L)^{\theta} u_t$. *Then, uniformly in* $s = 1, \ldots, m$ *with* $m = o(n)$,

$$\begin{array}{ll} (a) & -w_{\log(1-L)u}(\lambda_s) = J_n(e^{i\lambda_s}) w_u(\lambda_s) + r_{ns}, \\ (b) & E\sup_{\theta\in[-1/2,1/2]} \left|\lambda_j^{-\theta} w_{\log(1-L)y}(\lambda_s)\right|^2 = O\left((\log n)^8\right), \\ (c) & E\sup_{\theta\in[-1/2,1/2]} \left|\lambda_j^{-\theta} w_{(\log(1-L))^2 y}(\lambda_s)\right|^2 = O\left((\log n)^{10}\right), \end{array}$$

where $E\left|r_{ns}\right|^2 = O(s^{-1}(\log n)^8)$.

8.8 Lemma

Let $v_t = I\{t \geq 1\}$. *Then the following holds uniformly in* $s = 1, \ldots, m$ *with* $m = o(n)$ *and in* d:

$$\begin{array}{ll} (a) & w_{\Delta^d v}(\lambda_s) \\ = & \dfrac{e^{i\lambda_s}}{1 - e^{i\lambda_s}} \dfrac{1}{\sqrt{2\pi n}} \left[\left(1 - e^{i\lambda_s}\right)^d - \dfrac{n^{-d}}{\Gamma(1-d)} + O\left(n^{-d}s^{-1}\right)\right], \quad d \in [-1+\varepsilon, C], \\ (b) & -w_{\log(1-L)\Delta^d v}(\lambda_s) \end{array}$$

$$= \begin{cases} J_n(e^{i\lambda_s})w_{\Delta^d v}(\lambda_s) + O\left(s^{-1}n^{1/2-d}(\log n)^2\right), & d \in [-1+\varepsilon, 1], \\ J_n(e^{i\lambda_s})(1-e^{i\lambda_s})w_{\Delta^{d-1}v}(\lambda_s) + O\left(n^{1/2-d}(\log n)^2\right), & d \in [1,2], \end{cases}$$

$$= w_{\Delta^d v}(\lambda_s) \cdot O\left((\log n)^2\right), \quad d \notin (-\varepsilon, \varepsilon),$$

$$(c) \quad w_{(\log(1-L))^2 \Delta^d v}(\lambda_s)$$

$$= \begin{cases} J_n(e^{i\lambda_s})^2 w_{\Delta^d v}(\lambda_s) + O\left(s^{-1}n^{1/2-d}(\log n)^4\right), & d \in [-1+\varepsilon, 1], \\ J_n(e^{i\lambda_s})^2(1-e^{i\lambda_s})w_{\Delta^{d-1}v}(\lambda_s) + O\left(n^{1/2-d}(\log n)^4\right), & d \in [1,2], \end{cases}$$

$$= w_{\Delta^d v}(\lambda_s) \cdot O\left((\log n)^4\right), \quad d \notin (-\varepsilon, \varepsilon).$$

8.9 Proof

For part (a), from Lemma 8.1 (b), we have

$$w_{\Delta^d v}(\lambda_s) = \frac{1}{1-e^{i\lambda_s}}\left[w_{\Delta^{d+1}v}(\lambda_s) - \frac{e^{i\lambda_s}}{\sqrt{2\pi n}}\Delta^d v_n\right].$$

Observe that (note that $\frac{(-d)_0}{0!} = 1$)

$$\begin{aligned} \Delta^{d+1}v_t &= \Delta^d(1-L)v_t = \sum_{k=0}^{t-1}\frac{(-d)_k}{k!}L^k v_t - \sum_{k=0}^{t-1}\frac{(-d)_k}{k!}L^{k+1}v_t \\ &= \sum_{k=0}^{t-1}\frac{(-d)_k}{k!} - \sum_{k=0}^{t-2}\frac{(-d)_k}{k!} = \frac{(-d)_{t-1}}{(t-1)!}. \end{aligned}$$

It follows that

$$\begin{aligned} w_{\Delta^{d+1}v}(\lambda_s) &= \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n}\frac{(-d)_{t-1}}{(t-1)!}e^{it\lambda_s} = \frac{e^{i\lambda_s}}{\sqrt{2\pi n}}\sum_{k=0}^{n-1}\frac{(-d)_k}{k!}e^{ik\lambda_s} \\ &= \frac{e^{i\lambda_s}}{\sqrt{2\pi n}}\left[D_n\left(e^{i\lambda_s};d\right) - \frac{(-d)_n}{n!}\right], \end{aligned}$$

where $D_n(e^{i\lambda_s};d)$ is defined in Lemma 8.1. Therefore, we obtain

$$w_{\Delta^d v}(\lambda_s) = \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\frac{1}{\sqrt{2\pi n}}\left[D_n\left(e^{i\lambda_s};d\right) - \frac{(-d)_n}{n!} - \frac{(1-d)_{n-1}}{(n-1)!}\right].$$

The stated result follows from Lemma 8.2 and the fact that

$$\begin{aligned} \frac{(-d)_n}{n!} &= \frac{\Gamma(n-d)}{\Gamma(n+1)\Gamma(-d)} = \frac{1}{\Gamma(-d)}n^{-d-1}\left(1+O\left(n^{-1}\right)\right) = O\left(n^{-d-1}\right), \\ \frac{(1-d)_{n-1}}{(n-1)!} &= \frac{\Gamma(n-d)}{\Gamma(n)\Gamma(1-d)} = \frac{1}{\Gamma(1-d)}n^{-d}\left(1+O\left(n^{-1}\right)\right). \end{aligned}$$

For part (b), first we find a uniform bound for $d \in [-1+\varepsilon, 1]$. From Lemma 8.5, we have

$$-\log(1-L)\Delta^d v_t = J_n(L)\Delta^d v_t = J_n(e^{i\lambda_s})\Delta^d v_t + \widetilde{J}_{n\lambda_s}(e^{-i\lambda_s}L)(e^{-i\lambda_s}L-1)\Delta^d v_t.$$

Taking the dft leaves us with

$$-w_{\log(1-L)\Delta^d v}(\lambda_s) = J_n(e^{i\lambda_s})w_{\Delta^d v}(\lambda_s) - (2\pi n)^{-1/2}\widetilde{J}_{n\lambda_s}(e^{-i\lambda_s}L)\Delta^d v_n.$$

Since $\Delta^d v_{n-p} = (1-d)_{n-p-1}/(n-p-1)! = O((n-p)^{-d})$ for $d \in [-C, C]$, from Lemma 8.6 (b), the second term on the right is

$$\begin{aligned} & -(2\pi n)^{-1/2} \sum_{p=0}^{n-1} \widetilde{j}_{\lambda_s p} e^{-ip\lambda_s} \Delta^d v_{n-p} \\ = \;& O\left(n^{-1/2} \textstyle\sum_{p=0}^{n-1} |p|_+^{-1} n s^{-1} (n-p)^{-d}\right) = O\left(s^{-1} n^{1/2} \textstyle\sum_{p=0}^{n-1} |p|_+^{-1} (n-p)^{-d}\right). \end{aligned}$$

Observe that for any number m such that $\frac{1}{m} + \frac{m}{n} \to 0$ and $d \in [-C, 1]$

$$\begin{aligned} \textstyle\sum_{p=0}^{n-1} |p|_+^{-1} (n-p)^{-d} &= \textstyle\sum_0^{n-m} |p|_+^{-1} (n-p)^{-d} + \textstyle\sum_{n-m+1}^{n-1} |p|_+^{-1} (n-p)^{-d} \\ &\leq \begin{cases} n^{-d} \sum_0^{n-m} |p|_+^{-1} + (n-m)^{-1} \sum_1^m p^{-d} & d \in [-C, 0] \\ m^{-d} \sum_0^{n-m} |p|_+^{-1} + (n-m)^{-1} \sum_1^m p^{-d} & d \in [0, 1] \end{cases} \\ &= \begin{cases} O\left(n^{-d} \log n + n^{-1} m^{1-d}\right) & d \in [-C, 0] \\ O\left(m^{-d} \log n + n^{-1} m^{1-d} \log m\right) & d \in [0, 1], \end{cases} \\ &= O(n^{-d} (\log n)^2) \quad \text{uniformly in } d, \end{aligned} \tag{20}$$

by setting $m = n/\log n$. It follows that

$$-(2\pi n)^{-1/2} \widetilde{J}_{n\lambda_s}(e^{-i\lambda_s} L) \Delta^d v_n = O\left(s^{-1} n^{1/2-d} (\log n)^2\right). \tag{21}$$

For $d \in [1, 2]$, first Lemma 8.1 (b) gives

$$-w_{\Delta^d v}(\lambda_s) = -(1 - e^{i\lambda_s}) w_{\Delta^{d-1} v}(\lambda_s) + \frac{e^{i\lambda_s}}{\sqrt{2\pi n}} \Delta^{d-1} v_n. \tag{22}$$

Differentiating it with respect to d, we find

$$-w_{\log(1-L)\Delta^d v}(\lambda_s) = -(1 - e^{i\lambda_s}) w_{\log(1-L)\Delta^{d-1} v}(\lambda_s) + \frac{e^{i\lambda_s}}{\sqrt{2\pi n}} \log(1-L) \Delta^{d-1} v_n.$$

The first term on the right is

$$J_n(e^{i\lambda_s})(1 - e^{i\lambda_s}) w_{\Delta^{d-1} v}(\lambda_s) + O(n^{1/2-d} (\log n)^2).$$

From (20) and the fact that $d - 1 \leq 1$, the second term on the right is bounded by

$$n^{-1/2} \sum_{p=1}^{n-1} p^{-1} \Delta^{d-1} v_{n-p} = O\left(n^{-1/2} \textstyle\sum_{p=1}^{n-1} p^{-1} (n-p)^{1-d}\right) = O\left(n^{1/2-d} (\log n)^2\right),$$

giving the stated result.

For part (c), first, for $d \in [-C, 1]$, we have from Lemma 8.5

$$\begin{aligned} J_n(L)^2 &= \left[J_n(e^{i\lambda}) + \widetilde{J}_{n\lambda}(e^{-i\lambda} L)(e^{-i\lambda} L - 1)\right]^2 \\ &= J_n(e^{i\lambda})^2 + J_n(e^{i\lambda}) \widetilde{J}_{n\lambda}(e^{-i\lambda} L)(e^{-i\lambda} L - 1) + J_n(L) \widetilde{J}_{n\lambda}(e^{-i\lambda} L)(e^{-i\lambda} L - 1). \end{aligned}$$

It follows that

$$\begin{aligned} w_{(\log(1-L))^2 \Delta^d v}(\lambda_s) &= w_{J_n(L)^2 \Delta^d v}(\lambda_s) \\ &= J_n(e^{i\lambda})^2 w_{\Delta^d v}(\lambda_s) - J_n(e^{i\lambda}) (2\pi n)^{-1/2} \widetilde{J}_{n\lambda_s}(e^{-i\lambda_s} L) \Delta^d v_n \\ &\quad - J_n(L) (2\pi n)^{-1/2} \widetilde{J}_{n\lambda_s}(e^{-i\lambda_s} L) \Delta^d v_n. \end{aligned}$$

The second term is $O(s^{-1}n^{1/2-d}(\log n)^3)$ by (21), and from Lemma 8.6 (b), the third term is

$$\begin{aligned}
& -(2\pi n)^{-1/2}\sum_{q=1}^{n} q^{-1}\sum_{p=0}^{n-1}\widetilde{j}_{\lambda_s p}e^{-ip\lambda_s}\Delta^d v_{n-p-q} \\
= & -(2\pi n)^{-1/2}\sum_{p=0}^{n-1}\sum_{q=1}^{n-p-1}\widetilde{j}_{\lambda_s p}e^{-ip\lambda_s}q^{-1}\Delta^d v_{n-p-q} \\
= & \ O\left(s^{-1}n^{1/2}\sum_{p=0}^{n-1}|p|_+^{-1}\sum_{q=1}^{n-p-1}q^{-1}(n-p-q)^{-d}\right) \\
= & \ O\left(s^{-1}n^{1/2}\sum_{p=0}^{n-1}|p|_+^{-1}(n-p)^{-d}(\log n)^2\right) = O\left(s^{-1}n^{1/2-d}(\log n)^4\right). \quad (23)
\end{aligned}$$

For $d \in [1,2]$, taking the second derivative of $(-(22))$ with respect to d gives

$$w_{(\log(1-L))^2\Delta^d v}(\lambda_s) = (1-e^{i\lambda_s})w_{(\log(1-L))^2\Delta^{d-1}v}(\lambda_s) - \frac{e^{i\lambda_s}}{\sqrt{2\pi n}}(\log(1-L))^2\Delta^{d-1}v_n.$$

The first term on the right is

$$J_n(e^{i\lambda})^2(1-e^{i\lambda_s})w_{\Delta^{d-1}v}(\lambda_s) + O(n^{1/2-d}(\log n)^4),$$

and the second term on the right is

$$O\left(n^{-1/2}\sum_{p=1}^{n-1}p^{-1}\sum_{q=1}^{n-p-1}q^{-1}(n-p-q)^{1-d}\right) = O\left(n^{1/2-d}(\log n)^4\right),$$

giving the stated result. ■

9 Appendix B: Proofs

9.1 Proof of Theorem 4.2

We follow the approach developed by Shimotsu and Phillips (2002a), hereafter simply SP. Define $G(d) = G_0\frac{1}{m}\sum_1^m \lambda_j^{2d-2d_0}$ and $S(d) = R(d) - R(d_0)$. Define $\Theta_1^a = \{d : d_0 - \frac{1}{2} + \Delta \le d \le d_0 + \frac{1}{2}\}$, $\Theta_1^b = \{d : d_0 + \frac{1}{2} \le d \le \Delta_2\}$ and $\Theta_2 = \{d : \Delta_1 \le d \le d_0 - \frac{1}{2} + \Delta\}$, where Δ is defined in (13) and Θ_1^b and Θ_2 are possibly empty. In view of the arguments in Robinson (1995b), $\widehat{d} \to_p d_0$ if

$$\sup_{\Theta_1^a}|T(d)| \to_p 0, \ \Pr\left(\inf_{\Theta_1^b} S(d) \le 0\right) \to 0, \ \text{and} \ \Pr\left(\inf_{\Theta_2} S(d) \le 0\right) \to 0, \quad (24)$$

as $n \to \infty$, where

$$\begin{aligned}
T(d) = & \log\frac{\widehat{G}(d_0)}{G_0} - \log\frac{\widehat{G}(d)}{G(d)} - \log\left(\frac{1}{m}\sum_{j=1}^m j^{2d-2d_0}\Big/\frac{m^{2d-2d_0}}{2(d-d_0)+1}\right) \\
& +(2d-2d_0)\left[\frac{1}{m}\sum_1^m \log j - (\log m - 1)\right].
\end{aligned}$$

Robinson (1995b) shows that the fourth term on the right-hand side is $O(\log m/m)$ uniformly in $d \in \Theta_1^a \cup \Theta_1^b$ and

$$\sup_{\Theta_1^a \cup \Theta_1^b} \left| \frac{2(d-d_0)+1}{m} \sum_1^m \left(\frac{j}{m}\right)^{2d-2d_0} - 1 \right| = O\left(\frac{1}{m^{2\Delta}}\right). \tag{25}$$

A little algebra shows that

$$\begin{aligned} & \frac{\widehat{G}(d) - G(d)}{G(d)} \\ = \quad & \frac{\left[2(d-d_0)+1\right] m^{-1} \sum_1^m (j/m)^{2d-2d_0} \left[\lambda_j^{2d_0-2d} I_{\Delta^d(x-\widehat{x}_0)}(\lambda_j) - G_0\right]}{\left[2(d-d_0)+1\right] G_0 m^{-1} \sum_1^m (j/m)^{2d-2d_0}} = \frac{A(d)}{B(d)}. \end{aligned}$$

Therefore, by the fact that $\Pr(|\log Y| \geq \varepsilon) \leq 2\Pr(|Y-1| \geq \varepsilon/2)$ for any nonnegative random variable Y and $\varepsilon \leq 1$, $\sup_{\Theta_1^a} |T(d)| \to_p 0$ if

$$\sup_{\Theta_1^a} |A(d)/B(d)| \to_p 0. \tag{26}$$

Define $Y_t = (1-L)^d (X_t - X_0)$. Then

$$Y_t = (1-L)^{d-d_0}(1-L)^{d_0}(X_t - X_0) = (1-L)^{\theta} u_t I\{t \geq 1\},$$

where $\theta \equiv d - d_0$. Hereafter, we use the notation $Y_t \sim I(\alpha)$ when Y_t is generated by (1) with parameter α. So $Y_t \sim I(-\theta)$.

Because we have observations only for $t \geq 1$, it follows that

$$X_t - \widehat{X}_0 = X_t - X_0 + X_0 - \widehat{X}_0 = X_t - X_0 + \eta \cdot v_t,$$

where

$$\eta = X_0 - \widehat{X}_0, \quad v_t = I\{t \geq 1\}.$$

Taking the dft of both sides, the error in estimating X_0 is translated in the frequency domain as

$$w_{\Delta^d(x-\widehat{x}_0)}(\lambda_j) = w_{\Delta^d(x-x_0)}(\lambda_j) + \eta w_{\Delta^d v}(\lambda_j) = w_y(\lambda_j) + \eta w_{\Delta^d v}(\lambda_j). \tag{27}$$

Hence, with $g = 2(d-d_0)+1$, $A(d)$ can be written as

$$A(d) = \frac{g}{m} \sum_1^m \left(\frac{j}{m}\right)^{2\theta} \left[\lambda_j^{-2\theta} \left|w_y(\lambda_j) + \eta w_{\Delta^d v}(\lambda_j)\right|^2 - G_0\right].$$

Hereafter let I_{yj} denote $I_y(\lambda_j)$, w_{uj} denote $w_u(\lambda_j)$, and similarly for other dft's and periodograms. SP shows that

$$\sup_{\Theta_1^a} \left| \frac{g}{m} \sum_1^m \left(\frac{j}{m}\right)^{2\theta} \left[\lambda_j^{-2\theta} I_{yj} - G_0\right] \right| \to_p 0.$$

From the fact that $||A|^2 - |B|^2| \leq |A+B||A-B|$ and the Cauchy-Schwartz inequality we have

$$\begin{aligned} & E \sup_{\theta} \left| \lambda_j^{-2\theta} I_{yj} - \lambda_j^{-2\theta} \left| w_{yj} + \eta w_{\Delta^d vj} \right|^2 \right| \\ \leq \quad & \left(E \sup_{\theta} \left| 2\lambda_j^{-\theta} w_{yj} - \lambda_j^{-\theta} \eta w_{\Delta^d vj} \right|^2 \right)^{1/2} \left(E \sup_{\theta} \left| \lambda_j^{-\theta} \eta w_{\Delta^d vj} \right|^2 \right)^{1/2}. \end{aligned} \tag{28}$$

Before proceeding, we state a useful result as a lemma:

Lemma A *Uniformly in d,*

$$\lambda_j^{-\theta} w_{\Delta^d v j} = \begin{cases} C_1(\theta)\, n^{1/2-d_0} j^{d_0-1} \left[1 + O(j^{-d}) + O(\lambda_j)\right], & d \in [0, C], \\ C_2(\theta)\, n^{1/2-d_0} j^{-\theta-1} \left[1 + O(j^{d}) + O(\lambda_j)\right], & d \in [-1+\varepsilon, 0]. \end{cases}$$

where $C_1(\theta)$ and $C_2(\theta)$ do not depend on j and are bounded and bounded away from zero uniformly in θ.

Proof From Lemma 8.8 (a), $\lambda_j^{-\theta} w_{\Delta^d v j}$ is equal to

$$\begin{aligned} & \lambda_j^{-\theta} \frac{e^{i\lambda_j}}{1-e^{i\lambda_j}} \frac{1}{\sqrt{2\pi n}} \left[\left(1-e^{i\lambda_j}\right)^d - \frac{n^{-d}}{\Gamma(1-d)} + O\left(n^{-d} j^{-1}\right)\right] \\ = & \begin{cases} \lambda_j^{-\theta}(1-e^{i\lambda_j})^{d-1} (2\pi n)^{-1/2} \left[1 + O(j^{-d}) + O(\lambda_j)\right], & d \geq 0, \\ -\lambda_j^{-\theta}(1-e^{i\lambda_j})^{-1} (2\pi n)^{-1/2} n^{-d} \Gamma(1-d)^{-1} \left[1 + O(j^{d}) + O(\lambda_j)\right], & d \leq 0, \end{cases} \end{aligned}$$

and the stated result follows from the fact that $\theta = d - d_0$ and Lemma 8.2. ■

Because $d_0 \geq -1/2$, Lemma 8.3 gives

$$\eta = \frac{1}{n}\sum_{t=1}^{n} (X_t - X_0) = n^{-1}(1-L)^{-d_0-1} u_t \{t \geq 1\}, \quad E\eta^2 = O\left(n^{2d_0-1}\right). \tag{29}$$

It follows from Lemma A and (29) that

$$\eta \lambda_j^{-\theta} w_{\Delta^d v j} = \begin{cases} n^{1/2-d_0}\eta \cdot O\left(j^{d_0-1}\right), & d \geq 0, \\ n^{1/2-d_0}\eta \cdot O\left(j^{-\theta-1}\right), & d \leq 0, \end{cases} \quad E|n^{1/2-d_0}\eta|^2 < \infty, \tag{30}$$

where the $O(\cdot)$ terms are uniform in d. Equation (42) of SP gives

$$E \sup_\theta \left|\lambda_j^{-\theta} w_{yj}\right|^2 = O\left((\log n)^6\right), \quad j = 1, \ldots, m. \tag{31}$$

Therefore, for $\theta \in \Theta_1^a = \{-1/2 + \Delta \leq \theta \leq 1/2\}$, we obtain

$$(28) = O\left(j^{d_0-1}(\log n)^3 + j^{-1/2}(\log n)^3\right), \quad j = 1, \ldots, m,$$

and it follows that $\sup_{\Theta_1^a} |A(d)|$ is

$$o_p(1) + O_p\left(\frac{1}{m}\sum_1^m \left(\frac{j}{m}\right)^{2\Delta-1} \left(j^{d_0-1}(\log n)^3 + j^{-\frac{1}{2}}(\log n)^3\right)\right) = o_p(1).$$

From (25), $\inf_{\Theta_1^a} B(d) \geq 0.5 G_0$ for large n, and (26) follows.

Next we take care of $\Theta_1^b = \{\frac{1}{2} \leq \theta \leq \Delta_2 - d_0\}$. From the arguments in SP pp. 20-21 (equation (44)), $\Pr\left(\inf_{\Theta_1^b} S(d) \leq 0\right)$ tends to 0 if

$$\frac{1}{m}\sum_1^m I_{\Delta^{d_0}(x-\widehat{x}_0)j} - G_0 \to_p 0, \tag{32}$$

and for $\delta \in (0, 0.01)$,

$$\Pr\left(\inf_{\Theta_1^b}\left[\frac{1}{m}\sum_1^m a_j\left(\lambda_j^{-2\theta} I_{\Delta^d(x-\widehat{x}_0)j} - G_0\right)\right] \leq -\delta G_0\right) \to 0, \tag{33}$$

as $n \to \infty$, where

$$a_j = \begin{cases} (j/p)^M, & 1 \le j \le p, \quad M \ge \max\{2\Delta_2 - 2\Delta_1, 2\}, \\ j/p, & p < j \le m, \quad p = \exp\left(m^{-1}\sum_1^m \log j\right) \sim m/e \text{ as } m \to \infty. \end{cases} \tag{34}$$

(32) follows from the arguments for Θ_1^a with $\theta = 0$. We proceed to evaluate the limit of $\inf_\theta m^{-1}\sum_1^m a_j(\lambda_j^{-2\theta} I_{(x-\widehat{x}_0)j} - G_0)$ for subsets of Θ_1^b. For $\Theta_1^{b1} = \{\theta : \frac{1}{2} \le \theta \le \frac{3}{2}\}$, from the equation between (46) and (47) in SP p. 22 and (27), we have

$$\begin{aligned} \lambda_j^{-\theta} w_{(x-\widehat{x}_0)j} &= D_{nj}(\theta) w_{uj} - \lambda_j^{-\theta}(1-e^{i\lambda_j})(2\pi n)^{-1/2}\widetilde{U}_{\lambda_j n}(\theta - 1) \\ &\quad + \lambda_j^{-\theta}(2\pi n)^{-1/2} e^{i\lambda_j} Z_n + \eta\lambda_j^{-\theta} w_{\Delta^d v j}, \end{aligned} \tag{35}$$

where $D_{nj}(\theta) = e^{-\frac{\pi}{2}\theta i} + O(\lambda_j) + O(j^{-1/2})$. In view of the order of magnitude of $\eta\lambda_j^{-\theta} w_{\Delta^d v j}$ and $\lambda_j^{-\theta}(1-e^{i\lambda_j})(2\pi n)^{-1/2}\widetilde{U}_{\lambda_j n}(\theta-1)$ shown in (30) and Lemma 8.4, the arguments in SP pp. 22-25 go through with an extra $o_p(1)$ term $\eta\lambda_j^{-\theta} w_{\Delta^d v j}$, giving

$$\Pr\left(\inf_{\Theta_1^{b1}} m^{-1}\sum_1^m a_j(\lambda_j^{-2\theta} I_{(x-\widehat{x}_0)j} - G_0)\right) \le -\delta G_0 \to 0,$$

as $n \to \infty$. Since $\Delta_1 > -1/2$ and $\Delta_2 < 1$, $\theta < 3/2$ and $\Theta_1^b \subseteq \Theta_1^{b1}$, and we finish the analysis for Θ_1^b.

Now we consider $\Theta_2 = \{\theta : \Delta_1 - d_0 \le \theta \le -\frac{1}{2} + \Delta\}$. Note that (32) still holds. Then, from the arguments in SP pp. 26-27, $\Pr(\inf_{\Theta_2} S(d) \le 0)$ tends to 0 if

$$\Pr\left(\inf_{\Theta_2}\left[\frac{1}{m}\sum_1^m a_j\left(\lambda_j^{-2\theta} I_{\Delta^d(x-\widehat{x}_0)j} - G_0\right)\right] \le -\delta G_0\right) \to 0, \tag{36}$$

for $\delta \in (0, 0.01)$, where

$$a_j = \begin{cases} (j/p)^{2\Delta-1}, & 1 \le j \le p, \\ (j/p)^{2\Delta_1 - 2d_0}, & p < j \le m. \end{cases} \tag{37}$$

SP shows that

$$\begin{aligned} \textstyle\sum_1^m a_j &= O(m), \quad \textstyle\sum_1^m a_j^2 = O(m^{2-4\Delta}), \\ m^{-1}\textstyle\sum_1^m a_j j^\alpha &= O(m^\alpha \log m + m^{-2\Delta}\log m) \text{ uniformly in } \alpha \in [-C, C]. \end{aligned} \tag{38}$$

First, for $\Theta_2^a = \{\theta : -\frac{1}{2} \le \theta \le -\frac{1}{2} + \Delta\}$, from Lemma 8.4 and (27), we have

$$\begin{aligned} \lambda_j^{-\theta} w_{\Delta^d(x-\widehat{x}_0)j} &= \lambda_j^{-\theta} w_{yj} + \lambda_j^{-\theta}\eta w_{\Delta^d v j} \\ &= \lambda_j^{-\theta} D_n(e^{i\lambda_j};\theta) w_{uj} - (2\pi n)^{-1/2}\lambda_j^{-\theta}\widetilde{U}_{\lambda_j n}(\theta) + \lambda_j^{-\theta}\eta w_{\Delta^d v j}. \end{aligned} \tag{39}$$

Since $\sup_\theta |m^{-1}\sum_1^m a_j[\lambda_j^{-2\theta}|D_n(e^{i\lambda_j};\theta)|^2 I_{uj} - G_0]| \to_p 0$ in view of Lemma 8.2, apart from $o_p(1)$ and almost surely nonnegative terms, $m^{-1}\sum_1^m a_j[\lambda_j^{-2\theta} I_{\Delta^d(x-\widehat{x}_0)j} - G_0]$ consists of

$$m^{-1}\textstyle\sum_1^m a_j \lambda_j^{-2\theta}\eta^2 I_{\Delta^d v j} \tag{40}$$

$$+m^{-1}\textstyle\sum_1^m a_j \lambda_j^{-\theta} D_n(e^{i\lambda_j};\theta)^* w_{uj}^* (2\pi n)^{-1/2}\lambda_j^{-\theta}\widetilde{U}_{\lambda_j n}(\theta) \tag{41}$$

$$+m^{-1}\textstyle\sum_1^m a_j \lambda_j^{-\theta} D_n(e^{i\lambda_j};\theta)^* w_{uj}^* \lambda_j^{-\theta}\eta w_{\Delta^d v j}, \tag{42}$$

and complex conjugates of (41) and (42). From equation (68) in SP and (38) for any $D_{nj}(\theta)$ such that

$$D_{nj}(\theta) = e^{-\frac{\pi}{2}\theta i} + O(\lambda_j) + O(j^{-1/2}),$$

we have uniformly in $-1 \leq \alpha < C$

$$m^{-1}\sum_1^m a_j D_{nj}(\theta) w_{uj} O(j^\alpha) = O_p\left(m^\alpha \log m + m^{-2\Delta}\log m\right), \tag{43}$$

and

$$\begin{aligned} & m^{-1}\sum_1^m a_j D_{nj}(\theta) w_{uj} j^\alpha [1 + O(\lambda_j)] \\ = \;& O_p\left(m^{\alpha-1/2}\log m + m^{-2\Delta}\log m\right) + O_p\left(n^{-1}m^{\alpha+1}\right). \end{aligned} \tag{44}$$

(41) and (42) are $o_p(1)$ by Lemma 8.4, (30) and (43). (40) is almost surely nonnegative, and (36) follows.

We move to $\Theta_2^b = \{\theta : -3/2 \leq \theta \leq -1/2\}$. From SP p. 27 equation (62) and thereafter, $\lambda_j^{-\theta} w_{(x-\widehat{x}_0)j}$ is equal to

$$\begin{array}{l} D_{nj}(\theta) w_{uj} - \lambda_j^{-\theta}(1 - e^{i\lambda_j})^{-1}(2\pi n)^{-1/2}\widetilde{U}_{\lambda_j n}(\theta+1) \\ -\lambda_j^{-\theta}(1 - e^{i\lambda_j})^{-1}(2\pi n)^{-1/2} e^{i\lambda_j} Y_n + \eta\lambda_j^{-\theta} w_{\Delta^d vj}, \end{array} \tag{45}$$

where $D_{nj}(\theta) = e^{-\frac{\pi}{2}\theta i} + O(\lambda_j) + O(j^{-1/2})$. First consider the case $\theta \in [-1, -1/2]$. Apart from $o_p(1)$ and almost surely nonnegative terms, $m^{-1}\sum_1^m a_j[\lambda_j^{-2\theta} I_{(x-\widehat{x}_0)j} - G_0]$ consists of

$$-m^{-1}\sum_1^m a_j D_{nj}(\theta)^* w_{uj}^* \lambda_j^{-\theta}(1 - e^{i\lambda_j})^{-1}(2\pi n)^{-1/2}\widetilde{U}_{\lambda_j n}(\theta+1) \tag{46}$$

$$-m^{-1}\sum_1^m a_j D_{nj}(\theta)^* w_{uj}^* \lambda_j^{-\theta}(1 - e^{i\lambda_j})^{-1}(2\pi n)^{-1/2} e^{i\lambda_j} Y_n \tag{47}$$

$$-m^{-1}\sum_1^m a_j D_{nj}(\theta)^* w_{uj}^* \eta\lambda_j^{-\theta} w_{\Delta^d vj}, \tag{48}$$

and their complex conjugates. SP shows that (46) and (47) are $o_p(1)$ (see SP (66) and (67)). Using an argument similar to the one in the proof of Lemma A, we can derive from Lemma 8.8 that

$$\eta\lambda_j^{-\theta} w_{\Delta^d vj} = n^{1/2-d_0}\eta\left[C_1(\theta) j^{d_0-1} + C_2(\theta) j^{-\theta-1} + O(j^{-\theta-2})\right](1 + O(\lambda_j)), \tag{49}$$

where $C_1(\theta)$ and $C_2(\theta)$ are bounded and bounded away from zero uniformly in θ. In view of (43), (44), and (49) and that $n^{1/2-d_0}\eta = O_p(1)$, (48) is $O_p(m^{-2\Delta}\log m) + O_p(n^{-1}m)$, giving (36).

For $\theta \in [-3/2, -1]$, first observe that d is strictly negative because

$$\theta = d - d_0 \leq -1 \Rightarrow d \leq d_0 - 1 < 0.$$

From (45), apart from almost surely nonnegative terms and obviously $o_p(1)$ terms, $m^{-1}\sum_1^m a_j[\lambda_j^{-2\theta} I_{(x-\widehat{x}_0)j} - G_0]$ consists of

$$m^{-1}\sum_1^m a_j\left|\lambda_j^{-\theta}(1 - e^{i\lambda_j})^{-1}(2\pi n)^{-1/2} e^{i\lambda_j} Y_n - \eta\lambda_j^{-\theta} w_{\Delta^d vj}\right|^2 \tag{50}$$

$$-m^{-1}\sum_1^m a_j D_{nj}(\theta)^* w_{uj}^* \lambda_j^{-\theta}(1 - e^{i\lambda_j})^{-1}(2\pi n)^{-1/2}\widetilde{U}_{\lambda_j n}(\theta+1) \tag{51}$$

$$-m^{-1}\sum_1^m a_j D_{nj}(\theta)^* w_{uj}^* \lambda_j^{-\theta}\left[(1 - e^{i\lambda_j})^{-1}(2\pi n)^{-1/2} e^{i\lambda_j} Y_n - \eta w_{\Delta^d vj}\right] \tag{52}$$

$$\begin{array}{l} -m^{-1}\sum_1^m a_j \lambda_j^{-\theta}(1 - e^{-i\lambda_j})^{-1}(2\pi n)^{-1/2}\widetilde{U}_{\lambda_j n}(\theta+1)^* \\ \times\lambda_j^{-\theta}\left[(1 - e^{i\lambda_j})^{-1}(2\pi n)^{-1/2} e^{i\lambda_j} Y_n - \eta w_{\Delta^d vj}\right], \end{array} \tag{53}$$

and complex conjugates of (51)-(53). SP shows (51) is $o_p(1)$. Lemma A gives

$$\eta\lambda_j^{-\theta} w_{\Delta^d vj} = C_1(\theta)\, n^{1/2-d_0}\eta j^{-\theta-1}\left[1+O(j^{d_0-1})+O(\lambda_j)\right],$$

and a similar argument gives

$$\lambda_j^{-\theta}(1-e^{i\lambda_j})^{-1}(2\pi n)^{-1/2} e^{i\lambda_j} Y_n = C_2(\theta)\, n^{\theta+1/2} Y_n j^{-\theta-1}\left[1+O(\lambda_j)\right],$$

where $C_1(\theta)$ and $C_2(\theta)$ do not depend on j and are bounded and bounded away from zero uniformly in θ. It follows that (50) is almost surely larger than

$$m^{-1}\textstyle\sum_{\log m}^{m} a_j j^{-2\theta-2}\left|C_2(\theta)\, n^{\theta+1/2} Y_n - C_1(\theta)\, n^{1/2-d_0}\eta\right|^2\left[1+o(1)\right], \tag{54}$$

where the $o(1)$ term follows from the fact that $d_0-1<0$ and $j\geq \log m\to\infty$. From equation (69) in SP p. 28, there exists $\kappa>0$ such that

$$\min\{m^{-1}\textstyle\sum_{\log m}^{m/4} a_j j^{\alpha}, m^{-1}\sum_{3m/4}^{m} a_j j^{\alpha}\}\geq \kappa m^{\alpha}, \tag{55}$$

uniformly in $\alpha\in[-C,C]$. In equation (69) in SP, the summation begins at 1 instead of $\log m$, but this does not change the result. Therefore, there exists $\kappa>0$ such that

$$(54)\geq_{a.s.} \kappa m^{-2\theta-2}\left|C_2(\theta)\, n^{\theta+1/2} Y_n - C_1(\theta)\, n^{1/2-d_0}\eta\right|^2. \tag{56}$$

Using the same argument as in SP pp. 28-29, we can show that (52) and (53) are dominated by (56), and (36) follows. Since $|d-d_0|<3/2$, $\Theta_2\subseteq\Theta_2^a\cup\Theta_2^b$, thus we finish the proof for Θ_2 and also complete the proof. ■

9.2 Proof of Theorem 4.3

From Lemma 8.8 (a), the fact that $d>0$ and $d_0\geq 1/2$ we have

$$\lambda_j^{-\theta} w_{\Delta^d vj} = O\left(n^{1/2-d_0} j^{d_0-1}\right) = O\left((j/n)^{d_0-1/2} j^{-1/2}\right) = O\left(j^{-1/2}\right).$$

Hence we have uniform in d

$$\eta\lambda_j^{-\theta} w_{\Delta^d vj} = \eta\cdot O\left(j^{-1/2}\right),\quad E|\eta|^2 = E|u_1|^2<\infty. \tag{57}$$

Therefore, the arguments in the proof of Theorem 3.3 of SP go through without change by handling $\eta\lambda_j^{-\theta} w_{\Delta^d vj}$ similarly as $\lambda_j^{-\theta}(2\pi n)^{-1/2}\widetilde{U}_{\lambda_j n}(\theta)$, $\lambda_j^{-\theta}(1-e^{i\lambda_j})^{-1}(2\pi n)^{-1/2}\widetilde{U}_{\lambda_j n}(\theta+1)$, and $\lambda_j^{-\theta}(1-e^{i\lambda_j})(2\pi n)^{-1/2}\widetilde{U}_{\lambda_j n}(\theta-1)$. ■

9.3 Proof of Theorem 4.6

Without loss of generality, assume $\varepsilon<\Delta$. We need to treat the cases for different values of d_0 separately.

9.3.1 (a) $d_0 \geq 1$

When $d \in [c, \Delta_2]$, $\widehat{X}_0(d) = X_1$, therefore the required result follows from the proof of Theorem 4.3.

When $d \in [\Delta_1, c)$, $\widehat{X}_0(d) = \overline{X}$ and $\eta = X_0 - \overline{X}$. Since $\theta = d - d_0 \leq -\frac{1}{2} + \Delta$, we need to consider only Θ_2. Because $m^{-1}\sum_1^m [I_{\Delta^{d_0}(x-\widehat{x}_0(d_0))j} - G_0] \rightarrow_p 0$, $\Pr\left(\inf_{\Theta_2} S(d) \leq 0\right)$ tends to 0 if

$$\Pr\left(\inf_{\Theta_2}\left[\frac{1}{m}\sum_1^m a_j\left(\lambda_j^{-2\theta} I_{\Delta^d(x-\widehat{x}_0(d))j} - G_0\right)\right] \leq -\delta G_0\right) \rightarrow 0, \quad \text{as } n \rightarrow \infty \tag{58}$$

for $\delta \in (0, 0.01)$, where a_j is defined in (37). We proceed to establish (58) for subsets of Θ_2. For $\Theta_2^a = \{-1/2 \leq \theta \leq -1/2 + \Delta\}$, the decomposition (40)-(42) still holds with (41) being $o_p(1)$. When $d \leq -\varepsilon$, (40) is a.s. nonnegative and from Lemma A and (43), we have

$$(42) = n^{1/2-d_0}\eta O_p\left(m^{-2\Delta}\log m\right) = o_p(1),$$

giving (58). When $d \geq \varepsilon$, from Lemma A, (43), and (44), we obtain

$$\begin{aligned}
(42) &= n^{1/2-d_0}\eta\left[O_p\left(m^{d_0-3/2}\log m + m^{-2\Delta}\log m\right) + O_p\left(m^{d_0-1-\varepsilon}\log m\right)\right] \\
&\quad + n^{1/2-d_0}\eta O_p\left(n^{-1}m^{d_0}\right) \\
&= m^{d_0-1}n^{1/2-d_0}\eta\left[O_p\left(m^{-\varepsilon}\log m\right) + O_p\left(n^{-1}m\right)\right].
\end{aligned}$$

From Lemma A and (55), there exists $\kappa > 0$ such that

$$(40) \geq \kappa |m^{d_0-1}n^{1/2-d_0}\eta|^2 \quad a.s.$$

Therefore, (42) is dominated by (40), and (58) follows.

For $\Theta_2^b = \{-3/2 \leq \theta \leq -1/2\}$, we can use the same approach as the one for $\theta \in [-3/2, -1]$ in the proof of Theorem 4.2, i.e., the equation (50) and thereafter. When $d \leq -\varepsilon$, because d is strictly negative, the argument from (50) to (56) holds without change. When $d \geq \varepsilon$, (51) is still $o_p(1)$, and from Lemma A (50) is a.s. larger than

$$m^{-1}\textstyle\sum_{\log m}^m a_j\left|C_2(\theta)\, n^{\theta+1/2}Y_n j^{-\theta-1} - C_1(\theta)\, n^{1/2-d_0}\eta j^{d_0-1}\right|^2 [1 + o(1)].$$

Because $-\theta = d_0 - d$ and $d \geq \varepsilon$, the following holds for any numbers A and B and large m :

$$\begin{aligned}
m^{-1}\textstyle\sum_{\log m}^m a_j|Aj^{-\theta-1} + Bj^{d_0-1}|^2 &\geq m^{-1}\textstyle\sum_{\log m}^m a_j j^{-2\theta-2}|A + Bj^d|^2 \\
&\geq 0.5m^{-1}\textstyle\sum_{\log m}^m a_j j^{-2\theta-2}|A|^2, \\
m^{-1}\textstyle\sum_{\log m}^m a_j|Aj^{-\theta-1} + Bj^{d_0-1}|^2 &\geq m^{-1}\textstyle\sum_{\log m}^m a_j j^{2d_0-2}|Aj^{-d} + B|^2 \\
&\geq 0.5m^{-1}\textstyle\sum_{\log m}^m a_j j^{2d_0-2}|B|^2.
\end{aligned}$$

Therefore, from (55), for large m, (50) is a.s. larger than

$$\kappa|C_2(\theta)\, n^{\theta+1/2}Y_n m^{-\theta-1}|^2 + \kappa|C_1(\theta)\, n^{1/2-d_0}\eta m^{d_0-1}|^2, \tag{59}$$

for some $\kappa > 0$. Consequently, (52) and (53) are dominated by (59), and (58) follows. For $\Theta_2^c = \{-5/2 \leq \theta \leq -3/2\}$ and smaller values of θ, the expression of $\lambda_j^{-\theta}w_{yj}$ will contain ΔY_n, $\Delta^2 Y_n, \ldots$ as shown in SP p. 29, but the same reasoning gives (58).

9.3.2 (b) $d_0 \in [\frac{1}{2}, 1)$

The required result follows because $\widehat{d}$ is consistent both under $\widehat{X}_0 = X_1$ and $\widehat{X}_0 = \overline{X}$.

9.3.3 (c) $d_0 < \frac{1}{2}$

Divide Θ_1^a into the two, $\Theta_1^{a1} = \{-\frac{1}{2} + \Delta \le \theta \le \Delta\}$ and $\Theta_1^{a2} = \{\Delta \le \theta \le \frac{1}{2}\}$. $\widehat{d} \to_p d_0$ if

$$\sup_{\Theta_1^{a1}} |T(d)| \to_p 0 \text{ and } \Pr\left(\inf_{\Theta_1^{a2} \cup \Theta_1^b \cup \Theta_2} S(d) \le 0\right) \to 0 \quad \text{as } n \to \infty. \tag{60}$$

First we collect the results for Θ_1^{a2}. From the arguments in SP pp. 20-21, we have

$$S(d) = \log \widehat{D}(d) - \log \widehat{D}(d_0); \quad \widehat{D}(d) = m^{-1} \sum_1^m \left(\frac{j}{p}\right)^{2\theta} \lambda_j^{-2\theta} I_{\Delta^d(x - \widehat{x}_0(d))j}.$$

Therefore, $\Pr(\inf_{\Theta_1^{a2}} S(d) \le 0) \to 0$ if

$$\Pr\left(\inf_{\Theta_1^{a2}} \widehat{D}(d) - G_0 \le \delta G_0\right) \to 0, \quad \text{for } \delta \in (0, 0.01) \tag{61}$$

as $n \to \infty$. Observe that

$$\widehat{D}(d) - G_0 = m^{-1} \textstyle\sum_1^m (j/p)^{2\theta} [\lambda_j^{-2\theta} I_{\Delta^d(x-\widehat{x}_0(d))j} - G_0] + G_0[m^{-1} \sum_1^m (j/p)^{2\theta} - 1]. \tag{62}$$

Now

$$m^{-1} \textstyle\sum_1^m (j/p)^{2\theta} = (m/p)^{2\theta} m^{-1} \sum_1^m (j/m)^{2\theta} \sim e^{2\theta} / (2\theta + 1), \quad \text{as } m \to \infty,$$

by (25). Since $\log[e^{2\theta}/(2\theta + 1)] = 2\theta - \log(2\theta + 1) > 0$ for $\theta > 0$, we have for large m and small δ

$$\inf_{\Theta_1^{a2}} m^{-1} \textstyle\sum_1^m (j/p)^{2\theta} > 1 + 10\delta.$$

Hence the second term on the right of (62) is larger than $10\delta G_0$. The first term on the right of (62) is equal to

$$(m/p)^{2\theta} m^{-1} \textstyle\sum_1^m (j/m)^{2\theta} [\lambda_j^{-2\theta} I_{\Delta^d(x-\widehat{x}_0(d))j} - G_0]. \tag{63}$$

It is easy to check that $(j/m)^{2\theta}$ with $\theta \in \Theta_1^{a2}$ satisfies

$$\begin{aligned} \sup_\theta \textstyle\sum_1^m (j/m)^{2\theta} &= O(m), \quad \sup_\theta \textstyle\sum_1^m (j/m)^{4\theta} = O(m), \\ m^{-1} \textstyle\sum_1^m (j/m)^{2\theta} j^\alpha &\le m^{-1} \textstyle\sum_1^m (j/m)^{2\Delta} j^\alpha = O(m^\alpha \log m + m^{-1-2\Delta} \log m), \end{aligned}$$

and there exists $\kappa > 0$ such that

$$\min\{\inf_\theta m^{-1} \textstyle\sum_{\log m}^{m/4} (j/m)^{2\theta} j^\alpha, \inf_\theta m^{-1} \sum_{3m/4}^m (j/m)^{2\theta} j^\alpha\} \ge \kappa m^\alpha,$$

uniformly in $\alpha \in [-C, C]$.

For $d \in [\Delta_1, c)$, $\widehat{X}_0 = \overline{X}$. $\sup_{\Theta_1^{a1}} |T(d)| \to_p 0$ and $\Pr(\inf_{\Theta_1^b \cup \Theta_2} S(d) \le 0) \to 0$ follow from the proof of Theorem 4.2. For $\theta \in \Theta_1^{a2}$, observe that

$$(63) = [2\theta + 1]^{-1} (m/p)^{2\theta} A(d),$$

where $A(d)$ is defined at the beginning of the proof of Theorem 4.2. Since $\sup_{\Theta_1^a} |A(d)| \to_p 0$, it follows that $\sup_{\Theta_1^{a2}} |(63)| \to_p 0$ and $\Pr(\inf_{\Theta_1^{a2}} S(d) \le 0) \to 0$, giving (60).

For $d \in [c, \Delta_2]$, $\theta = d - d_0 \ge \Delta$ and $\theta \in \Theta_1^{a2} \cup \Theta_1^b$. First we take care of Θ_1^{a2}. From Lemma 8.4 and (27), we have

$$\lambda_j^{-\theta} w_{\Delta^d(x-\widehat{x}_0)j} = \lambda_j^{-\theta} D_n(e^{i\lambda_j}; \theta) w_{uj} - (2\pi n)^{-1/2} \lambda_j^{-\theta} \widetilde{U}_{\lambda_j n}(\theta) + \lambda_j^{-\theta} \eta w_{\Delta^d vj}.$$

From Lemma A and an argument similar to a previously stated one, all the possibly nonpositive terms in $m^{-1}\sum_1^m (j/m)^{2\theta}[\lambda_j^{-2\theta} I_{\Delta^d(x-\widehat{x}_0(d))j} - G_0]$ are dominated by $m^{-1}\sum_1^m (j/m)^{2\theta} |\lambda_j^{-\theta} \eta w_{\Delta^d vj}|^2$, and (61) follows. For $\Theta_1^b = \{\frac{1}{2} \le \theta < \Delta_2 - d_0\}$, as in the proof of Theorem 4.2, $\Pr(\inf_{\Theta_1^b} S(d) \le 0)$ tends to 0 if for $\delta \in (0, 0.01)$,

$$\Pr\left(\inf_{\Theta_1^b}\left[\frac{1}{m}\sum_1^m a_j\left(\lambda_j^{-2\theta} I_{\Delta^d(x-\widehat{x}_0(d))j} - G_0\right)\right] \le -\delta G_0\right) \to 0, \tag{64}$$

where a_j are defined in (34). From the results in SP pp. 21-23 (equations (i), (ii), (iv) and (v) and Lemma C), for any $D_{nj}(\theta) = e^{-\frac{\pi}{2}\theta i} + O(\lambda_j) + O(j^{-1/2})$ we have uniformly in $-C \le \alpha < C$

$$\begin{aligned} m^{-1}\textstyle\sum_1^m a_j D_{nj}(\theta) w_{uj} O(j^\alpha) &= O_p(m^\alpha \log m), \\ m^{-1}\textstyle\sum_1^m a_j D_{nj}(\theta) w_{uj} j^\alpha [1 + O(\lambda_j)] &= O_p(m^{\alpha-1/2}\log m) + O_p(n^{-1} m^{\alpha+1}). \end{aligned}$$

Furthermore, (55) still holds for a_j defined in (34), and from the fact that $d \notin (1-\varepsilon, 1+\varepsilon)$, we have for large m

$$\begin{aligned} & m^{-1}\textstyle\sum_{\log m}^m a_j \left|A j^{-\theta} + B j^{d_0-1}\right|^2 \\ = \; & m^{-1}\textstyle\sum_{\log m}^m a_j \left|A j^{d_0-d} + B j^{d_0-1}\right|^2 \\ \ge \; & 0.5 m^{-1}\textstyle\sum_{\log m}^m a_j j^{-2\theta} |A|^2 + 0.5 m^{-1}\sum_{\log m}^m a_j j^{2d_0-2} |B|^2 . \end{aligned}$$

For $\Theta_1^{b1} = \{\theta : \frac{1}{2} \le \theta \le \frac{3}{2}\}$, in view of the above results and a similar argument as before, all the possibly nonpositive terms in $m^{-1}\sum_1^m a_j[\lambda_j^{-2\theta} I_{\Delta^d(x-\widehat{x}_0(d))j} - G_0]$ are dominated by (c.f. (35))

$$m^{-1}\textstyle\sum_1^m a_j |\lambda_j^{-\theta}(2\pi n)^{-1/2} e^{i\lambda_j} Z_n + \lambda_j^{-\theta}\eta w_{\Delta^d vj}|^2, \quad Z_n = \sum_{t=1}^n Y_t,$$

and (64) follows. For smaller values of d, the expression of $\lambda_j^{-2\theta} I_{\Delta^d(x-\widehat{x}_0(d))j}$ will contain $\sum_{k=1}^n Z_k$, $\sum_{k=1}^n \sum_{t=1}^k Z_t, \ldots$, but the same line of reasoning establishes (64) and completes the proof. ■

9.4 Proof of Theorem 4.8

Theorem 4.6 holds under Assumptions 1′-3c′ and implies that with probability approaching 1, as $n \to \infty$, $\widehat{d}$ satisfies

$$0 = R'(\widehat{d}) = R'(d_0) + R''(d^*)(\widehat{d} - d_0), \tag{65}$$

where $|d^* - d_0| \leq |\widehat{d} - d_0|$. Let $M = \{d : (\log n)^4 |d - d_0| < \delta\}$ for a fixed δ. From the arguments in SP pp. 30-32, $R''(d^*) = 4 + o_p(1)$ holds if

$$\sup_M \left| \widetilde{G}_0(d) - \frac{1}{m} \sum_1^m j^{2\theta} \lambda_j^{-2\theta} I_{\Delta^d(x-\widehat{x}_0(d))j} \right| = o_p\left((\log n)^{-2}\right), \quad (66)$$

$$\sup_M \left| \widetilde{G}_1(d) - \frac{1}{m} \sum_1^m j^{2\theta} \lambda_j^{-2\theta} 2\,\mathrm{Re}\left[w_{\log(1-L)\Delta^d(x-\widehat{x}_0(d))j} w^*_{\Delta^d(x-\widehat{x}_0(d))j} \right] \right| = o_p\left((\log n)^{-1}\right), \quad (67)$$

$$\sup_M \left| \widetilde{G}_2(d) - \frac{1}{m} \sum_1^m j^{2\theta} \lambda_j^{-2\theta} W_x(L, d, j) \right| = o_p(1), \quad (68)$$

where

$$\begin{aligned}
\widetilde{G}_0(d) &= \frac{1}{m} \sum_1^m j^{2\theta} \lambda_j^{-2\theta} I_{yj}, \quad \widetilde{G}_1(d) = \frac{1}{m} \sum_1^m j^{2\theta} \lambda_j^{-2\theta} 2\,\mathrm{Re}\left[w_{\log(1-L)yj} w^*_{yj} \right], \\
\widetilde{G}_2(d) &= \frac{1}{m} \sum_1^m j^{2\theta} \lambda_j^{-2\theta} \left\{ 2\,\mathrm{Re}\left[w_{(\log(1-L))^2 yj} w^*_{yj} \right] + 2 I_{\log(1-L)yj} \right\}, \\
W_x(L, d, j) &= 2\,\mathrm{Re}\left[w_{(\log(1-L))^2 \Delta^d(x-\widehat{x}_0(d))j} w^*_{\Delta^d(x-\widehat{x}_0(d))j} \right] + 2 I_{\log(1-L)\Delta^d(x-\widehat{x}_0(d))j}.
\end{aligned}$$

Observe that from SP we have

$$\sup_M |j^{-2\theta} - 1| = O((\log n)^{-3}). \quad (69)$$

For (66), from (28) - (31) and (57), we obtain

$$\begin{aligned}
& E \sup_M \left| \lambda_j^{-2\theta} I_{yj} - \lambda_j^{-2\theta} I_{\Delta^d(x-\widehat{x}_0(d))j} \right| \\
= & \left\{ \begin{array}{ll} O\left(j^{d_0-1} (\log n)^3 + j^{-1/2} (\log n)^3 \right), & d_0 < c, \\ O\left(j^{-1/2} (\log n)^3 \right), & d_0 \geq c, \end{array} \right\} = O\left(j^{-1/4} (\log n)^3 \right) \quad (70)
\end{aligned}$$

(66) follows from (69) and (70).

In view of (69), (67) holds if

$$\begin{aligned}
& E \sup_M \left| \lambda_j^{-2\theta} w_{\log(1-L)yj} w^*_{yj} - \lambda_j^{-2\theta} w_{\log(1-L)\Delta^d(x-\widehat{x}_0(d))j} w^*_{\Delta^d(x-\widehat{x}_0(d))j} \right| \\
= & O\left(j^{-1/4} (\log n)^5 \right). \quad (71)
\end{aligned}$$

Observe that

$$\begin{aligned}
& \lambda_j^{-2\theta} w_{\log(1-L)\Delta^d(x-\widehat{x}_0(d))j} w^*_{\Delta^d(x-\widehat{x}_0(d))j} \\
= & \left[\lambda_j^{-\theta} w_{\log(1-L)yj} + \lambda_j^{-\theta} \eta w_{\log(1-L)\Delta^d vj} \right] \left[\lambda_j^{-\theta} w^*_{yj} + \lambda_j^{-\theta} \eta w^*_{\Delta^d vj} \right].
\end{aligned}$$

Therefore, (71) holds if

$$\begin{array}{l}
E \sup_M \left| \lambda_j^{-\theta} w_{\log(1-L)yj} \lambda_j^{-\theta} \eta w^*_{\Delta^d vj} \right|, \quad E \sup_M \left| \lambda_j^{-\theta} \eta w_{\log(1-L)\Delta^d vj} \lambda_j^{-\theta} w^*_{yj} \right|, \\
E \sup_M \left| \lambda_j^{-\theta} \eta w_{\log(1-L)\Delta^d vj} \lambda_j^{-\theta} \eta w^*_{\Delta^d vj} \right|,
\end{array} \quad (72)$$

are all $O(j^{-1/4}(\log n)^5)$. From (30), (57), and Lemma 8.8 (b), we have

$$\lambda_j^{-\theta} \eta w_{\Delta^d vj} = \xi_1 O\left(j^{-1/2+\Delta} \right), \quad \lambda_j^{-\theta} \eta w_{\log(1-L)\Delta^d vj} = \xi_2 O\left(j^{-1/2+\Delta} (\log n)^2 \right), \quad (73)$$

uniformly in $d \in M$, where $E|\xi_1|^2$, $E|\xi_2|^2 < \infty$. From (31), we have

$$E \sup_M \left| \lambda_j^{-\theta} w_{yj} \right|^2 = O\left((\log n)^6 \right), \quad j = 1, \dots m. \tag{74}$$

(72)$= O(j^{-1/2+\Delta} (\log n)^5)$ follows from (73), (74), and Lemma 8.7 (b).

In view of (69), (68) holds if

$$\begin{array}{l} E \sup_M \left| \lambda_j^{-2\theta} w_{(\log(1-L))^2 yj} w_{yj}^* - \lambda_j^{-2\theta} w_{(\log(1-L))^2 \Delta^d (x - \widehat{x}_0(d))j} w^*_{\Delta^d (x-\widehat{x}_0(d))j} \right|, \\ E \sup_M \left| \lambda_j^{-2\theta} I_{\log(1-L) yj} - \lambda_j^{-2\theta} I_{\log(1-L)\Delta^d (x-\widehat{x}_0(d))j} \right|, \end{array} \tag{75}$$

are $O(j^{-1/4} (\log n)^7)$. From (30), (57), and Lemma 8.8 (c), we have

$$E \sup_M \left| \lambda_j^{-\theta} \eta w_{(\log(1-L))^2 \Delta^d vj} \right|^2 = O\left(j^{-1+2\Delta} (\log n)^8 \right). \tag{76}$$

(75)$= O(j^{-1/4} (\log n)^7)$ follows from (73), (74), (76), and Lemma 8.7 (b) and (c). It follows that $R''(d^*) = 4 + o_p(1)$.

Now we find the limit distribution of

$$m^{1/2} R'(d_0) = m^{1/2} \left\{ \left[\widehat{G}_1(d_0) / \widehat{G}(d_0) \right] - 2 \frac{1}{m} \sum_1^m \log \lambda_j \right\},$$

where

$$\begin{aligned} \widehat{G}_1(d_0) &= \frac{1}{m} \sum_1^m 2 \operatorname{Re} \left[w_{\log(1-L)\Delta^{d_0}(x-\widehat{x}_0(d_0))j} w^*_{\Delta^{d_0}(x-\widehat{x}_0(d_0))j} \right], \\ \widehat{G}(d_0) &= \frac{1}{m} \sum_1^m I_{\Delta^{d_0}(x-\widehat{x}_0(d_0))j}. \end{aligned}$$

SP shows

$$m^{1/2} \left[\frac{G_1^*(d_0)}{G^*(d_0)} - 2 \frac{1}{m} \sum_1^m \log \lambda_j \right] \to_d N(0, 4),$$

where

$$G_1^*(d_0) = \frac{1}{m} \sum_1^m 2 \operatorname{Re} \left[w_{\log(1-L)uj} w_{uj}^* \right], \quad G^*(d_0) = \frac{1}{m} \sum_1^m I_{uj}.$$

Therefore, $m^{1/2} R'(d_0) \to_d N(0, 4)$ if

$$\begin{aligned} & m^{1/2} \left[\left[\widehat{G}_1(d_0) / \widehat{G}(d_0) \right] - \left[G_1^*(d_0) / G^*(d_0) \right] \right] \\ = \; & m^{1/2} \left[\frac{\widehat{G}_1(d_0) - G_1^*(d_0)}{\widehat{G}(d_0)} \right] + m^{1/2} G_1^*(d_0) \left[\frac{1}{\widehat{G}(d_0)} - \frac{1}{G^*(d_0)} \right], \end{aligned} \tag{77}$$

is $o_p(1)$. For the first term in (77), observe that $m^{1/2}[\widehat{G}_1(d_0) - G_1^*(d_0)]$ is equal to

$$\frac{1}{\sqrt{m}} \sum_1^m 2\eta \operatorname{Re} \left[w_{\log(1-L)uj} w^*_{\Delta^{d_0} vj} \right] + \frac{1}{\sqrt{m}} \sum_1^m 2\eta \operatorname{Re} \left[w_{\log(1-L)\Delta^{d_0} vj} w^*_{uj} \right] \tag{78}$$

$$+ \frac{1}{\sqrt{m}} \sum_1^m 2\eta^2 \operatorname{Re} \left[w_{\log(1-L)\Delta^{d_0} vj} w^*_{\Delta^{d_0} vj} \right]. \tag{79}$$

In view of Lemma 8.7 (a), (73), and that

$$w_{us} = C(e^{i\lambda_s})w_{\varepsilon s} + r_{ns}; \quad E\left|r_{ns}\right|^2 = O(n^{-1}), \quad s = 1, \ldots m, \tag{80}$$

(Hannan 1970, p. 248) the first term in (78) is

$$\frac{1}{\sqrt{m}}\sum_1^m 2\eta \operatorname{Re}\left[J_n\left(e^{i\lambda_j}\right)C(e^{i\lambda_j})w_{\varepsilon j}w^*_{\Delta^{d_0}vj}\right] \tag{81}$$
$$+O_p\left(\frac{1}{\sqrt{m}}\sum_1^m n^{-1/2}j^{-1/2+\Delta}\log n\right) + O_p\left(\frac{1}{\sqrt{m}}\sum_1^m j^{-1+\Delta}\left(\log n\right)^4\right).$$

The reminder terms are $O_p(n^{-1/2}m^{\Delta}\log n + m^{\Delta-1/2}(\log n)^4)$, and (81) is

$$O_p\left(\left(\frac{1}{m}\sum_1^m j^{-1+2\Delta}\left(\log n\right)^2\right)^{1/2}\right) = O_p\left(m^{-1/2+\Delta}\log n\right). \tag{82}$$

Similarly, we can show the second term in (78) is $O_p(n^{-1/2}m^{\Delta}(\log n)^2) + O_p(m^{-1/2+\Delta}(\log n)^2)$. Finally, from (73), we have

$$(79) = O_p\left(\frac{1}{\sqrt{m}}\sum_1^m j^{-1+2\Delta}(\log n)^2\right) = O_p\left(m^{2\Delta-1/2}(\log n)^2\right),$$

and $m^{1/2}[\widehat{G}_1(d_0) - G_1^*(d_0)] = o_p(1)$ follows. $\widehat{G}(d_0) = G_0 + o_p(1)$ from the proof of Theorems 4.2 and 4.3, and hence the first term in (77) is $o_p(1)$.

For the second term in (77), SP p. 33 shows that $m^{1/2}G_1^*(d_0) = 2m^{-1/2}\sum_1^m \log(\lambda_j)I_{uj} + o_p(1) = O_p(m^{1/2}\log n)$. Now

$$m^{1/2}\left[\widehat{G}\left(d_0\right) - G^*\left(d_0\right)\right] = \frac{1}{\sqrt{m}}\sum_1^m 2\eta \operatorname{Re}\left[w_{uj}w^*_{\Delta^{d_0}vj}\right] + \frac{1}{\sqrt{m}}\sum_1^m \eta^2 I_{\Delta^{d_0}vj}.$$

The second term on the right is $O_p(m^{-1/2+2\Delta})$, and, in view of (73) and (80), the first term on the right is

$$\frac{1}{\sqrt{m}}\sum_1^m 2\eta \operatorname{Re}\left[C(e^{i\lambda_j})w_{\varepsilon j}w^*_{\Delta^{d_0}vj}\right] + O_p\left(\frac{1}{\sqrt{m}}\sum_1^m j^{-1/2+\Delta}n^{-1/2}\right)$$
$$= \quad O_p\left(\left(\frac{1}{m}\sum_1^m j^{-1+2\Delta}\right)^{1/2}\right) + O_p\left(n^{-1/2}m^{\Delta}\right) = O_p\left(m^{-1/2+\Delta} + n^{-1/2}m^{\Delta}\right).$$

Since both $\widehat{G}\left(d_0\right)$ and $\widehat{G}\left(d_0\right)$ are $G_0 + o_p(1)$, it follows that

$$m^{1/2}\left\{\left[1/\widehat{G}\left(d_0\right)\right] - \left[1/G^*\left(d_0\right)\right]\right\} = o_p\left((\log n)^{-1}\right),$$

and the second term in (77) is $o_p(1)$. Therefore, we obtain $m^{1/2}R'(d_0) \to_d N(0,4)$ to complete the proof. ■

9.5 Proof of Theorem 5.1

From (15), taking the dft of $\Delta^d(\widehat{X}_t - \varphi(d))$ gives

$$\begin{aligned} w_{\Delta^d(\widehat{x}-\varphi(d))j} &= w_{\Delta^d x^0 j} - \psi_n(d) w_{\Delta^d v j} - \zeta_n(d_0) w_{\Delta^d t j}; \\ \psi_n(d) &= [\overline{X^0} - \zeta_n(d_0)\overline{t}] I\{d < c\} + [X_1^0 - \zeta_n(d_0)] I\{d \geq c\}. \end{aligned}$$

First we derive the order of

$$\zeta_n(d_0) = \left[\sum_{t=1}^n (t-\overline{t})^2\right]^{-1} \left(\sum_{t=1}^n t X_t^0 - \overline{t}\sum_{t=1}^n X_t^0\right).$$

By summation by parts,

$$\begin{aligned} \sum_{t=1}^n t X_t^0 &= \sum_{k=1}^{n-1} (k-(k+1)) \sum_{t=1}^k X_t^0 + n\sum_{t=1}^n X_t^0 \\ &= -\sum_{k=1}^{n-1}\sum_{t=1}^k X_t^0 + n\sum_{t=1}^n X_t^0 \\ &= -(1-L)^{-d_0-2} u_{n-1} I\{t \geq 1\} + n(1-L)^{-d_0-1} u_n I\{t \geq 1\}. \end{aligned}$$

Since $d_0 > -1/2$ and $\overline{t} \sim n/2$, we have

$$\sum_{t=1}^n t X_t^0 = O_P\left(n^{d_0+3/2}\right), \quad \sum_{t=1}^n (t-\overline{t}) X_t^0 = O_P\left(n^{d_0+3/2}\right).$$

Because $\sum_{t=1}^n (t-\overline{t})^2 = \sum_{t=1}^n t^2 - n(\overline{t})^2 \sim n^3/3 - n^3/4 = n^3/12$, it follows that

$$\zeta_n(d_0) = O_p(n^{d_0-3/2}), \quad \zeta_n(d_0)\overline{t} = O_p(n^{d_0-1/2}). \tag{83}$$

Therefore, of the terms in $\psi_n(d)$, $\overline{X^0} - \zeta_n(d_0)\overline{t}$ has the same order as $\overline{X^0}$. For the term involving $X_1^0 - \zeta_n(d_0)$, since $\lambda_j^{-\theta} w_{\Delta^d v j} = O(j^{-1/2})$ for $d > 0$ and $d_0 \geq 1/2$, we obtain for $d, d_0 \geq 1/2$

$$[X_1^0 - \zeta_n(d_0)]\lambda_j^{-\theta} w_{\Delta^d v j} = \eta \cdot O(j^{-1/2} + j^{d_0-2}), \quad E|\eta|^2 < \infty.$$

Hence, $\psi_n(d)\lambda_j^{-\theta} w_{\Delta^d v j}$ can be handled in the same manner as $\eta\lambda_j^{-\theta} w_{\Delta^d v j}$ in the proof of Theorems 4.2 - 4.6.

Now we evaluate the order of $w_{\Delta^d t j}$. Observe that

$$t = (1-L)^{-1} v_t, \quad w_{\Delta^d t j} = w_{\Delta^{d-1} v j}. \tag{84}$$

For $d \geq \varepsilon$, the expression for $w_{\Delta^d t j}$ follows from (84) and Lemma 8.8 (a). For $d \in (-1+\varepsilon, -\varepsilon)$, first observe that

$$w_{\Delta^d t j} = w_{\Delta^{d-1} v j} = \frac{1}{1-e^{i\lambda_j}}\left[w_{\Delta^d v j} - \frac{e^{i\lambda_j}}{\sqrt{2\pi n}}\Delta^{d-1} v_n\right].$$

From Lemma 8.8 and its proof, we have

$$w_{\Delta^d v j} = O\left(n^{1/2-d} j^{-1}\right), \quad \Delta^{d-1} v_n = \frac{n^{1-d}}{\Gamma(2-d)}\left(1 + O\left(n^{-1}\right)\right),$$

which gives

$$w_{\Delta^d tj} = \frac{1}{1-e^{i\lambda_j}}\left[-\frac{e^{i\lambda_j}}{\sqrt{2\pi n}}\frac{n^{1-d}}{\Gamma(2-d)} + O\left(n^{1/2-d}j^{-1}\right)\right].$$

It follows that

$$w_{\Delta^d tj} = \begin{cases} C_1(d)\, n^{3/2-d} j^{d-2}\left(1 + O\left(j^{1-d}\right) + O(\lambda_j)\right), & d \in [1+\varepsilon, C], \\ C_2(d)\, n^{3/2-d} j^{-1}\left(1 + O\left(j^{d-1}\right) + O(\lambda_j)\right), & d \in [\varepsilon, 1-\varepsilon], \\ C_3(d)\, n^{3/2-d} j^{-1}\left(1 + O\left(j^{-1}\right) + O(\lambda_j)\right), & d \in [-1+\varepsilon, -\varepsilon], \end{cases}$$

where $C_k(d)$ are generic functions that do not depend on j and are bounded and bounded away uniformly in d. Combining the above results, we obtain

$$\begin{aligned} &\zeta_n(d_0)\lambda_j^{-\theta} w_{\Delta^d tj} \\ = &\begin{cases} n^{3/2-d_0}\zeta_n(d_0)\cdot C_1(d)\, j^{d_0-2}\left(1 + O\left(j^{-\varepsilon}\right)\right), & d \in [1+\varepsilon, C], \\ n^{3/2-d_0}\zeta_n(d_0)\cdot C_2(d)\, j^{-\theta-1}\left(1 + O\left(j^{-\varepsilon}\right)\right), & d \in [-1+\varepsilon, , 1-\varepsilon]\backslash(-\varepsilon, \varepsilon). \end{cases} \end{aligned}$$

Therefore, we can apply the arguments in the proof of Theorems 4.2 - 4.6 to show consistency.

For asymptotic normality, a similar result is obtained for $w_{\log(1-L)\Delta^d tj}$ and $w_{(\log(1-L))^2\Delta^d tj}$. From the inspection of the proof of Theorem 4.8, the sufficient condition for asymptotic normality is (see (73) and (76)) that uniformly in $d \in M$ we have for $\alpha < -1/4$

$$\begin{aligned} \psi_n(d)\lambda_j^{-\theta} w_{\Delta^d vj}, \quad \zeta_n(d_0)\lambda_j^{-\theta} w_{\Delta^d tj} &= \xi O\left(j^\alpha\right), \\ \psi_n(d)\lambda_j^{-\theta} w_{\log(1-L)\Delta^d vj}, \quad \zeta_n(d_0)\lambda_j^{-\theta} w_{\log(1-L)\Delta^d tj} &= \xi O\left(j^\alpha(\log n)^2\right), \\ \psi_n(d)\lambda_j^{-\theta} w_{(\log(1-L))^2\Delta^d vj}, \quad \zeta_n(d_0)\lambda_j^{-\theta} w_{(\log(1-L))^2\Delta^d tj} &= \xi O\left(j^\alpha(\log n)^8\right), \end{aligned}$$

where ξ is a generic random variable with $E|\xi|^2 < \infty$. If $d_0 < 7/4$, the above conditions are satisfied, and the required result follows. ■

References

[1] Backus, D. K. and S. E. Zin, 1991, Long-memory inflation uncertainty: Evidence from the term structure of interest rates. *Journal of Money, Credit, and Banking* **25**: 681-700.

[2] Crato, N. and P. Rothman, 1994, Fractional integration analysis of long-run behavior for US macroeconomic time series. *Economics Letters* **45**: 287-291.

[3] Diebold, F. X. and G. D. Rudebusch, 1991, Is consumption too smooth? Long memory and the Deaton paradox. *Review of Economics and Statistics* **73**: 1-9.

[4] Hannan, E. J., 1970, *Multiple Time Series* (Wiley, New York).

[5] Hassler, U. and J. Wolters, 1995, Long memory in inflation rates: International evidence. *Journal of Business & Economic Statistics* **13**: 37-45.

[6] Haubrich, J. G., 1993, Consumption and fractional differencing: Old and new anomalies. *Review of Economics and Statistics* **75**: 767-72.

[7] Henry, M. and P. Zaffaroni, 2002, The long range dependence paradigm for macroeconomics and finance. Forthcoming in: P. Doukhan, G. Oppenheim and M. Taqqu, eds., *Long-range dependence: Theory and applications* (Birkhauser).

[8] Hurvich, C. M. and W. W. Chen, 2000, An efficient taper for potentially overdifferenced long-memory time series. *Journal of Time Series Analysis* **21**: 155-180.

[9] Kim, C. S. and P. C. B. Phillips, 1999, Log periodogram regression: the nonstationary case. Mimeographed, Cowles Foundation, Yale University.

[10] Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin, 1992, Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* **54**: 159-178.

[11] Maynard, A. and P. C. B. Phillips, 2001, Rethinking an old empirical puzzle: Econometric evidence on the forward discount anomaly. *Journal of Applied Econometrics* **16**: 671-708.

[12] Michelacci, C. and P. Zaffaroni, 2000, (Fractional) beta convergence. *Journal of Monetary Economics* **45**: 129-53.

[13] Nelson, C. R. and C. I. Plosser, 1982, Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics* **10**: 139-62.

[14] Phillips, P. C. B., 1999, Unit root log periodogram regression. Cowles Foundation Discussion Paper #1244, Yale University.

[15] Phillips, P. C. B. and K. Shimotsu, 2001, Local Whittle estimation in nonstationary and unit root cases. Cowles Foundation Discussion Paper #1266, Yale University.

[16] Robinson, P. M., 1995a, Log-periodogram regression of time series with long range dependence. *Annals of Statistics* **23**: 1048-1072.

[17] Robinson, P. M., 1995b, Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* **23**: 1630-1661.

[18] Shimotsu, K. and P. C. B. Phillips, 2002a, Exact local Whittle estimation of fractional integration. Cowles Foundation Discussion Paper #1367, Yale University.

[19] Shimotsu, K. and P. C. B. Phillips, 2002b, Local Whittle estimation of fractional integration and some of its variants. Mimeograph. Yale University.

[20] Schotman, P. and H. K. van Dijk, 1991, On Bayesian routes to unit roots. *Journal of Applied Econometrics* **6**: 387-401.

[21] Velasco, C., 1999, Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* **20**: 87-127.