Econometric Theory, **26**, 2010, 501–540.
doi:10.1017/S0266466609100075

EXACT LOCAL WHITTLE ESTIMATION OF FRACTIONAL INTEGRATION WITH UNKNOWN MEAN AND TIME TREND

KATSUMI SHIMOTSU
Hitotsubashi University

Recently, Shimotsu and Phillips (2005, *Annals of Statistics* 33, 1890–1933) 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 it has the same $N(0, \frac{1}{4})$ asymptotic distribution for all values of d, if the optimization covers an interval of width less than $9/2$ and the mean of the process is known. With the intent to provide a semiparametric estimator suitable for economic data, we extend the ELW estimator so that it accommodates an unknown mean and a polynomial time trend. We show that the two-step ELW estimator, which is based on a modified ELW objective function using a tapered local Whittle estimator in the first stage, has an $N(0, \frac{1}{4})$ asymptotic distribution for $d \in (-\frac{1}{2}, 2)$ (or $d \in (-\frac{1}{2}, \frac{7}{4})$ when the data have a polynomial trend). Our simulation study illustrates that the two-step ELW estimator inherits the desirable properties of the ELW estimator.

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 for modeling 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, 2003). Because of their flexibility in modeling temporal dependence, $I(d)$ processes can also help to reconcile implications from economic models with observed data. Indeed, their use has provided solutions for many empirical "puzzles" in economics and finance, e.g., consumption (Diebold and Rudebusch, 1991; Haubrich, 1993), term structure

The author thanks the co-editor and three anonymous referees for helpful and constructive comments. 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 2002 to August 2003. This research was supported by ESRC under grant R000223629.
Address correspondence to Katsumi Shimotsu, Department of Economics, Hitotsubashi University, 2–1, Naka, Kunitachi, Tokyo, 186–8601, Japan; e-mail: shimotsu@econ.hit-u.ac.jp.

 0266-4666/10 $15.00

(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 its misspecification. Two common statistical procedures in this class are log periodogram regression and local Whittle estimation (Robinson, 1995a, 1995b). Although these estimators are consistent for $d \in \left(\frac{1}{2}, 1\right)$ and asymptotically normally distributed for $d \in \left(\frac{1}{2}, \frac{3}{4}\right)$, they are also known to exhibit nonstandard behavior when $d > \frac{3}{4}$. For instance, they have a nonnormal limit distribution for $d \in \left[\frac{3}{4}, 1\right]$, and they converge to unity in probability and are inconsistent for $d > 1$ (Kim and Phillips, 2006; Phillips, 2007; Phillips and Shimotsu, 2004). To avoid inconsistency and an unreliable basis for inference when d may be larger than 3/4, a simple and commonly used procedure is to estimate d by taking first differences of the data, estimating $d-1$, and adding one to the estimate $\widehat{d-1}$. However, if the data are trend stationary, i.e., $I(d)$ with $d \in \left[0, \frac{1}{2}\right)$ around a linear time trend, taking a first difference of a time series reduces it to $I(d)$ with $d \in \left[-1, -\frac{1}{2}\right)$. 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, 2006).

Data tapering has been suggested (Velasco, 1999 [hereafter Vel]; Hurvich and Chen, 2000 [hereafter HC]) as a solution to extend the range of consistent estimation of d. Tapered estimators are invariant to a linear (and possibly higher order) time trend and asymptotically normal for $d \in \left(-\frac{1}{2}, \frac{3}{2}\right)$ (and for larger values of d if higher order tapers are used), but they have a larger variance (1.5 times or more) than the untapered estimator. As a result, there is currently no general purpose efficient estimation procedure when the value of d may take on values in the nonstationary zone beyond $3/4$, except for Abadir, Distaso, and Giraitis (2007).

Many economists and econometricians took part in the debate on whether economic time series are trend stationary or difference stationary. This debate remains 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 $d \geq \frac{1}{2}$ or $d < \frac{1}{2}$, because $I(d)$ processes become nonstationary when $d \geq \frac{1}{2}$. Gil-Alaña and Robinson (1997) applied the Robinson (1994) Lagrange multiplier (LM) test to macroeconomic data to test the null hypothesis that $d = d_0$ for various values of d_0, including $d = \frac{1}{2}$, and found that the results depend on how the short-run dynamics of the data are specified. Therefore, it is of great interest to investigate this issue using the semiparametric approach, which is agnostic about short-run dynamics. However, neither using the raw data or the differenced data or combining the two can answer whether $d \geq \frac{1}{2}$, because these procedures must assume either $d < \frac{3}{4}$ or $d > \frac{1}{2}$ prior to estimation.

Recently Shimotsu and Phillips (2005) (hereafter SP) 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 and has the same $N\left(0, \frac{1}{4}\right)$ limit distribution for all values of d if the optimization covers an interval of width less than 9/2 and the mean (initial value) of the process is known. As such, it 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 polynomial time trend. First, we examine the effect of an unknown mean (initial value) on ELW estimation. It is shown that (a) if an unknown mean is replaced by the sample average, then the ELW estimator is consistent for $d \in \left(-\frac{1}{2}, 1\right)$ and asymptotically normal for $d \in \left(-\frac{1}{2}, \frac{3}{4}\right)$, but simulations suggest that the estimator is inconsistent for $d > 1$; and (b) if an unknown mean is replaced by the first observation, then the ELW estimator is consistent for $d > 0$ and asymptotically normal for $d \in (0, 2)$, but consistency and asymptotic normality for $d \in \left(0, \frac{1}{2}\right)$ require a strong assumption on the number of periodogram ordinates used in estimation, and simulations suggest that the estimator is inconsistent for $d \leq 0$. An unknown mean needs to be estimated carefully in the ELW estimation.

In light of the preceding undesirable effect of unknown mean on the ELW estimation, we extend the ELW estimator so that it accommodates an unknown mean and a polynomial time trend. We modify the ELW objective function to estimate the mean by combining two estimators: the sample average and the first observation, depending on the value of d. The presence of a linear and/or quadratic time trend is dealt with by first detrending the data. We show that the two-step ELW estimator, which is based on the modified ELW objective function and uses a tapered estimator in the first stage, has the same $N\left(0, \frac{1}{4}\right)$ limit distribution for $d \in \left(-\frac{1}{2}, 2\right)$ ($d \in \left(-\frac{1}{2}, \frac{7}{4}\right)$ when the data are detrended). The finite-sample performance of the two-step ELW estimator inherits the desirable properties of the ELW estimator, apart from a small increase in bias and variance when the data are detrended. We also investigate the properties of the estimator that minimizes the modified objective function. The resulting feasible ELW estimator is shown to be consistent for $d > -\frac{1}{2}$, provided we exclude arbitrary small intervals around 0 and 1.

Abadir, Distaso, and Giraitis (2007) propose a fully extended local Whittle (FELW) estimator that uses a fully extended discrete Fourier transform. They show that the FELW estimator is consistent and has an $N\left(0, \frac{1}{4}\right)$ limit distribution for $d \in \left(-\frac{3}{2}, \infty\right)$. We view their estimator as being complementary to the one proposed in this paper for a variety of reasons. The FELW estimator has an advantage over the two-step ELW estimator in that it covers a wider range of d and it does not require estimating the mean.[1] However, the FELW estimator excludes the values of $d = \frac{1}{2}, \frac{3}{2}, \ldots,$ which results in holes in confidence intervals at these points, whereas our two-step approach does not. Furthermore, the two estimators are based on different models of the $I(d)$ process. The FELW estimator is based on Type I processes, whereas the two-step ELW estimator is based on Type II processes. Type I and II processes have their relative advantages and disadvantages; e.g., Type I processes are stationary for $d < \frac{1}{2}$ and more amenable to standard statistical analysis, whereas Type II processes can use a single model for all d and do not impose model discontinuity at $d = \frac{1}{2}, \frac{3}{2}, \ldots$[2]

The remainder of the paper is organized as follows. Section 2 briefly reviews ELW estimation. Section 3 analyzes the effect of an unknown mean and compares two estimators of the mean. Section 4 demonstrates the asymptotic properties of the two-step ELW estimator. Section 5 reports some simulation results and gives an empirical application using the extended Nelson–Plosser data. Proofs and some technical results are collected in Appendixes A and B.

2. A MODEL OF FRACTIONAL INTEGRATION AND ELW ESTIMATION

First we briefly review the ELW estimation developed by SP as it serves as the basis for the following analysis. Consider the fractionally integrated process X_t generated by the model

$$\Delta^d X_t = (1-L)^d X_t = u_t \mathbf{1}\{t \geq 1\}, \qquad t = 0, \pm 1, \ldots, \tag{1}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function and u_t is assumed to be stationary with zero mean and spectral density $f_u(\lambda)$ satisfying $f_u(\lambda) \sim G$ for $\lambda \sim 0$. Inverting and expanding the binomial in (1) gives a representation of X_t in terms of $u_1, \ldots, u_n$, which is valid for all values of d:

$$X_t = \Delta^{-d} u_t \mathbf{1}\{t \geq 1\} = (1-L)^{-d} u_t \mathbf{1}\{t \geq 1\} = \sum_{k=0}^{t-1} \frac{(d)_k}{k!} u_{t-k}, \qquad t = 0, \pm 1, \ldots,$$

where $(d)_k = \Gamma(d+k)/\Gamma(d)$ and $\Gamma(\cdot)$ is the gamma function.

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

$$w_a(\lambda_j) = (2\pi n)^{-1/2} \sum_{t=1}^{n} a_t e^{it\lambda_j}, \qquad \lambda_j = \frac{2\pi j}{n}, \qquad j = 1, \ldots, n, \tag{2}$$

$$I_a(\lambda_j) = |w_a(\lambda_j)|^2.$$

SP propose to estimate (d, G) by minimizing the objective function

$$Q_m(G, d) = \frac{1}{m} \sum_{j=1}^{m} \left[\log\left(G \lambda_j^{-2d}\right) + \frac{1}{G} I_{\Delta^d x}(\lambda_j) \right]. \tag{3}$$

Concentrating $Q_m(G,d)$ with respect to G, SP define the ELW estimator as

$$\widetilde{d} = \underset{d \in [\Delta_1, \Delta_2]}{\arg\min}\; R(d), \qquad (4)$$

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

$$R(d) = \log \widehat{G}(d) - 2d \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j, \qquad \widehat{G}(d) = \frac{1}{m} \sum_{j=1}^{m} I_{\Delta^d x}(\lambda_j).$$

In what follows, we distinguish the true values of d and G by d_0 and G_0. The ELW estimator has been shown to be consistent and asymptotically normally distributed for any $d_0 \in (\Delta_1, \Delta_2)$ if $\Delta_2 - \Delta_1 \le \frac{9}{2}$ and under fairly mild assumptions on m and the stationary component u_t.

Assumption 1. $f_u(\lambda) \sim G_0 \in (0,\infty)$ as $\lambda \to 0+$.

Assumption 2. In a neighborhood $(0,\delta)$ of the origin, $f_u(\lambda)$ is differentiable and $(d/d\lambda) \log f_u(\lambda) = O(\lambda^{-1})$ as $\lambda \to 0+$.

Assumption 3. $u_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$ with $\sum_{j=0}^{\infty} c_j^2 < \infty$, where $\mathrm{E}(\varepsilon_t | F_{t-1}) = 0$, $\mathrm{E}(\varepsilon_t^2 | F_{t-1}) = 1$ a.s., $t = 0, \pm 1, \ldots$, in which F_t is the σ-field generated by $\varepsilon_s, s \le t$, and there exists a random variable ε such that $\mathrm{E}\varepsilon^2 < \infty$ and for all $\eta > 0$ and some $K > 0$, $\Pr(|\varepsilon_t| > \eta) \le K \Pr(|\varepsilon| > \eta)$.

Assumption 4. $m^{-1} + m(\log m)^{1/2} n^{-1} + m^{-\gamma} \log n \to 0$ for any $\gamma > 0$.

Assumption 5. $\Delta_2 - \Delta_1 \le \frac{9}{2}$.

See SP for a comparison of the preceding assumptions with those in Robinson (1995b).

LEMMA 1 (SP, Thm. 2.1). *Suppose X_t is generated by (1) with $d_0 \in [\Delta_1, \Delta_2]$ and Assumptions 1–5 hold. Then $\widetilde{d} \to_p d_0$ as $n \to \infty$.*

Assumption 1′. Assumption 1 holds and also for some $\beta \in (0,2]$, $f_u(\lambda) = G_0(1 + O(\lambda^{\beta}))$, as $\lambda \to 0+$.

Assumption 2′. In a neighborhood $(0,\delta)$ of the origin, $C(e^{i\lambda})$ is differentiable and $(d/d\lambda) C(e^{i\lambda}) = O(\lambda^{-1})$ as $\lambda \to 0+$.

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

Assumption 4′. As $n \to \infty$, $m^{-1} + m^{1+2\beta} (\log m)^2 n^{-2\beta} + m^{-\gamma} \log n \to 0$ for any $\gamma > 0$.

Assumption 5′. Assumption 5 holds.

LEMMA 2 (SP, Thm. 2.2). *Suppose X_t is generated by (1) with $d_0 \in (\Delta_1, \Delta_2)$ and Assumptions 1′–5′ hold. Then $m^{1/2}(\widetilde{d} - d_0) \to_d N\left(0, \frac{1}{4}\right)$ as $n \to \infty$.*

3. ELW ESTIMATION WITH UNKNOWN MEAN

The asymptotic properties of the ELW estimator in Section 2 are derived under the assumption that X_t is generated by (1). However, when a researcher models an economic time series, typically its mean/initial condition is assumed to be unknown, and it is often accompanied by a linear time trend. In this section, we analyze the effect of an unknown mean/initial condition on the ELW estimation.

We consider estimating d when the data X_t are generated by

$$X_t = \mu_0 + X_t^0; \qquad X_t^0 = (1-L)^{-d_0} u_t \mathbf{1}\{t \geq 1\}, \tag{5}$$

where μ_0 is a nonrandom unknown finite number. Because $\mathrm{E}u_t = 0$, the initial condition μ_0 is also the mean of the process X_t in the sense $\mathrm{E}X_t = \mu_0$. Consider estimating μ_0 by $\widehat{\mu}$. One candidate for $\widehat{\mu}$ is the sample average $\overline{X} = n^{-1}\sum_{t=1}^{n} X_t$. For $d_0 > -\frac{1}{2}$, the error in estimating μ_0 by $\overline{X}$ is

$$\widehat{\mu} - \mu_0 = \overline{X} - \mu_0 = n^{-1}(1-L)^{-d_0-1} u_n \mathbf{1}\{t \geq 1\} = O_p(n^{d_0-1/2}). \tag{6}$$

Because the magnitude of the error increases as d_0 increases, the sample average is not a good estimate of μ_0 for large d_0. As shown by Adenstedt (1974), for a stationary process whose spectral density behaves like λ^{-2d} around the origin that includes a Type I $I(d)$ process with $d < \frac{1}{2}$, the best linear unbiased estimator (BLUE) of the mean is more efficient than the sample average.[3] When $d_0 < -\frac{1}{2}$, the BLUE has a variance that is $O(n^{2d_0-1})$. Samarov and Taqqu (1988) compare the variance of the BLUE with that of the sample average.

Note that, when $d_0 \geq \frac{1}{2}$, the variance of X_t^0 tends to infinity as $t \to \infty$ and the magnitude of X_t^0 dominates that of μ_0. Consequently, if $d_0 \geq \frac{1}{2}$, the signal on the value of d from X_t^0 dominates the noise from μ_0, and one can estimate d consistently from X_t without correcting for μ_0. In other words, there is no need to estimate μ_0. In a finite sample, however, it would be sensible to reduce the adverse effect of large μ_0 (10,000, say) by using the first observation X_1 as a proxy of μ_0. This leads to $\widehat{\mu} = X_1$, whose error in estimating μ_0 is[4]

$$\widehat{\mu} - \mu_0 = X_1 - \mu_0 = (1-L)^{-d_0} u_1 \mathbf{1}\{t \geq 1\} = u_1 = O_p(1). \tag{7}$$

Therefore, X_1 serves as another estimator of μ_0 for large d_0 and complements $\overline{X}$.

We state the results formally. Estimate μ_0 by $\widehat{\mu}$ and define the resulting estimator as

$$\widehat{d} = \arg\min_{d \in \Theta} R^{\diamond}(d), \tag{8}$$

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

$$R^{\diamond}(d) = \log \widehat{G}^{\diamond}(d) - 2d\frac{1}{m}\sum_{j=1}^{m} \log \lambda_j, \qquad \widehat{G}^{\diamond}(d) = \frac{1}{m}\sum_{j=1}^{m} I_{\Delta^d(x-\widehat{\mu})}(\lambda_j)$$

and $I_{\Delta^d(x-\widehat{\mu})}(\lambda_j)$ is the periodogram of $\Delta^d(X_t - \widehat{\mu})$. Define $\upsilon_t = \mathbf{1}\{t \geq 1\}$; then

$$w_{\Delta^d(x-\widehat{\mu})}(\lambda_j) = w_{\Delta^d(x-\mu_0)}(\lambda_j) + (\mu_0 - \widehat{\mu}) w_{\Delta^d \upsilon}(\lambda_j).$$

The asymptotics of the estimator depend on the relative magnitude of $w_{\Delta^d(x-\mu_0)}(\lambda_j)$ and $(\mu_0 - \widehat{\mu}) w_{\Delta^d \upsilon}(\lambda_j)$. The ELW estimator with $\widehat{\mu} = \overline{X}$ is consistent for $d_0 \in \left(-\frac{1}{2}, 1\right)$ and asymptotically normally distributed for $d_0 \in \left(-\frac{1}{2}, \frac{3}{4}\right)$, whereas the ELW estimator with $\widehat{\mu} = X_1$ is consistent and asymptotically normally distributed for $d_0 > 0$. The following theorems establish these results. Assumptions 6a and 6b nest Assumption 5.

Assumption 6a. $\Theta = [\Delta_1, \Delta_2]$ with $\frac{1}{2} < \Delta_1 < \Delta_2 < 1$.

THEOREM 1a. *Suppose X_t is generated by (5) with $d_0 \in [\Delta_1, \Delta_2]$, Assumptions 1–4 and 6a hold, and $\widehat{\mu} = \overline{X}$. Then $\widehat{d} \to_p d_0$ as $n \to \infty$.*

THEOREM 1b. *Suppose X_t is generated by (5) with $d_0 \in (\Delta_1, \Delta_2)$ and $\Delta_2 \leq \frac{3}{4}$, Assumptions 1′–4′ and 6a hold, and $\widehat{\mu} = \overline{X}$. Then $m^{1/2}(\widehat{d} - d_0) \to_d N\left(0, \frac{1}{4}\right)$ as $n \to \infty$.*

Assumption 6b. $\Theta = [\Delta_1, \Delta_2]$ with $0 < \Delta_1 < \Delta_2 < \infty$ and $\Delta_2 - \Delta_1 \leq \frac{9}{2}$.

THEOREM 2a. *Suppose X_t is generated by (5) with $d_0 \in [\Delta_1, \Delta_2]$, Assumptions 1–4 and 6b hold, $n^{1-2d_0} m^{-1+\eta} \log m \to 0$ for some $\eta > 0$, and $\widehat{\mu} = X_1$. Then $\widehat{d} \to_p d_0$ as $n \to \infty$.*

THEOREM 2b. *Suppose X_t is generated by (5) with $d_0 \in (\Delta_1, \min\{\Delta_2, 2\})$, Assumptions 1′–4′ and 6b hold, $n^{1-2d_0} m^{-1/2} \log n \to 0$, and $\widehat{\mu} = X_1$. Then $m^{1/2}(\widehat{d} - d_0) \to_d N(0, \frac{1}{4})$ as $n \to \infty$.*

Remark 1. We assume $\Delta_1 > -\frac{1}{2}$ in Theorems 1a and 1b, because the order of $\overline{X} - \mu_0$ is not given by (6) (indeed, it becomes $O_p(n^{-1} \log n)$) if $d_0 \leq -\frac{1}{2}$. 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. It may be possible to relax the restriction $\Delta_1 > -0.5$ if we use a BLUE in place of the sample average. The resulting estimator may suffer from complications, however, because the derivatives of the objective function involve the derivative of the BLUE with respect to d.

Remark 2. For Theorems 2a and 2b, any $O_p(1)$ variable will work as an estimator of μ_0. We choose X_1 to avoid (possibly) large error in estimating μ_0 in the worst case. As shown in (7), using X_1 as $\hat{\mu}$ gives $\hat{\mu} - \mu_0 = u_1$, whose order of magnitude will be smaller than that of X_n for moderately large n. On the other hand, if we use a constant (or an $O_p(1)$ random variable) c as $\hat{\mu}$, then the error associated with it is $\hat{\mu} - \mu_0 = c - \mu_0$, which may take a very large absolute value if c is chosen inappropriately. Then, the magnitude of $c - \mu_0$ may become as large as that of X_n for moderately large n, which has an adverse effect on the finite-sample performance of the estimator. Further, $X_t - X_1$ is shift-invariant, which would be desirable from an applied researcher's point of view.[5]

Remark 3. Theorems 1a–2b hold even if μ_0 is assumed to be an $O_p(1)$ random variable. The irrelevance of the initial condition for large d_0 can be highlighted through the unit root model. Suppose the data generating process (DGP) is $X_t = \mu_0 + X_t^0$, $t = 1, \ldots, n$, with $\mu_0 \neq 0$, $(1 - \rho L)X_t^0 = \varepsilon_t \mathbf{1}\{t \geq 1\}$, and $\varepsilon_t \sim iidN(0,1)$. Consider estimating ρ by regressing X_t on X_{t-1} without an intercept. When $|\rho| < 1$, the estimator $\hat{\rho} = (\sum_{t=2}^n X_{t-1}^2)^{-1} \sum_{t=2}^n X_{t-1} X_t$ converges to $[(1-\rho^2)^{-1} + \mu_0^2]^{-1}[\rho(1-\rho^2)^{-1} + \mu_0^2] \neq \rho$ and is not consistent for ρ. When $\rho = 1$, however, $\hat{\rho}$ converges to 1 because X_t dominates X_0 as $t \to \infty$. In our context, when $d_0 \geq \frac{1}{2}$, the variance of X_t^0 tends to infinity as $t \to \infty$, and X_t^0 dominates μ_0.

Remark 4. The additional assumptions on m in Theorems 2a and 2b are automatically satisfied when $d_0 \geq \frac{1}{2}$. When $d_0 \in (0, \frac{1}{2})$, these conditions require m to grow fast, and they become stronger for smaller d_0. This phenomenon occurs because, when $d_0 \in \left[0, \frac{1}{2}\right)$, both X_t^0 and $\widehat{\mu} - \mu_0$ are $O_p(1)$, but the leakage from the DFT of $\Delta^d(\widehat{\mu} - \mu_0)$ has a nonnegligible effect on the behavior of the periodogram ordinates for extremely small λ_j's. Trimming the first $\ell = \delta m$ periodogram ordinates for arbitrary small $\delta > 0$ will relax the condition to $n^{1-2d_0} m^{2d_0 - 2} \to 0$ for consistency.

Shimotsu and Phillips (2006) report a similar phenomenon with untapered local Whittle estimation; when the local Whittle estimator is applied to an $I(d_0)$ process with $d_0 \in \left[-1, -\frac{1}{2}\right)$, the consistency of the estimator requires m to grow fast. They report that Monte Carlo simulation bias can be as large as 0.25 when $d_0 = -1$, $n = 200$, and $m = 10$. The magnitude of the bias of the ELW estimator for $d_0 = 0$ in Table 1 is smaller, but the bias does manifest itself in some cases; e.g., the bias is 0.148 when $d_0 = 0$, $n = 4{,}096$, and $m = 30$ (not reported in Table 1).

Remark 5. In the model (5), the process is initialized at $t = 0$; however alternate initializations may be considered. Both versions of the ELW estimator are invariant to the initial condition when it does not depend on t, because both demeaning and subtracting X_1 annihilate it. When the initial condition depends on t, it may affect the asymptotics of the estimator. One example of such an initial condition is a distant past initialization (see Phillips and Lee, 1996; Canjels and Watson, 1997) that initializes the process at $[\kappa n]$ for a positive constant κ. This gives

$$X_t = \sum_{k=0}^{t+[n\kappa]} \frac{(d)_k}{k!} u_{t-k} = \sum_{k=0}^{t} \frac{(d)_k}{k!} u_{t-k} + \sum_{k=t+1}^{[\kappa n]} \frac{(d)_k}{k!} u_{t-k} = X_t^0 + X_t^\kappa.$$

When $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, this initialization locates X_t between Type I and Type II processes. Through an analogy with a Type I process discussed in Section 4.3, we conjecture that this initialization does not affect the asymptotic property of the

TABLE 1. Monte Carlo simulation bias: $n = 256$, $m = n^{0.65} = 36$

d_0	−0.4	0.0	0.4	0.8	1.2	1.6	2.0
$\widehat{\mu} = \overline{X}$	−0.0047	−0.0034	0.0000	0.0144	−0.0639	−0.3926	−0.8021
$\widehat{\mu} = X_1$	0.3066	0.0048	−0.0064	−0.0001	−0.0025	−0.0029	0.0000

ELW estimator, but proving it rigorously is beyond the scope of this paper. When $d \geq 1/2$, this type of initialization will affect the asymptotics, because X_t^0 and X_t^κ have the same order of magnitude. Phillips and Lee (1996) and Canjels and Watson (1997) derive its effect on the asymptotics of various estimators in the context of unit root models.

Remark 6. Heuristically, the lack of consistency for $d_0 = 0$ in Theorem 2a is explained as follows. When μ_0 is estimated by X_1, it follows from (7) that we estimate d of the series $X_t^0 - u_1$. We can then view u_1 as a persistent $O_p(1)$ noise added to X_t^0, because it is common for all t. When $d_0 \in \left(0, \frac{1}{2}\right)$, X_t^0 is $O_p(1)$, but is a persistent process, and the estimator picks up the signal about d_0 from X_t^0 correctly for some choices of m. When $d_0 = 0$, the persistence of u_1 prevents the estimator from detecting the signal about d_0 from X_t^0.

Remark 7. The semiparametric estimators that use the DFT of the original (not fractionally differenced) data are invariant to the initial condition, because the DFT of a constant term is 0 at nonzero Fourier frequencies. Omitting a linear trend in a stationary $I(d)$ process affects these estimators, however, because periodograms are not invariant to a linear trend. For example, Phillips and Shimotsu (2004) show the inconsistency of the local Whittle estimator when the data have a linear time trend.

Intriguingly, when $d_0 \in \left[\frac{1}{2}, 1\right)$, $\widehat{d}$ with $\widehat{\mu} = \overline{X}$ is still consistent, although $\overline{X}$ is not a consistent estimate of μ_0. Table 1 shows the finite-sample performance of the preceding two estimators. We generate the data according to (5) with $u_t \sim iidN(0,1)$ and $\mu_0 = 0$. The constants Δ_1 and Δ_2 are set to -1 and 3. Sample size and m are chosen to be $n = 256$ and $m = n^{0.65} = 36$, and 10,000 replications are used. The ELW estimator with $\widehat{\mu} = \overline{X}$ becomes negatively biased for large d_0, whereas the estimator with $\widehat{\mu} = X_1$ appears to be inconsistent when d_0 is negative. Consequently, the ELW estimator can become inconsistent if the error in estimating X_0 is not controlled properly.

4. TWO-STEP ELW ESTIMATION

4.1. Two-Step ELW Estimator

The results in Section 3 indicate that

1. $\overline{X}$ is an acceptable estimator of μ_0 for small d_0;

2. X_1 is an acceptable estimator of μ_0 for large d_0;
3. for $d_0 \in \left[\frac{1}{2}, \frac{3}{4}\right]$, both $\overline{X}$ and X_1 are acceptable estimators of μ_0.

Therefore, to estimate d when it has a wide range, we propose to estimate μ_0 with the following linear combination of $\overline{X}$ and X_1:

$$\widetilde{\mu}(d) = w(d)\overline{X} + (1 - w(d))X_1,$$

where $w(d)$ is a twice continuously differentiable weight function such that $w(d) = 1$ for $d \le \frac{1}{2}$ and $w(d) = 0$ for $d \ge \frac{3}{4}$. With this estimate of μ_0, consider the following modified ELW objective function:

$$R_F(d) = \log \widehat{G}_F(d) - 2d\frac{1}{m}\sum_{j=1}^{m} \log \lambda_j, \qquad \widehat{G}_F(d) = \frac{1}{m}\sum_{j=1}^{m} I_{\Delta^d(x - \widetilde{\mu}(d))}(\lambda_j). \quad \textbf{(9)}$$

One may consider estimating d by minimizing this objective function. Analyzing the resulting estimator encounters, however, the difficulty in proving its global consistency, as discussed in Section 4.4. We apply two-step estimation to circumvent this difficulty.[6]

Two-step estimation has a long history, dating back to the work by Fisher (1925). It has been analyzed by many authors, including LeCam (1956), Pfanzagl (1974), Janssen, Jureckova, and Veraverbeke (1985), and Robinson (1988). In the context of long-memory processes, Lobato (1999) and Lobato and Velasco (2000) use the two-step estimation method to simplify inference and avoid the problems associated with proving the consistency of the considered estimators for certain values of d.

Two-step estimation requires a $\sqrt{m}$-consistent first step estimator. We propose to use the tapered local Whittle estimators of Vel and HC in the first stage. The asymptotic theory of these estimators is derived under Type I long-range dependent processes that are defined as an infinite-order moving average of short-memory innovations for $d \in \left[-\frac{1}{2}, \frac{1}{2}\right)$ and as its partial sums for larger values of d. In Propositions B.1 and B.2 in Appendix B, we extend their theory to the case where X_t is generated by (11) from Section 4.2 (Type II processes) using the results in Robinson (2005). This result may be of interest itself because the asymptotic properties of these estimators have not been studied under Type II processes. Phillips and Shimotsu (2004) and Shimotsu and Phillips (2006) analyze the untapered local Whittle estimator under Type II processes.

To analyze the tapered estimators under Type II processes, we need to impose an additional assumption on $f_u(\lambda)$.

Assumption 7′. $f_u(\lambda)$ is bounded for $\lambda \in [0, \pi]$.

Assumption 7′ is used by Robinson (2005), who derives the stochastic order of the difference of the tapered DFTs of Type I and Type II processes. Not allowing $f_u(\lambda)$ to have poles outside the origin certainly restricts the class of the spectral density. However, it imposes no additional restrictions with respect to the smoothness of $f_u(\lambda)$ beyond Assumptions $1'$–$5'$. Propositions B.1 and B.2 in Appendix B show that the tapered estimators are asymptotically normally distributed and $\sqrt{m}$-consistent.

With the $\sqrt{m}$-consistency of the tapered estimators in hand, we are now ready to derive the limiting distribution of the two-step estimator. We focus on the tapered estimator by Vel with a third-order taper as the first-stage estimator because it allows for $d_0 \in \left(-\frac{1}{2}, \frac{5}{2}\right)$, is invariant to a linear and quadratic time trend, and requires a weaker assumption on $f_u(\lambda)$ than HC. Let $\widehat{d}_T$ denote this first-stage estimator and define the two-step ELW estimator, $\widehat{d}_{2ELW}$, as

$$\widehat{d}_{2ELW} = \widehat{d}_T - R'_F(\widehat{d}_T)/R''_F(\widehat{d}_T), \tag{10}$$

where $R_F(d)$ is the modified objective function defined in (9). Iterating the preceding procedure and updating the estimator by $\widehat{d}^{(2)}_{2ELW} = \widehat{d}_{2ELW} - R'_F(\widehat{d}_{2ELW})/R''_F(\widehat{d}_{2ELW})$ and similarly for $\widehat{d}^{(3)}_{2ELW}$ does not change the asymptotic distribution of the estimator, but we find that iterating procedure can substantially improve its finite-sample properties.

THEOREM 3. *Suppose X_t is generated by (5) with $d_0 \in (\Delta_1, \Delta_2)$ and $-\frac{1}{2} < \Delta_1 < \Delta_2 < 2$ and Assumptions $1'$–$5'$ and $7'$ hold. Then $m^{1/2}(\widehat{d}_{2ELW} - d_0) \to_d N(0, \frac{1}{4})$ as $n \to \infty$.*

Theorem 3 holds if the Hessian is replaced with 4 because $R''_F(\widehat{d}_T) \to_p 4$. In the simulations reported subsequently, we replaced $R''_F(\widehat{d}_T)$ with $\max\{R''_F(\widehat{d}_T), 2\}$ and found that it improves the finite-sample performance of the estimator. The lower bound on $R''_F(\widehat{d}_T)$ prevents the occurrence of extremely large values of $\widehat{d}_{2ELW}$.

One can also determine the range of d in the first step using a tapered estimator and then apply the nontapered estimator to properly differenced data. This pretest approach is a viable option, but its practical applicability depends on the significance level at which one can reject $d < \frac{1}{2}$ $\left(\text{or } d > \frac{1}{2}\right)$ in the first step. The two important factors in determining significance are the sample size and the distance between 1/2 and d of the data. With economic data, a priori, it is unclear whether these factors work favorably.

It might be possible to remove the restriction $d \neq \frac{1}{2}, \frac{3}{2}, \ldots$ of the FELW estimator by applying a similar strategy. For example, one may consider the following FELW counterpart of our approach: use (a) the ordinary DFT when $d \leq \frac{1}{2}$, (b) the extended DFT when $d \geq \frac{3}{4}$, and (c) their linear combination when $d \in \left[\frac{1}{2}, \frac{3}{4}\right]$. It is not clear, however, what the asymptotic distribution of the resulting estimator would be.

4.2. Two-Step ELW Estimation with Unknown Mean and Time Trend

In this section, we extend the two-step ELW estimation to cases where the data have a polynomial time trend, in addition to an unknown mean:

$$X_t = \mu_0 + \beta_{10} t + \beta_{20} t^2 + \cdots + \beta_{k0} t^k + X_t^0; \qquad X_t^0 = (1-L)^{-d_0} u_t \mathbf{1}\{t \geq 1\}. \quad \mathbf{(11)}$$

We propose to estimate d by regressing X_t on $(1, t, \ldots, t^k)$ and then applying the two-step estimation to the residuals $\widehat{X}_t$. As shown in the proof of Theorem 4 in Appendix A, the residuals can be expressed as

$$\widehat{X}_t = X_t^0 + \Xi_{0n}(d_0) + \Xi_{1n}(d_0) t + \cdots + \Xi_{kn}(d_0) t^k,$$

where $\Xi_{kn}(d_0)$ are random variables. Because $\sum_{t=1}^n \widehat{X}_t = 0$ by construction, the estimate of μ_0 from the residuals takes the form

$$\varphi(d) = (1 - w(d)) \widehat{X}_1.$$

The following theorem establishes the asymptotics. Note that asymptotic normality now requires d_0 to be smaller than 7/4, because the order of magnitude of the initial condition of $\widehat{X}_t$, $\Xi_{0n}(d_0)$, depends on d_0. The terms $\Xi_{1n}(d_0)t, \ldots, \Xi_{kn}(d_0)t^k$ have the same order of magnitude as $\Xi_{0n}(d_0)$.

THEOREM 4. *Suppose X_t is generated by (11) with $d_0 \in (\Delta_1, \Delta_2)$ and $-\frac{1}{2} < \Delta_1 < \Delta_2 \leq \frac{7}{4}$, Assumptions 1′–5′ and 7′ hold, and $\widehat{X}_t - \varphi(d)$ is used in place of $X_t - \widetilde{\mu}(d)$ in defining $R_F(d)$ in (9). Then $m^{1/2}(\widehat{d}_{2ELW} - d_0) \to_d N\left(0, \frac{1}{4}\right)$ as $n \to \infty$.*

A simulation result (available upon request) indicates that the two-step ELW estimator loses its consistency if one underspecifies the order of the polynomial k. Overspecifying the order of the polynomial does not change the asymptotic properties of the estimator but increases its finite-sample mean square errors (MSEs).

4.3. Two-Step ELW Estimation under Type I Processes

In this section, we discuss the effect of the specification of $I(d)$ processes on the asymptotics of the ELW estimators. Suppose Y_t is generated by a Type I $I(d_0)$ process plus an initial condition

$$Y_t = Y_t^0 + \mu_0, \qquad Y_t^0 = (1-L)^{-s} U_t^{(s)} \mathbf{1}\{t \geq 1\}, \qquad U_t^{(s)} = (1-L)^{-d_0+s} u_t,$$

where u_t satisfies Assumptions 1′–3′, $d_0 > -\frac{1}{2}$, and $s = \left[d_0 + \frac{1}{2}\right]$.

Consider the case where $\mu_0 = 0$ first. We conjecture that the two-step ELW estimator has the same asymptotic properties under Type I processes, albeit a rigorous proof is beyond the scope of this paper. First, it is known that Type I and Type II processes with $|d| < \frac{1}{2}$ are asymptotically equivalent (Marinucci and Robinson, 1999) and that the effect of their difference in their initialization becomes negligible as $t \to \infty$. Second, the untapered local Whittle estimator has $N\left(0, \frac{1}{4}\right)$ asymptotic distribution under both Type I (Robinson, 1995b) and Type II (Shimotsu and Phillips, 2006) processes. Therefore, we conjecture that the asymptotic equivalence between these processes will also apply to the asymptotic distribution of the semiparametric estimators.

Note that the ELW estimator uses the periodograms of the dth difference of the data with truncation at $t = 0$. In what follows, we show that the dth differences of Type I and Type II $I(d_0)$ processes truncated at $t = 0$ are asymptotically equivalent for $d \in [d_0 - \varepsilon, d_0 + \varepsilon]$ and small $\varepsilon > 0$. It suffices to consider this range of d because we use a two-step method. For illustration, focus on the case when $d_0 \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ and $Y_t = (1 - L)^{-d_0} u_t$. Taking the dth difference of Y_t with truncation gives

$$(1-L)^d Y_t \mathbf{1}\{t \geq 1\} = \sum_{k=0}^{t-1} \frac{(-d)_k}{k!} Y_{t-k} = \sum_{k=0}^{\infty} \frac{(-d)_k}{k!} Y_{t-k} - \sum_{k=t}^{\infty} \frac{(-d)_k}{k!} Y_{t-k}. \quad \textbf{(12)}$$

The first term on the right is $(1-L)^d (1-L)^{-d_0} u_t = (1-L)^{d-d_0} u_t$, which is a Type I $I(d_0 - d)$ process and asymptotically equivalent to a Type II $I(d_0 - d)$ process. For the second term, let γ_k denote the kth autocovariance of Y_t and assume that it satisfies $\gamma_k = O(k^{2d_0 - 1})$ for $d_0 \neq 0$ and $\sum_{-\infty}^{\infty} |\gamma_k| < \infty$ for $d_0 = 0$. Then, a tedious but routine calculation gives

$$\mathrm{E}\left[\sum_{k=t}^{\infty} \frac{(-d)_k}{k!} Y_{t-k}\right]^2 = O\left(\sum_{k=t}^{\infty} k^{-d-1} \sum_{l=t}^{\infty} l^{-d-1} |\gamma_{l-k}|\right) = O\left(t^{-2d-1} + t^{2d_0 - 2d - 1}\right).$$

Because $d_0 \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ and $|d - d_0| \leq \varepsilon$, this is $o(1)$ as $t \to \infty$, and the asymptotic equivalence of the two dth differenced processes follows.

When $\mu_0 \neq 0$, the two-step ELW estimation estimates μ_0 by a linear combination of the sample average and the first observation. Using Type I specification does not affect the asymptotic behavior of the two-step ELW estimator, because Type I and Type II processes have the same stochastic order and the basic intuition used in Type II specification carries through. Specifically, if we estimate μ_0 by the sample average of Y_t, then the estimation error is $n^{-1} \sum_{t=1}^{n} Y_t^0$, which is n^{-1} times a Type I $I(d_0 + 1)$ process and is $O_p(n^{d_0 - 1/2})$ under weak regularity conditions; see Marinucci and Robinson (1999) and the references therein. Therefore, the order of the error is the same as in (6). If we estimate μ_0 by Y_1, then $Y_1 - \mu_0 = Y_1^0 = U_1^{(s)} = O_p(1)$, and the order of the error is the same as in (7). Similarly, the effect of detrending polynomial trends is the same, because the partial sums of Type I and Type II processes have the same stochastic order.

4.4. Feasible ELW Estimator

We may consider estimating d by directly minimizing the objective function of the two-step ELW estimator. Define the resulting feasible ELW estimator as

$$\widehat{d}_F = \arg\min_{d\in\Theta} R_F(d), \tag{13}$$

where Θ is the space of the admissible values of d. This estimator is consistent and asymptotically normal for $d_0 > -\frac{1}{2}$, although we need to exclude a small interval around 0 and 1.

Assumption 6c. For arbitrary small $\nu > 0$, $\Theta = [\Delta_1, \Delta_2]\backslash((-\nu, \nu) \cup (1-\nu, 1+\nu))$ with $-\frac{1}{2} < \Delta_1 < \Delta_2 < 2$.

THEOREM 5a. *Suppose X_t is generated by (5) with $d_0 \in \Theta$ and Assumptions 1–5 and 6c hold. Then $\widehat{d}_F \to_p d_0$ as $n \to \infty$.*

THEOREM 5b. *Suppose X_t is generated by (5) with $d_0 \in Int(\Theta)$ and Assumptions 1′–5′ and 6c hold. Then $m^{1/2}(\widehat{d}_F - d_0) \to_d N\left(0, \frac{1}{4}\right)$ as $n \to \infty$.*

The exclusion of $(-\nu, \nu) \cup (1-\nu, 1+\nu)$ is necessary because of the difficulty in proving the global consistency of the estimator. The consistency is proved by showing that $R_F(d) - R_F(d_0)$ is uniformly bounded away from 0 when $d \neq d_0$. When d is close to d_0, $R_F(d) - R_F(d_0)$ converges to a nonrandom function whose minimum is achieved at d_0. When d is not close to d_0, in particular when $|d - d_0| \geq \frac{1}{2}$, $R_F(d) - R_F(d_0)$ does not converge to a nonrandom function, and we need an alternate way to bound it away from 0.[7] One of the necessary steps in proving the lower bound is to show, for some $\zeta > 0$,

$$m^{-1} \sum_{j=[\kappa m]}^{m} |A_j - B_j|^2 \geq \zeta\{m^{-1} \sum_{j=[\kappa m]}^{m} (|A_j|^2 + |B_j|^2)\}, \tag{14}$$

where κ is a fixed number between 0 and 1, A_j is a function of $\Delta^d X_t^0$, and B_j is a function of $w_{\Delta^d \upsilon}(\lambda_j)$. Their explicit formula is given by (A.35) in Appendix A. Note that (14) does not hold if $A_j = B_j \neq 0$. For (14) to hold, A_j and/or B_j must vary sufficiently as j changes so that $A_j - B_j$ is bounded away from 0 for sufficiently many j's. When d is close to 0, the two leading terms of $w_{\Delta^d \upsilon}(\lambda_j)$, $(1 - e^{i\lambda_j})^d$ and $-n^{-d}/\Gamma(1-d)$ (see Lemma B.2(i)), are both close to 1, which makes it very hard to establish that $w_{\Delta^d \upsilon}(\lambda_j)$ has sufficient variation. A similar difficulty arises when d is close to 1. SP also needed to use a nonstandard approach to show $\inf_d R(d) - R(d_0) > 0$ for $|d - d_0| \geq \frac{1}{2}$ but were able to show it for $\frac{1}{2} \leq |d - d_0| \leq \frac{9}{2}$. In a way, the presence of $w_{\Delta^d \upsilon}(\lambda_j)$ aggravates the difficulty in showing the global consistency in SP. Note that Theorems 1a, 1b, and 2a, 2b do not suffer the same problem as Theorems 5a, 5b. This is because the error in estimating μ_0 is small and dominated by X_t^0 for the combinations of d and d_0 considered in the previous theorems, which makes showing (14) less difficult.

5. SIMULATIONS AND AN EMPIRICAL APPLICATION

This section reports some simulation results. Here X_t is generated by (5) with $\mu_0 = 0$, and Δ_1 and Δ_2 are set to -1 and 3. The form of the weight function $w(d)$ for $d \in \left[\frac{1}{2}, \frac{3}{4}\right]$ is chosen to be $(1/2)[1 + \cos(4\pi d)]$. We use 10,000 replications. In two-step estimation, analytic derivatives are used to compute $R_F'(d)$ and $R_F''(d)$. The terms involving $\partial \widetilde{\mu}(d)/\partial d$ are omitted from the derivatives, because they are negligible in the limit. The procedure (10) is iterated (with updating) 10 times.

We compare the two-step ELW estimator with the tapered estimator by HC, which has the smallest limiting variance $\left(1.5/(4m) \text{ for } d \in \left(-\frac{1}{2}, \frac{3}{2}\right)\right)$ among the tapered estimators. We focus on $d \in \left(-\frac{1}{2}, \frac{3}{2}\right)$ in most of our simulations, because this is the range of d that is relevant for many economic applications.

Table 2 compares the two-step estimator with the tapered estimator for the value of $d \in \left(-\frac{1}{2}, \frac{3}{2}\right)$ and with varying short-run dynamics of u_t. The sample size and m are chosen to be $n = 512$ and $m = n^{0.65} = 57$, and u_t is modeled as an AR(1) with the parameter ρ. This table corresponds to Table 1 of HC. The bias of the two estimators is very similar and not affected by the changes in d for a given value of ρ. For a given value of d, the bias of both estimators increases as ρ increases. The variance of the two-step estimator is smaller than that of the tapered estimator for any parameter combination, corroborating the theoretical result.

Tables 3 and 4 compare the ELW estimator, the two-step ELW estimator with and without linear detrending, and the tapered estimator.[8] The estimation of the mean has little negative effect on the bias and standard deviation of the ELW estimator. Also, the MSE of the ELW estimator and the two-step estimator are virtually the same for $n = 512$. If the data are detrended prior to estimation, the two-step estimator suffers from a mild increase in standard deviation and a small

TABLE 2. Simulation results: $n = 512$, $m = n^{0.65} = 57$

		2ELW		Tapered estimator	
d	ρ	bias	var	bias	var
0.0	0.0	−0.0022	0.0058	0.0020	0.0094
0.0	0.5	0.0994	0.0061	0.1130	0.0098
0.0	0.8	0.4133	0.0072	0.4404	0.0114
0.4	0.0	0.0001	0.0058	−0.0030	0.0097
0.4	0.5	0.1003	0.0060	0.1055	0.0099
0.4	0.8	0.4160	0.0072	0.4381	0.0113
0.8	0.0	−0.0003	0.0058	−0.0066	0.0095
0.8	0.5	0.0988	0.0060	0.1014	0.0098
0.8	0.8	0.4125	0.0073	0.4325	0.0113
1.2	0.0	−0.0006	0.0057	−0.0057	0.0093
1.2	0.5	0.0990	0.0061	0.1022	0.0099
1.2	0.8	0.4117	0.0070	0.4302	0.0108

TABLE 3. Simulation results: $n = 128$, $m = n^{0.65} = 23$

	ELW			2ELW		
d	bias	s.d.	MSE	bias	s.d.	MSE
−0.4	0.0043	0.1369	0.0188	−0.0004	0.1383	0.0191
0.0	−0.0007	0.1397	0.0195	0.0001	0.1385	0.0192
0.4	0.0004	0.1404	0.0197	0.0052	0.1381	0.0191
0.8	−0.0008	0.1395	0.0195	0.0031	0.1338	0.0179
1.0	0.0006	0.1405	0.0197	0.0015	0.1377	0.0190
1.2	−0.0004	0.1390	0.0193	−0.0003	0.1386	0.0192
1.6	0.0023	0.1381	0.0191	0.0031	0.1380	0.0191
	2ELW with detrending			Tapered estimator		
d	bias	s.d.	MSE	bias	s.d.	MSE
−0.4	−0.0108	0.1340	0.0181	0.0434	0.1740	0.0322
0.0	−0.0444	0.1481	0.0239	0.0115	0.1757	0.0310
0.4	−0.0426	0.1550	0.0258	−0.0042	0.1783	0.0318
0.8	−0.0168	0.1536	0.0239	−0.0164	0.1787	0.0322
1.0	−0.0034	0.1442	0.0208	−0.0163	0.1783	0.0321
1.2	−0.0002	0.1398	0.0195	−0.0193	0.1757	0.0312
1.6	0.0132	0.1342	0.0182	−0.0074	0.1732	0.0301

TABLE 4. Simulation results: $n = 512$, $m = n^{0.65} = 57$

	ELW			2ELW		
d	bias	s.d.	MSE	bias	s.d.	MSE
−0.4	−0.0023	0.0765	0.0059	−0.0039	0.0764	0.0059
0.0	−0.0021	0.0774	0.0060	−0.0020	0.0774	0.0060
0.4	−0.0022	0.0772	0.0060	−0.0003	0.0765	0.0059
0.8	−0.0016	0.0771	0.0059	−0.0008	0.0762	0.0058
1.0	−0.0024	0.0768	0.0059	−0.0024	0.0767	0.0059
1.2	−0.0005	0.0768	0.0059	−0.0004	0.0769	0.0059
1.6	−0.0008	0.0772	0.0060	−0.0007	0.0772	0.0060
	2ELW with detrending			Tapered estimator		
d	bias	s.d.	MSE	bias	s.d.	MSE
−0.4	−0.0078	0.0759	0.0058	0.0131	0.0962	0.0094
0.0	−0.0214	0.0815	0.0071	0.0037	0.0977	0.0096
0.4	−0.0190	0.0818	0.0071	−0.0049	0.0984	0.0097
0.8	−0.0059	0.0802	0.0065	−0.0069	0.0985	0.0097
1.0	−0.0035	0.0774	0.0060	−0.0086	0.0973	0.0095
1.2	0.0001	0.0769	0.0059	−0.0058	0.0966	0.0094
1.6	0.0060	0.0770	0.0060	−0.0011	0.0957	0.0092

negative bias for $d = 0.0 \sim 0.8$. Overall, the finite-sample performance of the two-step estimator 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.

As explained after Theorem 3, the simulations in Tables 1–4 are conducted using $\max\{R''_F(d), 2\}$ instead of $R''_F(d)$ to avoid the undesirable effects of very small values of $R''_F(d)$, which result in extremely large values of the updated d. Table 5 compares the two-step estimator with four different choices of the denominator in

TABLE 5. Simulation results with different Hessians

	$n = 128, m = n^{0.65} = 23$					
	$R''_F(d)$			$\max\{R''_F(d), 2\}$		
d	bias	s.d.	MSE	bias	s.d.	MSE
0.0	−0.6832	9.6644	93.8666	0.0046	0.1363	0.0186
0.4	−0.9285	12.6518	160.9299	0.0042	0.1370	0.0188
0.8	−1.6182	14.5551	214.4700	0.0035	0.1343	0.0181
1.2	−1.4070	14.6224	215.7950	−0.0001	0.1385	0.0192
	$\max\{R''_F(d), 3\}$			4		
d	bias	s.d.	MSE	bias	s.d.	MSE
0.0	0.0047	0.1363	0.0186	0.0051	0.1363	0.0186
0.4	0.0043	0.1370	0.0188	0.0047	0.1371	0.0188
0.8	0.0037	0.1340	0.0180	0.0043	0.1338	0.0179
1.2	−0.0001	0.1385	0.0192	0.0004	0.1384	0.0192
	$n = 512, m = n^{0.65} = 57$					
	$R''_F(d)$			$\max\{R''_F(d), 2\}$		
d	bias	s.d.	MSE	bias	s.d.	MSE
0.0	−0.0022	0.0764	0.0058	−0.0022	0.0764	0.0058
0.4	−0.0002	0.0769	0.0059	−0.0002	0.0769	0.0059
0.8	−0.0015	0.0761	0.0058	−0.0015	0.0761	0.0058
1.2	−0.0014	0.0771	0.0059	−0.0014	0.0771	0.0059
	$\max\{R''_F(d), 3\}$			4		
d	bias	s.d.	MSE	bias	s.d.	MSE
0.0	−0.0022	0.0764	0.0058	−0.0022	0.0764	0.0058
0.4	−0.0002	0.0769	0.0059	−0.0002	0.0769	0.0059
0.8	−0.0015	0.0761	0.0058	−0.0015	0.0761	0.0058
1.2	−0.0014	0.0771	0.0059	−0.0014	0.0771	0.0059

TABLE 6. Simulation results of the FELW estimator

	$n = 128, m = n^{0.65} = 23$					
	FELW			FELW with detrending		
d	bias	s.d.	MSE	bias	s.d.	MSE
−0.4	0.0068	0.1351	0.0183	−0.0046	0.1407	0.0198
0.0	−0.0167	0.1356	0.0187	−0.0485	0.1417	0.0224
0.4	−0.0099	0.1469	0.0217	−0.0616	0.1595	0.0292
0.8	−0.0272	0.1354	0.0191	−0.0329	0.1467	0.0226
1.0	−0.0329	0.1367	0.0198	−0.0332	0.1376	0.0200
1.2	−0.0375	0.1371	0.0202	−0.0374	0.1376	0.0203
1.6	−0.0443	0.1325	0.0195	−0.0397	0.1325	0.0191
	$n = 512, m = n^{0.65} = 57$					
	FELW			FELW with detrending		
d	bias	s.d.	MSE	bias	s.d.	MSE
−0.4	0.0014	0.0799	0.0064	−0.0020	0.0816	0.0067
0.0	−0.0066	0.0773	0.0060	−0.0219	0.0800	0.0069
0.4	0.0005	0.0853	0.0073	−0.0243	0.0884	0.0084
0.8	−0.0090	0.0775	0.0061	−0.0091	0.0777	0.0061
1.0	−0.0127	0.0765	0.0060	−0.0127	0.0765	0.0060
1.2	−0.0109	0.0770	0.0060	−0.0109	0.0770	0.0060
1.6	−0.0150	0.0822	0.0070	−0.0102	0.0808	0.0066

the updating formula (10): $R''_F(d)$, $\max\{R''_F(d), 2\}$, $\max\{R''_F(d), 3\}$, and 4.[9] The results for $\max\{R''_F(d), 2\}$, $\max\{R''_F(d), 3\}$, and 4 are almost identical for all n. When $n = 128$, $R''_F(d)$ suffers from a relatively larger standard deviation and MSE than the other variants. When $n = 512$, the performances of the four versions are identical.

Table 6 reports the simulation results for the fully extended (FELW) estimator of Abadir et al. (2007), with and without detrending to compare it with the two-step estimator. The first panel of Table 6 uses the same DGP as Table 3, and the second panel of Table 6 uses the same DGP as Table 4. Comparing Table 6 with Tables 3 and 4, it can be seen that detrending increases the MSE of both the two-step and fully extended estimator by an increase in their variance. The MSE of the two-step and fully extended estimator is similar, with neither dominating the other.

Table 7 shows the performance of the ELW estimator and the two-step ELW estimator under Type I processes with $n = 128$ and 512 to examine the conjecture in Section 4.2. When $n = 128$, the variance of both estimators appears to be slightly larger than their variance under Type II processes reported in Table 3. The results with $n = 512$ are very similar to the corresponding ones in Table 4.

TABLE 7. Simulation results with Type I processes

	$n = 128, m = n^{0.65} = 23$					
	ELW			2ELW		
d	bias	s.d.	MSE	bias	s.d.	MSE
−0.4	0.0023	0.1473	0.0217	0.0129	0.1405	0.0199
0.0	−0.0004	0.1391	0.0193	0.0003	0.1386	0.0192
0.4	0.0050	0.1559	0.0243	0.0110	0.1379	0.0191
0.8	0.0008	0.1418	0.0201	0.0048	0.1349	0.0182
1.0	−0.0004	0.1393	0.0194	0.0003	0.1375	0.0189
1.2	0.0014	0.1401	0.0196	0.0016	0.1385	0.0192
1.6	−0.0013	0.1482	0.0220	−0.0010	0.1477	0.0218
	$n = 512, m = n^{0.65} = 57$					
	ELW			2ELW		
d	bias	s.d.	MSE	bias	s.d.	MSE
−0.4	0.0010	0.0794	0.0063	0.0057	0.0785	0.0062
0.0	−0.0025	0.0781	0.0061	−0.0024	0.0781	0.0061
0.4	0.0022	0.0797	0.0064	0.0011	0.0756	0.0057
0.8	−0.0016	0.0775	0.0060	−0.0013	0.0766	0.0059
1.0	−0.0019	0.0765	0.0059	−0.0020	0.0766	0.0059
1.2	−0.0008	0.0774	0.0060	−0.0007	0.0774	0.0060
1.6	0.0001	0.0796	0.0063	0.0002	0.0795	0.0063

TABLE 8. Estimates of d for U.S. economic data: $m = n^{0.7}$

	n	LW	2ELW	95% asy. CI
Real GNP	80	1.077	1.126	[0.912, 1.340]
Nominal GNP	80	1.273	1.303	[1.089, 1.517]
Real per capita GNP	80	1.077	1.128	[0.914, 1.342]
Industrial production	129	0.821	0.850	[0.671, 1.029]
Employment	99	0.968	1.000	[0.800, 1.200]
Unemployment rate	129	0.951	0.980	[0.801, 1.159]
GNP deflator	100	1.374	1.398	[1.202, 1.594]
CPI	129	1.273	1.287	[1.109, 1.466]
Nominal wage	89	1.300	1.351	[1.147, 1.555]
Real wage	89	1.047	1.089	[0.885, 1.293]
Money stock	100	1.460	1.501	[1.305, 1.697]
Velocity of money	120	0.953	0.993	[0.808, 1.179]
Bond yield	89	1.091	1.108	[0.903, 1.312]
Stock prices	118	0.900	0.958	[0.772, 1.143]

As an empirical illustration, the two-step 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 (LW) estimator, and adding unity to the estimate $\widehat{d-1}$. This procedure is invariant to the linear trend. For the two-step ELW estimates, 95% asymptotic confidence intervals are constructed by adding and subtracting $1.96 \times 1/\sqrt{4m}$ to the estimates. Table 8 shows the results based on $m = n^{0.7}$. The two estimates are fairly close to each other. For real measures such as real gross national product (GNP), real per capita GNP, and employment, the estimates are close to 1. For price variables such as the GNP deflator, consumer price index (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 $H_0 : d = 0$ is accepted in none of the series. Crato and Rothman (1994) obtained a similar result using the autoregressive fractionally integrated moving average (ARFIMA) model; therefore it appears that the case for trend stationarity is weaker than has been suggested from the KPSS test of Kwiatkowski, Phillips, Schmidt, and Shin (1992).

NOTES

1. The two-step ELW estimator also assumes that the spectral density has no poles outside the origin. This restriction excludes, e.g., seasonal long memory.
2. Section 7 of Shimotsu and Phillips (2006) discusses the difference between Type I and II processes in more detail.
3. We thank a referee for bringing this to our attention.
4. Using an arbitrary random variable Z as $\widehat{\mu}$ results in the same order of the error.
5. We thank a referee for bringing this to our attention.
6. The idea of two-step estimation was originally suggested by an anonymous reader of SP, albeit in a different context.
7. In the proof of Theorem 3a, we use the fact that $d \notin (-\nu,\nu) \cup (1-\nu, 1+\nu)$ in showing the necessary results for Θ_1^a, even though $|\theta|$ may be smaller than $\frac{1}{2}$ in Θ_1^a. We can prove the necessary results for Θ_1^a without using $d \notin (-\nu,\nu) \cup (1-\nu, 1+\nu)$, although the derivation is more tedious.
8. For the tapered estimator, the result for $d = 1.6$ is only for reference, because the tapered estimator with taper of order 1 is asymptotically normal only for $d < \frac{2}{3}$.
9. When $n = 128$, iterating the procedure (10) with $R_F''(d)$ failed to work in a small number of (less than 0.3%) simulation draws. These instances arose when the updated d took extremely large values (>500), resulting in numerical error in the evaluation of $R_F''(d)$. The statistics are computed excluding such trials.

REFERENCES

Abadir, K.M., W. Distaso, & L. Giraitis (2007) Nonstationarity-extended local Whittle estimation. *Journal of Econometrics* 141, 1353–1384.

Adenstedt, R. (1974) On large-sample estimation for the mean of a stationary random sequence. *Annals of Mathematical Statistics* 2, 1095–1107.

Backus, D.K. & 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.

Canjels, E. & M.W. Watson (1997) Estimating deterministic trends in the presence of serially correlated errors. *Review of Economics and Statistics* 79, 184–200.

Crato, N. & P. Rothman (1994) Fractional integration analysis of long-run behavior for US macroeconomic time series. *Economics Letters* 45, 287–291.

Diebold, F.X. & G.D. Rudebusch (1991) Is consumption too smooth? Long memory and the Deaton paradox. *Review of Economics and Statistics* 73, 1–9.

Fisher, R.A. (1925) Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society* 22, 700–725.

Gil-Alaña, L.A. & P.M. Robinson (1997) Testing of unit root and other nonstationary hypotheses in macroeconomic time series. *Journal of Econometrics* 80, 241–268.

Hassler, U. & J. Wolters (1995) Long memory in inflation rates: International evidence. *Journal of Business & Economic Statistics* 13, 37–45.

Haubrich, J.G. (1993) Consumption and fractional differencing: Old and new anomalies. *Review of Economics and Statistics* 75, 767–772.

Henry, M. & P. Zaffaroni (2003) The long range dependence paradigm for macroeconomics and finance. In P. Doukhan, G. Oppenheim, & M. Taqqu, eds., *Long-Range Dependence: Theory and Applications*, pp. 419–438. Birkhauser.

Hurvich, C.M. & W.W. Chen (2000) An efficient taper for potentially overdifferenced long-memory time series. *Journal of Time Series Analysis* 21, 155–180.

Janssen, P., J. Jureckova, & N. Veraverbeke (1985) Rate of convergence of one- and two-step M-estimators with applications to maximum likelihood and Pitman estimators. *Annals of Statistics* 13, 1222–1229.

Kim, C.S. & P.C.B. Phillips (2006) Log Periodogram Regression: The Nonstationary Case. Cowles Foundation Discussion paper 1587, Yale University.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, & 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.

LeCam, L. (1956) On the asymptotic theory of estimation and testing hypotheses. In J. Neyman, ed., *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* 1, pp. 129–156. University of California Press.

Lobato, I.N. (1999) A semiparametric two-step estimator in a multivariate long memory model. *Journal of Econometrics* 90, 129–153.

Lobato, I.N. & C. Velasco (2000) Long memory in stock-market trading volume. *Journal of Business & Economic Statistics* 18, 410–427.

Marinucci, D. & P.M. Robinson (1999) Alternative forms of fractional Brownian notion. *Journal of Statistical Planning and Reference* 80, 111–122.

Maynard, A. & P.C.B. Phillips (2001) Rethinking an old empirical puzzle: Econometric evidence on the forward discount anomaly. *Journal of Applied Econometrics* 16, 671–708.

Michelacci, C. & P. Zaffaroni (2000) (Fractional) beta convergence. *Journal of Monetary Economics* 45, 129–153.

Nelson, C.R. & C.I. Plosser (1982) Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics* 10, 139–162.

Pfanzagl, J. (1974) Asymptotic optimum estimation and test procedures. In J. Hájek, ed., *Proceedings of the Prague Symposium on Asymptotics*, vol. 1, pp. 201–272. Charles University.

Phillips, P.C.B. (2007) Unit root log periodogram regression. *Journal of Econometrics* 138, 104–124.

Phillips, P.C.B. & C.C. Lee (1996) Efficiency gains from quasi-differencing under nonstationarity. In P.M. Robinson & M. Rosenblatt, eds., *Essays in Memory of E. J. Hannan*, pp. 300–314. Springer.

Phillips, P.C.B. & K. Shimotsu (2004) Local Whittle estimation in nonstationary and unit root cases. *Annals of Statistics* 32, 656–692.

Robinson, P.M. (1988) The stochastic difference between econometric statistics. *Econometrica* 56, 531–548.

Robinson, P.M. (1994) Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association* 89, 1420–1437.

Robinson, P.M. (1995a) Log-periodogram regression of time series with long range dependence. *Annals of Statistics* 23, 1048–1072.
Robinson, P.M. (1995b) Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* 23, 1630–1661.
Robinson, P.M. (2005) The distance between rival nonstationary fractional processes. *Journal of Econometrics* 128, 283–300.
Samarov, A. & M.S. Taqqu (1988) On the efficiency of the sample mean in long memory noise. *Journal of Time Series Analysis* 9, 191–200.
Schotman, P. & H.K. van Dijk (1991) On Bayesian routes to unit roots. *Journal of Applied Econometrics* 6, 387–401.
Shimotsu, K. & P.C.B. Phillips (2005) Exact local Whittle estimation of fractional integration. *Annals of Statistics* 33, 1890–1933.
Shimotsu, K. & P.C.B. Phillips (2006) Local Whittle estimation of fractional integration and some of its variants. *Journal of Econometrics* 130, 209–233.
Velasco, C. (1999) Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* 20, 87–127.

APPENDIX A: Proofs

In this and the following section, x^* denotes the complex conjugate of x; C and ε denote generic constants such that $C \in (1,\infty)$ and $\varepsilon \in (0,1)$ unless specified otherwise, and they may take different values in different places; and ξ_n denotes a generic random variable that does not depend on j and satisfies $\mathrm{E}|\xi_n|^2 < \infty$. Henceforth, let $I_{\Delta xj}$ denote $I_{\Delta x}(\lambda_j)$, w_{uj} denote $w_u(\lambda_j)$, and similarly for other DFTs and periodograms.

A.1. Proof of Theorem 1a. Assume $\mu_0 = 0$ without loss of generality. We follow the approach developed by SP. Define $S^\diamond(d) = R^\diamond(d) - R^\diamond(d_0)$. For arbitrary small $0 < \Delta < \frac{1}{8}$, define $\Theta_1 = \left\{d_0 - \frac{1}{2} + \Delta \le d \le d_0 + \frac{1}{2}\right\}$ and $\Theta_2 = \left\{d \in \left[\Delta_1, d_0 - \frac{1}{2} + \Delta\right] \cup \left[d_0 + \frac{1}{2}, \Delta_2\right]\right\}$, Θ_2 being possibly empty. For $\frac{1}{2} > \rho > 0$, define $N_\rho = \{d : |d - d_0| < \rho\}$. From SP (pp. 1900–1901), we have

$$\Pr\left(|\widehat{d} - d_0| \ge \rho\right) \le \Pr\left(\inf_{d \in \Theta_1 \setminus N_\rho} S^\diamond(d) \le 0\right) + \Pr\left(\inf_{\Theta_2} S^\diamond(d) \le 0\right). \quad \textbf{(A.1)}$$

As in SP (between eqns. (13) and (14), p. 1902), define $\theta = d - d_0$ and

$$Y_t(\theta) = (1-L)^d X_t = (1-L)^{d-d_0}(1-L)^{d_0} X_t = (1-L)^\theta u_t \mathbf{1}\{t \ge 1\}.$$

Note that $R^\diamond(d)$ is constructed by replacing $I_{\Delta^d xj}$ in the objective function of SP, $R(d)$, with $I_{\Delta^d(x-\widehat{\mu})j}$. Because $w_{\Delta^d(x-\widehat{\mu})j} = w_{\Delta^d xj} - \widehat{\mu} w_{\Delta^d vj} = w_{yj} - \widehat{\mu} w_{\Delta^d vj}$, the theorem is proved by replacing w_{yj} in SP with $w_{yj} - \widehat{\mu} w_{\Delta^d vj}$ and showing the results in SP carry through. We only state the main steps and refer the readers to SP for further details.

As in equation (15) of SP, define $A(d) = (2(d-d_0)+1)m^{-1}\sum_{j=1}^m (j/m)^{2\theta}[\lambda_j^{-2\theta} I_{yj} - G_0]$. To show that the first probability on the right of (A.1) tends to 0, we need to replace I_{yj} in $A(d)$ with $|w_{yj} - \widehat{\mu} w_{\Delta^d vj}|^2$ and show that $\sup_{\Theta_1} |A(d)| \to 0$ still holds. Because

$$I_{yj} - |w_{yj} - \widehat{\mu} w_{\Delta^d vj}|^2 = 2\widehat{\mu}\mathrm{Re}[w_{yj} w^*_{\Delta^d vj}] - \widehat{\mu}^2 I_{\Delta^d vj}, \quad \textbf{(A.2)}$$

it suffices to show, for any finite k,

$$\sup_{\theta\in\Theta_1}\left|\widehat{\mu}\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{2\theta}\lambda_j^{-2\theta}w_{yj}w_{\Delta^d vj}^{*}\right| = o_p((\log n)^{-k}), \tag{A.3}$$

$$\sup_{\theta\in\Theta_1}\left|\widehat{\mu}^2\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{2\theta}\lambda_j^{-2\theta}I_{\Delta^d vj}\right| = o_p((\log n)^{-k}). \tag{A.4}$$

We proceed to derive the order of $\widehat{\mu}w_{\Delta^d vj}$ and show (A.3) and (A.4). Because $w_{\Delta^d vj} = O(j^{d-1}n^{1/2-d})$ for $d \geq 0$ and $O(j^{-1}n^{1/2-d})$ for $d \leq 0$ from Lemma B.2 in Appendix B, we obtain

$$\lambda_j^{-\theta}w_{\Delta^d vj} = \begin{cases} O(n^{1/2-d_0}j^{d_0-1}), & d\in[0,\Delta_2], \\ O(n^{1/2-d_0}j^{-\theta-1}), & d\in[-1+\varepsilon,0] \end{cases} \tag{A.5}$$

uniformly in d and $j = 1,\ldots,m$. Observe that $\widehat{\mu} = n^{-1}\sum_{t=1}^{n}X_t = n^{-1}(1-L)^{-d_0-1}u_t \mathbf{1}\{t\geq 1\}$ with $d_0 > -\frac{1}{2}$. We can show that $\mathrm{E}\left[(1-L)^{-d_0-1}u_t\mathbf{1}\{t\geq 1\}\right]^2 = O\left(n^{2d_0+1}\right)$ easily from Lemma A.5(a2) of Phillips and Shimotsu (2004), and it follows that $\mathrm{E}\widehat{\mu}^2 = O\left(n^{2d_0-1}\right)$ and

$$\widehat{\mu}\cdot\lambda_j^{-\theta}w_{\Delta^d vj} = \begin{cases} \xi_n\cdot O\left(j^{d_0-1}\right), & d\geq 0, \\ \xi_n\cdot O\left(j^{-\theta-1}\right), & d\leq 0, \end{cases} \tag{A.6}$$

where $O(\cdot)$ terms are uniform in d and $j = 1,\ldots,m$. We also have, uniformly in $\alpha\in[-C,C]$ (note that $\Theta_1 = \left\{-\frac{1}{2}+\Delta\leq\theta\leq\frac{1}{2}\right\}$),

$$\sup_{\theta\in\Theta_1}\left|\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{2\theta}j^{\alpha}\right| \leq m^{\alpha}\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{2\Delta-1+\alpha} = O(m^{\alpha}\log m + m^{-2\Delta}\log m), \tag{A.7}$$

where the order of magnitude follows from considering the cases where $2\Delta-1+\alpha\geq -1$ and $2\Delta-1+\alpha\leq -1$ separately. Therefore, (A.4) follows from (A.6), (A.7), and the fact that $d_0 < 1$ and $|\theta|\leq\frac{1}{2}$ in Θ_1. For (A.3), its left-hand side is bounded by $(\sup_{\theta\in\Theta_1} m^{-1}\sum_{j=1}^{m}(j/m)^{2\theta}\lambda_j^{-2\theta}I_{yj})^{1/2}\cdot(\sup_{\theta\in\Theta_1}\widehat{\mu}^2 m^{-1}\sum_{j=1}^{m}(j/m)^{2\theta}\lambda_j^{-2\theta}I_{\Delta^d vj})^{1/2}$. The first term is $O_p(1)$ uniformly in $\theta\in\Theta_1$ because $\sup_{\Theta_1}|A(d)| = o_p(1)$ and $\sup_{\Theta_1}\left|(2\theta+1)m^{-1}\sum_{j=1}^{m}(j/m)^{2\theta}-1\right| = o(1)$ as shown in SP (p. 1903). Hence (A.3) follows from (A.4).

We now show that the second probability on the right of (A.1) tends to 0. As in SP, let $\kappa\in(0,1)$ and let $\sum'$ denote the sum over $j = [\kappa m],\ldots,m$. From the argument on pages 1904–1905 of SP that leads to their equation (23), the second probability on the right of (A.1) tends to 0 if there exists $\delta > 0$ such that

$$\Pr\left(\inf_{\Theta_2}\left(m^{-1}{\sum}'(j/p)^{2\theta}\left(\lambda_j^{-2\theta}|w_{yj}-\widehat{\mu}w_{\Delta^d vj}|^2 - G_0\right)\right)\leq -3\delta G_0\right)\to 0, \tag{A.8}$$

where $p = \exp(m^{-1}\sum_{j=1}^{m}\log j)\sim m/e$ as $m\to\infty$.

We show (A.8) for subsets of Θ_2. Define $\eta = 1 - d_0 > 0$ and split Θ_2 into two, $\Theta_2^a = \{\theta \geq -1 + \eta/2\} \cap \Theta_2$ and $\Theta_2^b = \{\theta \leq -1 + \eta/2\} \cap \Theta_2$. First, SP show (eqn. (23), p. 1905) that (A.8) holds if $|w_{yj} - \widehat{\mu} w_{\Delta^d vj}|^2$ is replaced by $I_y(\lambda_j)$. Second, if $\theta \in \Theta_2^a$ or $d > 0$, we have

$$\sup_{\Theta_2} \left| \widehat{\mu} m^{-1} \sum{}' (j/p)^{2\theta} \lambda_j^{-2\theta} w_{yj} w_{\Delta^d vj}^* \right| + \sup_{\Theta_2} \left| \widehat{\mu}^2 m^{-1} \sum{}' (j/p)^{2\theta} \lambda_j^{-2\theta} I_{\Delta^d vj} \right| = o_p(1),$$

from using the bound in (A.6) with $d_0 \leq 1 - \eta$ and $-\theta - 1 \leq -\eta/2$ and proceeding as in the proof of (A.3) and (A.4) with Lemma 5.4 of SP. Thus, (A.8) holds for $\theta \in \Theta_2^a$ or $d > 0$.

If $\theta \in \Theta_2^b$ and $d < 0$, we cannot use (A.6) because $-\theta - 1$ may take a positive value and its left-hand side is not $o_p(1)$. Note that $|\theta| = |d - d_0| \leq \frac{3}{2}$ because $d, d_0 \in (-\frac{1}{2}, 1)$. Define $\Theta_2^3 = \left\{ -\frac{3}{2} \leq \theta \leq -\frac{1}{2} \right\}$ as in SP (p. 1909); then Θ_2^b is a subset of Θ_2^3. We show the required result by replacing $\lambda_j^{-\theta} w_{yj}$ in the corresponding proof for Θ_2^3 in SP (eqn. (45), p. 1909) with $\lambda_j^{-\theta} w_{yj} - \lambda_j^{-\theta} \widehat{\mu} w_{\Delta^d vj}$ and showing that their argument carries through. Replacing $\lambda_j^{-\theta} (2\pi n)^{-1/2} e^{i\lambda_j} (1 - e^{i\lambda_j})^{-1} Y_n(\theta)$ on the right of (45) of SP with

$$\lambda_j^{-\theta} (2\pi n)^{-1/2} e^{i\lambda_j} (1 - e^{i\lambda_j})^{-1} Y_n(\theta) - \lambda_j^{-\theta} \widehat{\mu} w_{\Delta^d vj}, \tag{A.9}$$

we find that (47) in SP needs to be replaced with

$$m^{-1} \sum{}' (j/p)^{2\theta} \lambda_j^{-2\theta} \left| (2\pi n)^{-1/2} e^{i\lambda_j} (1 - e^{i\lambda_j})^{-1} Y_n(\theta) - \widehat{\mu} w_{\Delta^d vj} \right|^2 \tag{A.10}$$

and their equations (49) and (50) have additional terms

$$+ 2\text{Re}[m^{-1} \sum{}' (j/p)^{2\theta} \overline{U}_{nj}(\theta) \lambda_j^{-\theta} \widehat{\mu} w_{\Delta^d vj}], \tag{A.11}$$

$$- 2\text{Re}[m^{-1} \sum{}' (j/p)^{2\theta} D_{nj}(\theta)^* w_{uj}^* \lambda_j^{-\theta} \widehat{\mu} w_{\Delta^d vj}], \tag{A.12}$$

where $D_{nj}(\theta)$ and $\overline{U}_{nj}(\theta)$ are defined on p. 1909 of SP. Then, in view of the bounds of (48)–(50) in SP provided on page 1910 of SP, (A.8) holds by Lemma B.1 if

$$(A.10) \geq \zeta m^{-2\theta-2} n^{2\theta+1} Y_n(\theta)^2 + \zeta n^{1-2d+2\theta} m^{-2-2\theta} \widehat{\mu}^2 \tag{A.13}$$

for some $\zeta > 0$ and

$$(A.11) + (A.12) = n^{1/2-d+\theta} m^{-1-\theta} \widehat{\mu} \cdot O_p(m^{-\eta/2} \log n + mn^{-1}). \tag{A.14}$$

Loosely speaking, if (A.13) and (A.14) hold, then (A.11) and (A.12) are dominated by (A.10). It remains to show (A.13) and (A.14). Note that $d \leq -\nu$. For (A.13), applying Lemma B.3(ii) with $Q_2 = Y_n(\theta)$, $Q_1 = 0$, and $Q_0 = -\widehat{\mu}$ gives

$$\begin{aligned} (A.10) &= p^{-2\theta} (2\pi)^{-2\theta} n^{2\theta} m^{-1} \sum{}' \left| (2\pi n)^{-1/2} e^{i\lambda_j} (1 - e^{i\lambda_j})^{-1} Y_n(\theta) - \widehat{\mu} w_{\Delta^d vj} \right|^2 \\ &\geq \zeta m^{-2\theta} n^{2\theta} [n m^{-2} Y_n(\theta)^2 + n^{1-2d} m^{-2} \widehat{\mu}^2], \end{aligned}$$

giving (A.13). Equation (A.14) follows from applying Lemma B.4 with $\alpha = d$ to (A.11) and (A.12), because $\overline{U}_{nj}(\theta)$ and $D_{nj}(\theta)$ satisfy the assumptions of Lemma B.4 from equations (39) and (31) of SP, respectively. Thus (A.8) holds, and we complete the proof. ■

A.2. Proof of Theorem 1b. Assume $\mu_0 = 0$ without loss of generality. Theorem 1a holds under the current conditions and implies that with probability approaching 1, as $n \to \infty$, $\widehat{d}$ satisfies

$$0 = R^{\diamond\prime}(\widehat{d}) = R^{\diamond\prime}(d_0) + R^{\diamond\prime\prime}(\overline{d})(\widehat{d} - d_0), \tag{A.15}$$

where $|\overline{d} - d_0| \le |\widehat{d} - d_0|$. Again the theorem is proved by replacing w_{yj} in SP with $w_{yj} - \widehat{\mu} w_{\Delta^d vj}$ and showing that the results in SP carry through. Fix $\rho > 0$ and let $M = \{d : (\log n)^4 |d - d_0| < \rho\}$. Note that $\sup_{\Theta_1} |A(d)| = o_p((\log n)^{-10})$ still holds even if we replace $I_y(\lambda_j)$ in $A(d)$ with $|w_{yj} - \widehat{\mu} w_{\Delta^d vj}|^2$, because the order of the additional terms shown in (A.3) and (A.4) is smaller than $(\log n)^{-10}$. Therefore, $\Pr(\overline{d} \notin M)$ tends to 0 in view of equation (55) of SP and the argument surrounding it. Thus we assume $d \in M$ in what follows.

First we show $R^{\diamond\prime\prime}(\overline{d}) \to_p 4$. Define $\widehat{G}(d) = m^{-1} \sum_{j=1}^m I_{\Delta^d xj} = m^{-1} \sum_{j=1}^m I_{yj}$ as in SP and define

$$a_n(d) = \frac{1}{m} \sum_{j=1}^m \left\{ -2\widehat{\mu} \mathrm{Re}\left[w_{yj} w^*_{\Delta^d vj}\right] + \widehat{\mu}^2 I_{\Delta^d vj} \right\}, \tag{A.16}$$

so that $G^{\diamond}(d) = \widehat{G}(d) + a_n(d)$. Then $\widetilde{G}_0(d)$, $\widetilde{G}_1(d)$, and $\widetilde{G}_2(d)$ defined on page 1913 of SP have additional terms $(2\pi/n)^{-2\theta} a_n(d)$, $(2\pi/n)^{-2\theta} \partial a_n(d)/\partial d$, and $(2\pi/n)^{-2\theta} \partial^2 a_n(d)/\partial d^2$, respectively. In view of the results in SP (pp. 1915–1916) leading to their equation (60), $R^{\diamond}(\overline{d})'' \to_p 4$ holds if we show that these three terms are all $o_p((\log n)^{-2})$ uniformly in $d \in M$. First, $\sup_M |(2\pi/n)^{-2\theta} a_n(d)| = o_p((\log n)^{-10})$ follows from (A.3), (A.4), and $\sup_{\theta \in M} m^{2|\theta|} < \infty$. For $(2\pi/n)^{-2\theta} \partial a_n(d)/\partial d$, note that

$$\frac{\partial}{\partial d} a_n(d) = \frac{1}{m} \sum_{j=1}^m \left\{ -2\widehat{\mu} \frac{\partial}{\partial d} \mathrm{Re}\left[w_{yj} w^*_{\Delta^d vj}\right] + \widehat{\mu}^2 \frac{\partial}{\partial d} I_{\Delta^d vj} \right\}. \tag{A.17}$$

From Lemma B.2(i) and (ii), the order of $\partial w_{\Delta^d vj}/\partial d = -w_{\log(1-L)\Delta^d vj}$ is no larger than $\log n$ times the order of $w_{\Delta^d vj}$. Furthermore, from Lemma 5.9(a) of SP, the order of $\partial w_{yj}/\partial d$ is no larger than $(\log n)^2$ times the order of w_{yj}. Therefore, the order of $(2\pi/n)^{-2\theta} \partial a_n(d)/\partial d$ is no larger than $(\log n)^2$ times that of $(2\pi/n)^{-2\theta} a_n(d)$. Similarly, the order of $(2\pi/n)^{-2\theta} \partial^2 a_n(d)/\partial d^2$ is no larger than $(\log n)^4$ times that of $(2\pi/n)^{-2\theta} a_n(d)$ in view of Lemma 5.9(c) of SP and Lemma B.2(iii). Therefore, the three additional terms are all $o_p((\log n)^{-2})$ uniformly in $d \in M$, and we establish $R^{\diamond\prime\prime}(\overline{d}) \to_p 4$.

The proof is completed by showing $m^{1/2} R^{\diamond\prime}(d_0) \to_d N(0,4)$. Because $G^{\diamond}(d) = \widehat{G}(d) + a_n(d)$, we have

$$m^{1/2} R^{\diamond\prime}(d_0) = m^{1/2} \left[\frac{\partial G^{\diamond}(d)/\partial d\big|_{d_0}}{G^{\diamond}(d_0)} - 2\frac{1}{m} \sum_{j=1}^m \log \lambda_j \right],$$

$$= m^{1/2} \left[\frac{\partial G(d)/\partial d\big|_{d_0} + \partial a_n(d)/\partial d\big|_{d_0}}{\widehat{G}(d_0) + a_n(d_0)} - 2\frac{1}{m} \sum_{j=1}^m \log \lambda_j \right].$$

Because SP shows $\widehat{G}(d_0) \to_p G_0$, $m^{1/2}\{\partial G(d)/\partial d|_{d_0} / \widehat{G}(d_0) - 2m^{-1} \sum_{j=1}^m \log \lambda_j\} \to_d N(0,4)$, and $m^{-1} \sum_{j=1}^m \log \lambda_j = O(\log n)$, the required result follows if

$$a_n(d_0) = o_p(m^{-1/2} (\log n)^{-1}), \qquad \partial a_n(d)/\partial d\big|_{d_0} = o_p(m^{-1/2}). \tag{A.18}$$

Note that $a_n(d_0) = m^{-1}\sum_{j=1}^{m}\{-2\widehat{\mu}\mathrm{Re}[w_{uj}w^*_{\Delta^{d_0}vj}] + \widehat{\mu}^2 I_{\Delta^{d0}vj}\}$ and from (A.6) we can write

$$\widehat{\mu}\cdot w_{\Delta^{d0}vj} = \xi_n \cdot O(j^{-\alpha}), \tag{A.19}$$

with $\alpha > \frac{1}{4}$. Using (A.19) and $w_{uj} = C(e^{i\lambda_j})w_{\varepsilon j} + r_{nj}$ with $\mathrm{E}|r_{nj}|^2 = O(j^{-1}\log n)$ uniformly in $j = 1,\ldots,m$ (Robinson, 1995b), we have

$$\begin{aligned}\left|\frac{a_n(d_0)}{2}\right| &\le \left|\widehat{\mu}\frac{1}{m}\sum_{j=1}^{m} C(e^{i\lambda_j})w_{\varepsilon j}w^*_{\Delta^{d_0}vj}\right| + \left|\widehat{\mu}\frac{1}{m}\sum_{j=1}^{m} r_{nj}w^*_{\Delta^{d_0}vj}\right| + \widehat{\mu}^2\frac{1}{m}\sum_{j=1}^{m} I_{\Delta^{d0}vj}\\ &= O_p\left(\left(\frac{1}{m^2}\sum_{j=1}^{m} j^{-2\alpha}\right)^{1/2} + \frac{1}{m}\sum_{j=1}^{m} j^{-\alpha-1/2}\log n + \frac{1}{m}\sum_{j=1}^{m} j^{-2\alpha}\right)\\ &= O_p((m^{-\alpha-1/2} + m^{-2\alpha} + m^{-1})\log m\log n) = o_p(m^{-1/2}(\log n)^{-1}),\end{aligned}$$

where the last equality follows from $\alpha > \frac{1}{4}$. For $\partial a_n(d)/\partial d|_{d_0}$, we have

$$\begin{aligned}\left.\frac{\partial}{\partial d}a_n(d)\right|_{d_0} &= 2\widehat{\mu}\frac{1}{m}\sum_{j=1}^{m}\mathrm{Re}\left[w_{\log(1-L)uj}w^*_{\Delta^{d_0}vj}\right]\\ &\quad + 2\widehat{\mu}\frac{1}{m}\sum_{j=1}^{m}\mathrm{Re}\left[w_{uj}w^*_{\log(1-L)\Delta^{d_0}vj}\right] + \widehat{\mu}^2\frac{1}{m}\sum_{j=1}^{m}\left.\frac{\partial}{\partial d}I_{\Delta^d vj}\right|_{d_0}.\end{aligned}$$

It follows easily from Lemma B.2 that the orders of $w^*_{\log(1-L)\Delta^{d_0}vj}$ and $\partial I_{\Delta^d vj}/\partial d|_{d_0}$ are $\log n$ times the orders of $w^*_{\Delta^{d_0}vj}$ and $I_{\Delta^{d_0}vj}$, respectively. Therefore, the second and third terms on the right of the preceding equation are $o_p(m^{-1/2})$ in view of the order of $a_n(d_0)$. For the first term on the right, Lemma 5.9(a) of SP shows that $w_{\log(1-L)uj} = -J(e^{i\lambda_j})w_{uj} + R_{nj}$ with $J(e^{i\lambda_j}) = O(\log n)$ and $\mathrm{E}|R_{nj}|^2 = O(j^{-1}(\log n)^4)$ uniformly in $j = 1,\ldots,m$. Therefore, it follows from a similar argument as before that the first term on the right is $o_p(m^{-1/2})$; thus $\partial a_n(d)/\partial d|_{d_0} = o_p(m^{-1/2})$, and we complete the proof. ■

A.3. Proof of Theorems 2a and 2b. From (A.5) and the fact that $d \ge 0$, we have $\lambda_j^{-\theta}w_{\Delta^d vj} = O(n^{1/2-d_0}j^{d_0-1})$. Combining it with $\mathrm{E}\widehat{\mu}^2 = \mathrm{E}|u_1|^2 < \infty$, we have, in place of (A.6),

$$\widehat{\mu}\cdot\lambda_j^{-\theta}w_{\Delta^d vj} = \xi_n\cdot O(n^{1/2-d_0}j^{d_0-1}), \tag{A.20}$$

uniformly in d. If $d_0 \ge \frac{1}{2}$, then $\widehat{\mu}\lambda_j^{-\theta}w_{\Delta^d vj} = \xi_n\cdot O((j/n)^{d_0-1/2}j^{-1/2}) = \xi_n\cdot O(j^{-1/2})$, whose order is no larger than that of $\lambda_j^{-\theta}(2\pi n)^{-1/2}\widetilde{U}_{\lambda_j n}(\theta)$ in the equation between (20) and (21) of SP and that of $\overline{U}_{nj}(\theta)$ in (30) and (39) of SP. Therefore, if we replace w_{yj} in SP with $w_{yj} - \widehat{\mu}\lambda w_{\Delta^d vj}$, the proof of the consistency of SP carries through. For the asymptotic normality for $d_0 \ge \frac{1}{2}$, we can use the proof of Theorem 1b without changes, because $O(j^{-1/2})$ is no larger than the maximum of the right-hand side of (A.6).

To show consistency for $d_0 \in \left(0, \frac{1}{2}\right)$, we need to modify the proof of Theorem 1a. Split Θ_1 into two, $\Theta_1^a = \Theta_1 \cap \{d : |\theta| \le \eta\}$ and $\Theta_1^b = \Theta_1 \backslash \Theta_1^a$, where η is the constant specified in the statement of Theorem 2a. Then consistency of $\widehat{d}$ follows if we show

$$\Pr(\inf_{\Theta_1^a} S^{\diamond}(d) \le 0) + \Pr(\inf_{\Theta_1^b \cup \Theta_2} S^{\diamond}(d) \le 0) \to 0, \quad \text{as } n \to \infty. \tag{A.21}$$

For the set Θ_1^a, we can strengthen the bound in (A.7) to

$$\sup_{\theta \in \Theta_1^a} |m^{-1} \sum_{j=1}^{m} (j/m)^{2\theta} j^{\alpha}| = O(m^{\alpha} \log m + m^{-1+2\eta} \log m) \tag{A.22}$$

uniformly in $\alpha \in [-C, C]$. Then, it follows from (A.20), (A.22), and $d_0 < \frac{1}{2}$ that

$$\sup_{\theta \in \Theta_1^a} \left| \widehat{\mu}^2 m^{-1} \sum_{j=1}^{m} (j/m)^{2\theta} \lambda_j^{-2\theta} I_{\Delta^d vj} \right| = O_p(n^{1-2d_0} m^{-1+2\eta} \log m). \tag{A.23}$$

Therefore, the first probability in (A.21) tends to 0 by applying the argument of the proof of Theorem 1a for Θ_1.

The second probability of (A.21) tends to 0 if there exists $\delta > 0$ such that

$$\Pr\left(\inf_{\Theta_1^b \cup \Theta_2} \left(m^{-1} {\sum}' (j/p)^{2\theta} (\lambda_j^{-2\theta} |w_{yj} - \widehat{\mu} w_{\Delta^d vj}|^2 - G_0) \right) \le -3\delta G_0 \right) \to 0. \tag{A.24}$$

This is because the algebra on pages 1904–1905 of SP leading to (A.8) remains unchanged even if Θ_2 is replaced with $\Theta_1^b \cup \Theta_2$, and we can replace the equation between (22) and (23) in SP with $\inf_{\Theta_1^b \cup \Theta_2} G_0 (m^{-1} \sum' (j/p)^{2\theta} - 1) > 4\delta G_0$ using Lemma B.5.

We proceed to show (A.24) for subsets of $\Theta_1^b \cup \Theta_2$. First, note that it follows from (A.20) and Lemma 5.4 of SP that

$$\sup_{\Theta_1^b \cup \Theta_2} \left| \widehat{\mu}^2 m^{-1} {\sum}' (j/p)^{2\theta} \lambda_j^{-2\theta} I_{\Delta^d vj} \right| = O_p(n^{1-2d_0} m^{2d_0 - 2}) = O_p(m^{-2\eta}). \tag{A.25}$$

Consequently, we can show that (A.24) holds for Θ_2 by applying the proof of Theorem 1a for Θ_2.

It remains to show (A.24) for Θ_1^b. Write

$$m^{-1} {\sum}' (j/p)^{2\theta} (\lambda_j^{-2\theta} |w_{yj} - \widehat{\mu} w_{\Delta^d vj}|^2 - G_0) = L_{1n}(d) + L_{2n}(d) + L_{3n}(d), \tag{A.26}$$

where $L_{1n}(d) = m^{-1} \sum' (j/p)^{2\theta} (\lambda_j^{-2\theta} I_{yj} - G_0)$, $L_{2n}(d) = -2\widehat{\mu} m^{-1} \sum' (j/p)^{2\theta} \lambda_j^{-2\theta}$ $\mathrm{Re}[w_{yj} w_{\Delta^d vj}^*]$, and $L_{3n}(d) = \widehat{\mu}^2 m^{-1} \sum' (j/p)^{2\theta} \lambda_j^{-2\theta} I_{\Delta^d vj}$. For $L_{1n}(d)$, we can apply the argument from line 7, page 1905, of SP without change to conclude that $\sup_{\Theta_1^b} |L_{1n}(d)| = o_p(1)$. We have $\sup_{\Theta_1^b} |L_{3n}(d)| = o_p(1)$ from (A.25), and the bound of $L_{2n}(d)$ follows from the bound of $L_{1n}(d)$, $L_{3n}(d)$, and Cauchy–Schwarz inequality. This completes the proof of consistency for $d_0 \in \left(0, \frac{1}{2}\right)$.

Proof of asymptotic normality for $d_0 \in \left(0, \frac{1}{2}\right)$ follows the proof of Theorem 1b. We use the bound (A.23) in place of (A.4) to show $R^{\diamond\prime\prime}(\overline{d}) \to_p 4$. To show $m^{1/2} R^{\diamond\prime}(d_0) \to_d N(0, 4)$, we simply repeat the proof of Theorem 1b replacing (A.6) with (A.20); then the stated result follows by $n^{1-2d_0} m^{-1} = o(m^{-1/2} (\log n)^{-1})$. ■

A.4. Proof of Theorem 3. From the standard proof of the two-step estimator, the stated result follows if (a) $R''_F(d) \to_p 4$ for any d such that $|d - d_0| \le |\widehat{d}_T - d_0|$ and (b) $m^{1/2} R'_F(d_0) \to_d N(0,4)$. Define $M = \{d : |d - d_0| \le (\log n)^{-4}\}$; then $\Pr(\widehat{d}_T \notin M) \to 0$ from Proposition B.1.

We proceed to analyze the limit of $R'_F(d_0)$ and $R''_F(\bar{d})$, where $\bar{d} \in M$. First, observe that $w_{\Delta^d(x - \widetilde{\mu}(d))j} = w_{\Delta^d xj} - \widetilde{\mu}(d) w_{\Delta^d vj}$; hence

$$
\begin{aligned}
&(\partial/\partial d) w_{\Delta^d(x-\widetilde{\mu}(d))j} \\
&\quad = (\partial/\partial d) w_{\Delta^d xj} - \widetilde{\mu}(d)(\partial/\partial d) w_{\Delta^d vj} - [(\partial/\partial d)\widetilde{\mu}(d)] w_{\Delta^d vj} \\
&\quad = (\partial/\partial d) w_{\Delta^d xj} - \widetilde{\mu}(d)(\partial/\partial d) w_{\Delta^d vj} - (\partial/\partial d) w(d)(\overline{X} - X_1) w_{\Delta^d vj}. \qquad \textbf{(A.27)}
\end{aligned}
$$

The second term on the right of (A.27) does not affect the asymptotics of $R'_F(d_0)$ and $R''_F(\bar{d})$ because we simply need to replace $\widetilde{\mu}(d)$ with $\overline{X}$ $\left(\text{if } d_0 < \frac{1}{2}\right)$ or X_1 (if $d_0 \ge \frac{3}{4}$) or their linear combination (if $d_0 \in \left[\frac{1}{2}, \frac{3}{4}\right)$) and apply the proof of Theorems 1b (specifically, the argument following (A.17) and (A.18)) and 2b. The third term on the right of (A.27) is 0 for $d \notin \left(\frac{1}{2}, \frac{3}{4}\right)$ because $(\partial/\partial d) w(d) = 0$. For $d \in \left(\frac{1}{2}, \frac{3}{4}\right)$, the order of the third term on the right of (A.27) is bounded by that of the second term on the right of (A.27), because $\widetilde{\mu}(d)$ is a linear combination of $\overline{X}$ and X_1, $(\partial/\partial d) w(d)$ is uniformly bounded, and the order of $w_{\Delta^d vj}$ is bounded by that of $(\partial/\partial d) w_{\Delta^d vj} = w_{\log(1-L)\Delta^d vj}$ from Lemma B.2(i) and (ii). Therefore, this term does not affect the limit of $R'_F(d_0)$ and $R''_F(\bar{d})$ either. A similar argument applies to the second derivatives of $\widetilde{\mu}(d)$, and the proof of Theorems 1b and 2b carries through. Consequently,

$$
m^{1/2} R'_F(d_0) \to_d N(0,4), \qquad \sup_{d \in M} |R''_F(d) - 4| \to_p 0, \qquad \textbf{(A.28)}
$$

and the required result follows. ■

A.5. Proof of Theorem 4. Because the tapered estimators are invariant to polynomial trends, it suffices to show that the limit of $R'_F(d_0)$ and $R''_F(\bar{d})$, where $\bar{d} \in M$, is not affected by detrending. In light of the proofs of Theorems 1b and 2b, recalling equation (A.19) from the former, this is the case if

$$
\lambda_j^{-\theta} w_{\Delta^d(\widehat{x} - \varphi(d))j} = \lambda_j^{-\theta} w_{\Delta^d x^0 j} + r_n, \qquad \textbf{(A.29)}
$$

where

$$
r_n = \xi_n \cdot O(j^{-\alpha}), \quad \text{where } O(j^{-\alpha}) \text{ is uniformly in } d \in M \text{ with } \alpha > 1/4, \qquad \textbf{(A.30)}
$$

and the derivatives of $w_{\Delta^d(\widehat{x} - \varphi(d))j}$ admit an analogous expression with the same order of the remainder term.

We proceed to show (A.29). To simplify the notation, let $\Xi_{\cdot n}$ denote $\Xi_{\cdot n}(d_0)$, suppressing their dependence on d_0. We give the proof only for $k = 2$. The proof for larger k follows the same argument, apart from more tedious algebra. A routine calculation gives

$$
\widehat{X}_t = X_t^0 - T_{kn} M_{kn}^{-1} X_{kn},
$$

where $T_{kn} = (1, (t/n), (t/n)^2)$ and

$$
M_{kn} = \begin{pmatrix} 1 & n^{-2}\sum_1^n t & n^{-3}\sum_1^n t^2 \\ n^{-2}\sum_1^n t & n^{-3}\sum_1^n t^2 & n^{-4}\sum_1^n t^3 \\ n^{-3}\sum_1^n t^2 & n^{-4}\sum_1^n t^3 & n^{-5}\sum_1^n t^4 \end{pmatrix}, \qquad X_{kn} = \begin{pmatrix} n^{-1}\sum_1^n X_t^0 \\ n^{-2}\sum_1^n t X_t^0 \\ n^{-3}\sum_1^n t^2 X_t^0 \end{pmatrix}.
$$

First we show $E||X_{kn}||^2 = O(n^{2d_0-1})$. Recall that $\sum_{t=1}^n X_t^0 = (1-L)^{-d_0-1}u_n \mathbf{1}\{t \geq 1\}$. Because $d_0 > -\frac{1}{2}$, clearly $E[n^{-1}\sum_1^n X_t^0]^2 = O(n^{2d_0-1})$. Summation by parts gives $\sum_{t=1}^n tX_t^0 = -\sum_{k=1}^{n-1}\sum_{t=1}^k X_t^0 + n\sum_{t=1}^n X_t^0$, and it follows that $E[\sum_{t=1}^n tX_t^0]^2 = O(n^{2d_0+3})$. Similarly, we can derive $E[\sum_{t=1}^n t^\alpha X_t^0]^2 = O(n^{2d_0+2\alpha+1})$ for any positive integer α, and $E||X_{kn}||^2 = O(n^{2d_0-1})$ follows.

Because M_n converges to a finite and invertible matrix, we can write $\widehat{X}_t$ as

$$\widehat{X}_t = X_t^0 + \Xi_{0n} + \Xi_{1n}t + \Xi_{2n}t^2,$$

where $E|\Xi_{0n}|^2 = O(n^{2d_0-1})$, $E|\Xi_{1n}|^2 = O(n^{2d_0-3})$, and $E|\Xi_{2n}|^2 = O(n^{2d_0-5})$. Taking the DFT of $\Delta^d(\widehat{X}_t - \varphi(d))$ and multiplying it by $\lambda_j^{-\theta}$, we obtain

$$\lambda_j^{-\theta} w_{\Delta^d(\widehat{x}-\varphi(d))j} = \lambda_j^{-\theta} w_{\Delta^d x^0 j} + [\Xi_{0n} - \varphi(d)]\lambda_j^{-\theta} w_{\Delta^d vj} + \Xi_{1n}\lambda_j^{-\theta} w_{\Delta^d tj} + \Xi_{2n}\lambda_j^{-\theta} w_{\Delta^d t^2 j}. \quad \textbf{(A.31)}$$

We proceed to check that the second to fourth terms on the right of (A.31) satisfy the condition (A.30). First, we derive the order of the second term. When $d_0 < \frac{1}{2}$, we have $\varphi(d) = 0$ for $d \in M$. Thus, in view of the order of $\lambda_j^{-\theta} w_{\Delta^d vj}$ given by (A.5), it follows that

$$[\Xi_{0n} - \varphi(d)]\lambda_j^{-\theta} w_{\Delta^d vj} = \Xi_{0n} O(n^{1/2-d_0}(j^{d_0-1} + j^{-1})) = \xi_n \cdot O(j^{d_0-1} + j^{-1}). \quad \textbf{(A.32)}$$

When $d_0 > \frac{3}{4}$, we have $\varphi(d) = \widehat{X}_1$ for $d \in M$, and hence $\Xi_{0n} - \varphi(d) = -X_1^0 - \Xi_{1n} - \Xi_{2n} = O_p(1 + n^{d_0-3/2})$. Thus, for $d \in M$,

$$[\Xi_{0n} - \varphi(d)]\lambda_j^{-\theta} w_{\Delta^d vj} = \xi_n \cdot O((1 + n^{d_0-3/2})n^{1/2-d_0} j^{d_0-1}) = \xi_n \cdot O(j^{-1/2} + j^{d_0-2}), \quad \textbf{(A.33)}$$

where the last equality follows because $d_0 > \frac{1}{2}$. When $d_0 \in \left(\frac{1}{2}, \frac{3}{4}\right)$, the order of the $O(\cdot)$ term is given by the sum of the preceding two, which is $O(j^{-\alpha})$ with $\alpha > \frac{1}{4}$ because $d_0 < \frac{7}{4}$. From (A.32) and (A.33), the second term on the right of (A.31) satisfies the condition (A.30).

Now we derive the order of the third and fourth terms on the right of (A.31). Observe that $t^\alpha = (1-L)^{-\alpha} v_t$ for any positive integer α. Hence $w_{\Delta^d tj} = w_{\Delta^{d-1} vj}$ and $w_{\Delta^d t^2 j} = w_{\Delta^{d-2} vj}$. Applying Lemma B.2(i) gives (recall that $d \leq 2 - \nu$)

$$w_{\Delta^d tj} = \begin{cases} O(n^{(3/2)-d} j^{d-2}), & d \geq 1+\nu, \\ -e^{i\lambda_j}(1-e^{i\lambda_j})^{-1}(2\pi n)^{-1/2}\Gamma(2-d)^{-1} n^{1-d}[1 + O(j^{-\nu})], & d \leq 1-\nu, \end{cases}$$

$$w_{\Delta^d t^2 j} = -e^{i\lambda_j}(1-e^{i\lambda_j})^{-1}(2\pi n)^{-1/2}\Gamma(3-d)^{-1} n^{2-d}[1 + O(j^{-\nu})].$$

Therefore, in view of the order of Ξ_{1n} and Ξ_{2n}, we obtain

$$\Xi_{1n}\lambda_j^{-\theta} w_{\Delta^d tj} = \xi_n \cdot O(j^{d_0-2}) \quad \text{if } d \geq 1+\nu, \qquad \xi_n \cdot O(j^{-\theta-1}) \quad \text{if } d \leq 1-\nu,$$

$$\Xi_{2n}\lambda_j^{-\theta} w_{\Delta^d t^2 j} = \xi_n \cdot O(j^{-\theta-1}).$$

Hence, the third and fourth terms on the right of (A.31) satisfy the condition (A.30). For the derivatives of $w_{\Delta^d(\widehat{x}-\varphi(d))j}$, we can obtain a similar approximation in which the remainder term is multiplied by $(\log n)^2$, which satisfies the condition (A.30). Therefore, the required result follows. ■

A.6. Proof of Theorem 5a. Take ν to be smaller than $2-\Delta_2>0$ without the loss of generality. We need to treat the cases for different values of d_0 and d separately. When $d_0 \in [\frac{1}{2},1)$, the required result follows from the proofs of Theorems 1a and 2a, because $\widehat{d}$ is consistent both under $\widehat{\mu}=X_1$ and $\widehat{\mu}=\overline{X}$. When $d_0<\frac{1}{2}$ and $d\in[\Delta_1,\frac{1}{2}]$, the proof of Theorem 1a applies because $\widetilde{\mu}(d)=\overline{X}$. When $d_0\geq 1$ and $d\in[\frac{3}{4},\Delta_2]$, the proof of Theorem 2a applies because $\widetilde{\mu}(d)=X_1$. It leaves us with the consideration of the following two cases:

$$\text{(i)}\quad d_0<\frac{1}{2}\quad\text{and}\quad d\in\left[\frac{1}{2},\Delta_2\right],\qquad\text{(ii)}\quad d_0\geq 1\quad\text{and}\quad d\in\left[\Delta_1,\frac{3}{4}\right]. \tag{A.34}$$

Note that (i) implies $\theta=d-d_0\geq\frac{1}{2}-d_0>0$ and (ii) implies $\theta\leq\frac{3}{4}-1\leq-\frac{1}{4}$. With a slight abuse of notation, define $\eta=\min\{\frac{1}{2}-d_0,\frac{1}{4}\}>0$ and define $\Theta_1^b=\Theta_1\cap\{|\theta|>\eta\}$ as in the proof of Theorem 2a for $d_0<\frac{1}{2}$. Because $\theta\in\Theta_1^b\cup\Theta_2$ if (i) or (ii) is true, $\Pr(\inf_{\Theta_1^b\cup\Theta_2}S_F(d)\leq 0)\to 0$ suffices for consistency of $\widehat{d}_F$.

Consider Θ_1^b first. We show $\Pr(\inf_{\Theta_1^b}S_F(d)\leq 0)\to 0$ by using the proof of Theorem 2a for Θ_1^b and showing that (A.24) holds for Θ_1^b if $|w_{yj}-\widehat{\mu}w_{\Delta^d vj}|^2$ in (A.24) is replaced by $|w_{yj}-\widetilde{\mu}(d)w_{\Delta^d vj}|^2$. We obtain a decomposition similar to (A.26) with $\widetilde{\mu}(d)$ replacing $\widehat{\mu}$ and $\sup_{\Theta_1^b}|L_{1n}(d)|=o_p(1)$. For $L_{2n}(d)$, it follows from equation (14) of SP, Lemma 5.2(b) of SP, and the equation between (20) and (21) of SP that $\lambda_j^{-\theta}w_{yj}=D_{nj}(\theta)w_{uj}+\overline{U}_{nj}(\theta)$, where $D_{nj}(\theta)$ and $\overline{U}_{nj}(\theta)$ satisfy the assumptions of Lemma B.4. Therefore, applying Lemma B.4 with $\alpha=d$ gives

$$L_{2n}(d)=\widetilde{\mu}(d)\cdot\begin{cases} n^{1/2-d_0}m^{d_0-1}\cdot O_p(n^{-1}m+m^{-\nu}\log n), & d\geq\nu,\\ n^{1/2-d_0}m^{-\theta-1}\cdot O_p(n^{-1}m+m^{-\nu}\log n), & d\leq-\nu.\end{cases}$$

Define $D^-=[\Delta_1,-\nu]$ and $D^+=[\nu,1-\nu]\cup[1+\nu,\Delta_2]$, so that $\Theta\subset D^-\cup D^+$. Applying Lemma B.3(i) to $L_{3n}(d)$ with $Q_2=Q_1=0$ and $Q_0=\widetilde{\mu}(d)$, we find that $L_{3n}(d)$ is bounded from below by, for some $\eta>0$,

$$\eta\widetilde{\mu}(d)^2n^{1-2d_0}m^{2d_0-2}\quad\text{for } d\in D^+,\qquad \eta\widetilde{\mu}(d)^2n^{1-2d_0}m^{-2\theta-2}\quad\text{for } d\in D^-.$$

Hence, $\Pr(\inf_{\Theta_1^a}[L_{2n}(d)+L_{3n}(d)]\leq\xi)\to 0$ for any $\xi>0$ from Lemma B.1, and (A.24) follows.

For $\Theta_2^3=\{-\frac{3}{2}\leq\theta\leq-\frac{1}{2}\}$, we show $\Pr(\inf_{\Theta_2^3}S_F(d)\leq 0)\to 0$ by using the argument of the proof of Theorem 1a in this Appendix and showing that (A.8) holds for Θ_2^3 if $|w_{yj}-\widehat{\mu}w_{\Delta^d vj}|^2$ in (A.8) is replaced by $|w_{yj}-\widetilde{\mu}(d)w_{\Delta^d vj}|^2$. The algebra leading to (A.10)–(A.12) still holds with $\widetilde{\mu}(d)$ in place of $\widehat{\mu}$. Thus (A.8) holds for Θ_2^3 if we can replace (A.13) with

$$\begin{aligned} m^{-1}&{\sum}'(j/p)^{2\theta}\lambda_j^{-2\theta}|(2\pi n)^{-1/2}e^{i\lambda_j}(1-e^{i\lambda_j})^{-1}Y_n(\theta)-\widetilde{\mu}(d)w_{\Delta^d vj}|^2\\ &\geq\begin{cases}\zeta m^{-2\theta-2}n^{2\theta+1}Y_n(\theta)^2+\zeta n^{1-2d+2\theta}m^{2d-2-2\theta}\widetilde{\mu}(d)^2, & d\in D^+,\\ \zeta m^{-2\theta-2}n^{2\theta+1}Y_n(\theta)^2+\zeta n^{1-2d+2\theta}m^{-2-2\theta}\widetilde{\mu}(d)^2, & d\in D^-\end{cases}\end{aligned} \tag{A.35}$$

and replace (A.14) with

$$
\begin{aligned}
&|m^{-1}\sum{}'(j/p)^{2\theta}\overline{U}_{nj}(\theta)\lambda_j^{-\theta}\widetilde{\mu}(d)w_{\Delta^d vj}| + |m^{-1}\sum{}'(j/p)^{2\theta}D_{nj}(\theta)^* w_{uj}^*\lambda_j^{-\theta}\widetilde{\mu}(d)w_{\Delta^d vj}| \\
&\quad = \begin{cases} n^{1/2-d+\theta}m^{d-1-\theta}\widetilde{\mu}(d)\cdot O_p(m^{-\nu}\log n + mn^{-1}), \ d \in D^+, \\ n^{1/2-d+\theta}m^{-1-\theta}\widetilde{\mu}(d)\cdot O_p(m^{-\nu}\log n + mn^{-1}), \ \ d \in D^-. \end{cases} \qquad \textbf{(A.36)}
\end{aligned}
$$

Because d is bounded away from 0, 1, and 2 by $\nu > 0$, applying Lemma B.3(ii) with $Q_2 = Y_n(\theta)$, $Q_1 = 0$, and $Q_0 = -\widetilde{\mu}(d)$ gives (A.35). Equation (A.36) follows from Lemma B.4.

For the other subsets of Θ_2, $\Pr(\inf_\theta S_F(d) \le 0) \to 0$ is shown by showing that (A.9) in the consistency proof of SP holds for those subsets if I_{yj} in (A.9) in SP is replaced with $|w_{yj} - \widetilde{\mu}(d)w_{\Delta^d vj}|^2$. For example, for $\Theta_2^5 = \{\frac{3}{2} \le \theta \le \frac{5}{2}\}$, the proof in SP begins from page 1910. If we replace $\lambda_j^{-\theta}w_{yj}$ in line 9, page 1910, of SP with $\lambda_j^{-\theta}w_{yj} - \lambda_j^{-\theta}\widetilde{\mu}(d)w_{\Delta^d vj}$, then, in place of (52) of SP, we have

$$
\begin{aligned}
&m^{-1}\sum{}'(j/p)^{2\theta}\lambda_j^{-2\theta}\Bigg|(2\pi n)^{-1/2}e^{i\lambda_j}(1-e^{i\lambda_j})\sum_1^n Z_t(\theta) \\
&\quad + (2\pi n)^{-1/2}e^{i\lambda_j}Z_n(\theta) + \widetilde{\mu}(d)w_{\Delta^d vj}\Bigg|^2. \qquad \textbf{(A.37)}
\end{aligned}
$$

Applying Lemma B.3(ii) with $Q_2 = \sum_1^n Z_t(\theta)$, $Q_1 = Z_n(\theta)$, and $Q_0 = -\widetilde{\mu}(d)$ gives the lower bound of (A.37). The terms involving the cross products of w_{uj}, $\overline{U}_{nj}(\theta)$, and $\widetilde{\mu}(d)w_{\Delta^d vj}$ are dominated by (A.37) from Lemma B.4. For the other terms in $m^{-1}\sum'(j/p)^{2\theta}\lambda_j^{-2\theta}I_{yj}$, the result on pages 1910–1911 of SP holds without change, and (23) of SP holds. ■

A.7. Proof of Theorem 5b. Assume $\mu_0 = 0$ without loss of generality. Define $N_\varepsilon = \{d : |d - d_0| < \varepsilon\}$. From Theorem 5a, $\widehat{d}_F \in N_\varepsilon$ with probability approaching 1 for arbitrarily small $\varepsilon > 0$. As in the proof of Theorem 1b, define $M = \{d : (\log n)^4|d - d_0| < \rho\}$ for a fixed $\rho > 0$. Then, because $\widetilde{\mu}(d)$ is a weighted average of $\overline{X}$ and X, $\Pr(\widehat{d}_F \notin M) \to 0$ follows from the proof of Theorem 1b (the argument following equation (A.15)) if $d_0 < \frac{1}{2}$, from the proof of Theorem 2b (the argument following equation (A.20)) if $d_0 \ge \frac{3}{4}$, and from combining both if $d_0 \in [\frac{1}{2}, \frac{3}{4})$.

The asymptotic distribution of $\widehat{d}_F$ follows from $\sup_{d\in M}|R_F''(d) - 4| \to_p 0$ and $m^{1/2}R_F'(d_0) \to_d N(0, 4)$, which are shown in the proof of Theorem 3. ■

APPENDIX B: Technical Lemmas

Lemma B.1 is a generalized restatement of equation (40) in SP and is stated as a lemma because it is repeatedly used in the proofs. Lemma B.2 gives the approximation formula for the DFT of the deterministic process $v_t = \mathbf{1}\{t \ge 1\}$. Lemmas B.3 and B.5 extend Lemmas 5.10 and 5.5 of SP, respectively. Lemma B.3 is used in the proof of consistency for establishing the lower bound of the objective function when $|d - d_0| \ge \frac{1}{2}$. Lemma B.5 is used in the proof of Theorem 2a. Propositions B.1 and B.2 show the asymptotic properties of the tapered estimator under Type II processes.

LEMMA B.1. *Let $\theta \in \Theta$ be a parameter and assume $m \to \infty$ as $n \to \infty$. Suppose two random variables $A_n(\theta)$ and $B_n(\theta)$ satisfy (i) $A_n(\theta) \geq \eta X_n(\theta)^2$ uniformly in $\theta \in \Theta$ for some $\eta > 0$ and (ii) $B_n(\theta) = X_n(\theta)R_n(\theta)$, where $\sup_\theta |R_n(\theta)| = O_p(k_n)$ with $k_n^2 \log m \to 0$. Then, for any $\zeta > 0$,*

$$\Pr\left(\inf_{\theta\in\Theta}[A_n(\theta)+B_n(\theta)] \leq -\zeta\right) \to 0 \quad \text{as } n \to \infty.$$

Proof. This result is a generalized restatement of equation (40) in SP. The terms $A_n(\theta)$ and $X_n(\theta)$ correspond to (35) and $m^{-\theta}n^{\theta-1/2}Z_n(\theta)$ in SP, respectively, and $B_n(\theta)$ corresponds to (37) + (38) in SP. The proof follows from repeating the argument on pages 1908–1909 of SP. ■

LEMMA B.2. *Let $v_t = \mathbf{1}\{t \geq 1\}$. Then the following result holds uniformly in $j = 1, \ldots, m$ with $m = o(n)$ and in d:*

(*i*) $w_{\Delta^d vj}$

$$= \begin{cases} e^{i\lambda_j}(1-e^{i\lambda_j})^{-1}(2\pi n)^{-1/2}[(1-e^{i\lambda_j})^d - n^{-d}/\Gamma(1-d) \\ \quad + O(n^{-d}j^{-1})], \qquad d \in \left[-\frac{1}{2}, C\right], \\ -e^{i\lambda_j}(1-e^{i\lambda_j})^{-1}(2\pi n)^{-1/2}\Gamma(1-d)^{-1}n^{-d}[1+O(j^{-1/2})], \\ \quad d \in [-C, -\frac{1}{2}], \end{cases}$$

(*ii*) $-w_{\log(1-L)\Delta^d vj}$

$$= \begin{cases} J_n(e^{i\lambda_j})w_{\Delta^d vj} + O(j^{-1}n^{1/2-d}\log n), & d \in [-C,1], \\ J_n(e^{i\lambda_j})w_{\Delta^d vj} + O(n^{1/2-d}\log n), & d \in [1,2], \end{cases}$$

(*iii*) $w_{(\log(1-L))^2\Delta^d vj}$

$$= \begin{cases} J_n(e^{i\lambda_j})^2 w_{\Delta^d vj} + O(j^{-1}n^{1/2-d}(\log n)^2), & d \in [-C,1], \\ J_n(e^{i\lambda_j})^2 w_{\Delta^d vj} + O(n^{1/2-d}(\log n)^2), & d \in [1,2], \end{cases}$$

where $J_n(e^{i\lambda_j}) = \sum_{k=1}^n k^{-1}e^{i\lambda_j} = O(\log n)$.

Proof. For part (i), first, from Lemma 5.1(b) of SP, we have

$$w_{\Delta^d vj} = (1-e^{i\lambda_j})^{-1}\left[w_{\Delta^{d+1}vj} - e^{i\lambda_j}(2\pi n)^{-1/2}\Delta^d v_n\right]. \tag{B.1}$$

From the proof of Lemma A.7 of Phillips and Shimotsu (2004, line 10, p. 676), we have

$$\Delta^{d+1}v_t = (1-L)^{d+1}v_t = (-d)_{t-1}/(t-1)!. \tag{B.2}$$

Therefore, the first term in the brackets in (B.1) can be expressed as

$$w_{\Delta^{d+1}vj} = \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^n \frac{(-d)_{t-1}}{(t-1)!}e^{it\lambda_j} = \frac{e^{i\lambda_j}}{\sqrt{2\pi n}}\sum_{k=0}^{n-1}\frac{(-d)_k}{k!}e^{ik\lambda_j}.$$

Define $D_n(e^{i\lambda}; d) = \sum_{k=0}^{n}[(-d)_k/k!]e^{ik\lambda}$ as defined in Lemma 5.1 of SP; then it follows from (B.1) that

$$w_{\Delta^d vj} = \frac{e^{i\lambda_j}}{1-e^{i\lambda_j}}(2\pi n)^{-1/2}\big[D_{n-1}(e^{i\lambda_j}; d) - \Delta^d v_n\big].$$

The stated result for $d \in \big[-\frac{1}{2}, C\big]$ follows from the approximation of $D_n(e^{i\lambda_j}; d)$ shown by Lemma A.2 of Phillips and Shimotsu (2004) and the fact that (see (B.2))

$$\Delta^d v_n = \frac{(1-d)_{n-1}}{(n-1)!} = \frac{\Gamma(n-d)}{\Gamma(n)\Gamma(1-d)} = \frac{1}{\Gamma(1-d)} n^{-d}(1+O(n^{-1})). \tag{B.3}$$

For $d \in \big[-\frac{3}{2}, -\frac{1}{2}\big]$, it follows from (B.1), the result for $d \in \big[-\frac{1}{2}, C\big]$, and (B.3) that

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

The results for smaller d follow from (B.1) and induction.

For part (ii), first we find a uniform bound for $d \in [-C, 1]$. Define $J_n(L) = \sum_{k=1}^{n}(1/k)L^k$. Lemma 5.7(a) of SP gives

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

where $\widetilde{J}_{n\lambda_j}(e^{-i\lambda}L) = \sum_{p=0}^{n-1}\widetilde{j}_{\lambda p}e^{-ip\lambda}L^p$ and $\widetilde{j}_{\lambda p} = \sum_{p+1}^{n}(1/k)e^{ik\lambda}$. Taking its DFT leaves us with

$$-w_{\log(1-L)\Delta^d vj} = J_n(e^{i\lambda_j})w_{\Delta^d vj} - (2\pi n)^{-1/2}\widetilde{J}_{n\lambda_j}(e^{-i\lambda_j}L)\Delta^d v_n. \tag{B.4}$$

Define $|x|_+ = \max\{x, 1\}$. Because $\Delta^d v_{n-p} = O((n-p)^{-d})$ from (B.3) and $\widetilde{j}_{\lambda p} = O(|p|_+^{-1}nj^{-1})$ from Lemma 5.8(b) of SP, the second term on the right of (B.4) is

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

Uniformly in $d \in [-C, 1]$, we have

$$\begin{aligned} \sum_{p=0}^{n-1}|p|_+^{-1}(n-p)^{-d} &\le \sum_{p=0}^{n/2}|p|_+^{-1}(n-p)^{-d} + \sum_{p=n/2}^{n-1}|p|_+^{-1}(n-p)^{-d} \\ &\le Cn^{-d}\sum_{p=0}^{n/2}|p|_+^{-1} + (n/2)^{-1}\sum_{p=n/2}^{n-1}(n-p)^{-d} \\ &= O(n^{-d}\log n). \end{aligned} \tag{B.5}$$

Therefore, the second term on the right of (B.4) is $O(j^{-1}n^{1/2-d}\log n)$, and the stated result follows. The order of $J_n(e^{i\lambda_j})$ is shown in Lemma 5.8(a) of SP.

For $d \in [1,2]$, Lemma 5.1(b) of SP gives

$$-w_{\Delta^d vj} = -(1-e^{i\lambda_j})w_{\Delta^{d-1}vj} - e^{i\lambda_j}(2\pi n)^{-1/2}\Delta^{d-1}v_n. \tag{B.6}$$

Differentiating it with respect to d, we find

$$-w_{\log(1-L)\Delta^d vj} = -(1-e^{i\lambda_j})w_{\log(1-L)\Delta^{d-1}vj} - e^{i\lambda_j}(2\pi n)^{-1/2}\log(1-L)\Delta^{d-1}v_n. \tag{B.7}$$

From (B.3), (B.5), and the fact that $d-1 \le 1$, the second term on the right of (B.7) is bounded by

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

Substituting the result for $d \in [-C,1]$ to $w_{\log(1-L)\Delta^{d-1}vj}$ on the right of (B.7) and then applying (B.6) gives the stated result.

For part (iii), first, for $d \in [-C,1]$, we have from Lemma 5.7(a) of SP

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

It follows that

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

The second term is $J_n(e^{i\lambda})$ times the second term on the right of (B.4); hence it is $O(j^{-1}n^{1/2-d}(\log n)^2)$. The third term is

$$\begin{aligned} &-(2\pi n)^{-1/2}\sum_{q=1}^{n} q^{-1}\sum_{p=0}^{n-q-1}\widetilde{j}_{\lambda_j p}e^{-ip\lambda_j}\Delta^d v_{n-p-q} \\ &\quad = O\left(j^{-1}n^{1/2}\sum_{q=1}^{n} q^{-1}\sum_{p=0}^{n-q-1}|p|_+^{-1}(n-q-p)^{-d}\right) \\ &\quad = O\left(j^{-1}n^{1/2}\sum_{q=1}^{n} q^{-1}(n-q)^{-d}\log n\right) = O(j^{-1}n^{1/2-d}(\log n)^2). \end{aligned} \tag{B.8}$$

For $d \in [1,2]$, taking the second derivative of $(-(\text{B.6}))$ with respect to d gives

$$w_{(\log(1-L))^2\Delta^d vj} = (1-e^{i\lambda_j})w_{(\log(1-L))^2\Delta^{d-1}vj} - e^{i\lambda_j}(2\pi n)^{-1/2}(\log(1-L))^2\Delta^{d-1}v_n. \tag{B.9}$$

The second term on the right of (B.9) is

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

and the first term on the right of (B.9) is, from the result for $d \in [-C, 1]$,

$$J_n(e^{i\lambda})^2(1-e^{i\lambda_j})w_{\Delta^{d-1}vj} + O(n^{1/2-d}(\log n)^2) = J_n(e^{i\lambda})^2 w_{\Delta^d vj} + O(n^{1/2-d}(\log n)^2),$$

giving the stated result. ■

LEMMA B.3. *Let $Q_k, k = 0,1,2$, be any real numbers, $v_t = \mathbf{1}\{t \geq 1\}$, $\kappa \in (0, \frac{1}{8})$, and $m = o(n)$. Then, there exists $\eta > 0$ not depending on Q_k such that, uniformly in $d \in \left\{\left[-\frac{1}{2}, -\varepsilon\right] \cup [\varepsilon, 1-\varepsilon] \cup [1+\varepsilon, 2-\varepsilon]\right\}$ and for sufficiently large n,*

$$(i)\ m^{-1}\sum_{j=[\kappa m]}^{m} |(2\pi n)^{-1/2}e^{i\lambda_j}(1-e^{i\lambda_j})Q_2 + (2\pi n)^{-1/2}e^{i\lambda_j}Q_1 + w_{\Delta^d vj}Q_0|^2$$
$$\geq \begin{cases} \eta(n^{-3}m^2Q_2^2 + n^{-1}Q_1^2 + n^{1-2d}m^{2d-2}Q_0^2), \\ \qquad d \in \{[\varepsilon, 1-\varepsilon] \cup [1+\varepsilon, 2-\varepsilon]\}, \\ \eta(n^{-3}m^2Q_2^2 + n^{-1}Q_1^2 + n^{1-2d}m^{-2}Q_0^2), \\ \qquad d \in \left\{\left[-\frac{1}{2}, -\varepsilon\right]\right\}. \end{cases}$$

$$(ii)\ m^{-1}\sum_{j=[\kappa m]}^{m} |(2\pi n)^{-1/2}e^{i\lambda_j}(1-e^{i\lambda_j})^{-1}Q_2$$
$$+ (2\pi n)^{-1/2}e^{i\lambda_j}(1-e^{i\lambda_j})^{-2}Q_1 + w_{\Delta^d vj}Q_0|^2$$
$$\geq \begin{cases} \eta(nm^{-2}Q_2^2 + n^3m^{-4}Q_1^2 + n^{1-2d}m^{2d-2}Q_0^2), \\ \qquad d \in \{[\varepsilon, 1-\varepsilon] \cup [1+\varepsilon, 2-\varepsilon]\}, \\ \eta(nm^{-2}Q_2^2 + n^3m^{-4}Q_1^2 + n^{1-2d}m^{-2}Q_0^2), \\ \qquad d \in \left\{\left[-\frac{1}{2}, -\varepsilon\right]\right\}. \end{cases}$$

Proof. The proof follows the approach of the proof of Lemma 5.10 of SP. For part (i), first define

$$A(\lambda_j) = (1-e^{i\lambda_j})Q_2 + Q_1 + (2\pi n)^{1/2}e^{-i\lambda_j}w_{\Delta^d vj}Q_0,$$

so the right-hand side of (i) is $m^{-1}\sum_{j=[\kappa m]}^{m}(2\pi n)^{-1/2}e^{i\lambda_j}A(\lambda_j)$. Then part (i) for $d \geq \varepsilon$ follows if, for sufficiently large n,

$$m^{-1}\sum_{j=[\kappa m]}^{m} |A(\lambda_j)|^2 \geq \eta(n^{-2}m^2Q_2^2 + Q_1^2 + n^{2-2d}m^{2d-2}Q_0^2). \tag{B.10}$$

We consider the case with $d \in [1+\varepsilon, 2-\varepsilon]$ in detail. The other cases follow the same line of argument. Because $d \geq \varepsilon$ implies that $j^{-d} = o(1)$ as $m \to \infty$ uniformly in $j \geq [\kappa m]$, we can refine the approximation of $w_{\Delta^d v j}$ in Lemma B.2(i) as

$$\begin{aligned} w_{\Delta^d v j} &= e^{i\lambda_j}(2\pi n)^{-1/2}(1-e^{i\lambda_j})^{d-1}(1+o(1)) \\ &= e^{i\lambda_j}(2\pi n)^{-1/2}e^{-(\pi/2)(d-1)i}\lambda_j^{d-1}(1+o(1)), \end{aligned} \quad \textbf{(B.11)}$$

uniformly in $j=[\kappa m],\dots,m$, where the second equality follows from Lemma 5.2 of SP. Define $\widetilde{c}(d)=\cos(-\pi(d-1)/2)$ and $\widetilde{s}(d)=\sin(-\pi(d-1)/2)$; then it follows that

$$A(\lambda_j) = -i\lambda_j Q_2 + o(\lambda_j)Q_2 + Q_1 + \widetilde{c}(d)\lambda_j^{d-1}Q_0 + i\widetilde{s}(d)\lambda_j^{d-1}Q_0 + o(\lambda_j^{d-1})Q_0,$$

$$|A(\lambda_j)|^2 = \left[Q_1+\widetilde{c}(d)\lambda_j^{d-1}Q_0\right]^2 + \left[\lambda_j Q_2 - \widetilde{s}(d)\lambda_j^{d-1}Q_0\right]^2 + r_{nj},$$

where $r_{nj}=o(\lambda_j^2)Q_2^2+o(\lambda_j)Q_1Q_2+o(\lambda_j^d)Q_2Q_0+o(\lambda_j^{d-1})Q_1Q_0+o(\lambda_j^{2d-2})Q_0^2$. Note that $m^{-1}\sum_{j=[\kappa m]}^m r_{nj}=o(m^2n^{-2}Q_2^2+Q_1^2+m^{2d-2}n^{2-2d}Q_0^2)$. Therefore, (B.10) follows if we show that, either for $j=[\kappa m],\dots,[m/4]$ or $j=[3m/4],\dots,m$,

$$\left[Q_1+\widetilde{c}(d)\lambda_j^{d-1}Q_0\right]^2 \geq \eta\left[Q_1^2+\widetilde{c}(d)^2\lambda_{m/2}^{2d-2}Q_0^2\right], \quad \textbf{(B.12)}$$

and either for $j=[\kappa m],\dots,[m/4]$ or $j=[3m/4],\dots,m$,

$$\left[\lambda_j Q_2-\widetilde{s}(d)\lambda_j^{d-1}Q_0\right]^2 \geq \eta\left[\lambda_{m/2}^2 Q_2^2+\widetilde{s}(d)^2\lambda_{m/2}^{2d-2}Q_0^2\right]. \quad \textbf{(B.13)}$$

We proceed to show (B.12). Assume $\widetilde{c}(d)\geq 0$ without the loss of generality. When $\operatorname{sgn}(Q_1)=\operatorname{sgn}(Q_0)$, the result follows immediately, and so assume $Q_1<0$ and $Q_0>0$ without the loss of generality. Now suppose $Q_1+\widetilde{c}(d)\lambda_{m/2}^{d-1}Q_0\geq 0$ and consider $j\geq[3m/4]$. Because $d-1\geq\varepsilon$, we have $\lambda_{3m/4}^{d-1}=(3/2)^{d-1}\lambda_{m/2}^{d-1}\geq(1+2\xi)\lambda_{m/2}^{d-1}$ for some $\xi>0$ uniformly in d; thus $Q_1+\widetilde{c}(d)\lambda_{3m/4}^{d-1}Q_0\geq 2\xi\widetilde{c}(d)\lambda_{m/2}^{d-1}Q_0\geq -2\xi Q_1>0$. Because λ_j^{d-1} is an increasing function of j, we have, for $j=[3m/4],\dots,m$,

$$Q_1+\widetilde{c}(d)\lambda_j^{d-1}Q_0\geq\xi\left(-Q_1+\widetilde{c}(d)\lambda_{m/2}^{d-1}Q_0\right),$$

and (B.12) follows because both $-Q_1$ and $\widetilde{c}(d)\lambda_{m/2}^{d-1}Q_0$ are positive. Now suppose $Q_1+\widetilde{c}(d)\lambda_{m/2}^{d-1}Q_0<0$ and consider $j\leq[m/4]$. Then $\lambda_{m/4}^{d-1}=(1/2)^{d-1}\lambda_{m/2}^{d-1}\leq(1-2\xi)\lambda_{m/2}^{d-1}$ for some $\xi\in(0,1/4)$ uniformly in d, and it follows that $Q_1+\widetilde{c}(d)\lambda_{m/4}^{d-1}Q_0\leq Q_1+(1-2\xi)\widetilde{c}(d)\lambda_{m/2}^{d-1}Q_0\leq\xi(Q_1-\widetilde{c}(d)\lambda_{m/2}^{d-1}Q_0)<0$. Therefore, we have, for $j=[\kappa m],\dots,[m/4]$,

$$Q_1+\widetilde{c}(d)\lambda_j^{d-1}Q_0\leq\xi\left(Q_1-\widetilde{c}(d)\lambda_{m/2}^{d-1}Q_0\right),$$

and (B.12) follows. Because $Q_1+\widetilde{c}(d)\lambda_{m/2}^{d-1}Q_0$ can be only ≥ 0 or <0, we established (B.12). Expression (B.13) is obtained by writing down $[\lambda_j Q_2-\widetilde{s}(d)\lambda_j^{d-1}Q_0]^2=\lambda_j^2[Q_2-\widetilde{s}(d)\lambda_j^{d-2}Q_0]^2$ and proceeding in the same manner with $d-2\leq\varepsilon$.

The other cases in part (i) follow the same argument. The essential element is that there is sufficient variation in $Q_1+\widetilde{c}(d)\lambda_j^{d-1}Q_0$ and $Q_2-\widetilde{s}(d)\lambda_j^{d-2}Q_0$ as j changes, which is guaranteed by bounding away $|d-1|$ and $|d-2|$ from 0. Part (ii) follows from the same argument, because there is sufficient variation in λ_j^{-1}, λ_j^{-2}, and λ_j^{d-1} if $|d-2|\geq\varepsilon$, $|d-1|\geq\varepsilon$, and $|d|\geq\varepsilon$. ■

LEMMA B.4. *Suppose $D_{nj}(\theta)$ and $\overline{U}_{nj}(\theta)$ satisfy*

$$\begin{cases} D_{nj}(\theta) = e^{-(\pi/2)\theta i} + O(\lambda_j) + O(j^{-1/2}) & \text{uniformly in } \theta \in \left[-\frac{1}{2}, \frac{1}{2}\right], \\ \mathrm{E}\sup_{\theta \in \left[-\frac{1}{2}, \frac{1}{2}\right]} |\overline{U}_{nj}(\theta)|^2 = O(j^{-1}(\log n)^2). \end{cases} \tag{B.14}$$

Let $\kappa \in (0, 1/8)$ and $m = o(n)$. Then, uniformly in $\theta \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ and $\alpha \in [-C, -\varepsilon] \cup [\varepsilon, 1-\varepsilon] \cup [1+\varepsilon, 2-\varepsilon]$ with $\varepsilon < \frac{1}{2}$,

$$\left| \frac{1}{m} \sum_{j=[\kappa m]}^{m} \left(\frac{j}{m}\right)^{2\theta} D_{nj}(\theta) w_{uj} \lambda_j^{-\theta} w_{\Delta^\alpha vj} \right| + \left| \frac{1}{m} \sum_{j=[\kappa m]}^{m} \left(\frac{j}{m}\right)^{2\theta} \overline{U}_{nj}(\theta) \lambda_j^{-\theta} w_{\Delta^\alpha vj} \right|$$

$$= \begin{cases} n^{1/2-\alpha+\theta} m^{\alpha-1-\theta} \cdot O_p(n^{-1}m + m^{-\varepsilon} \log n), & \alpha \geq \varepsilon, \\ n^{1/2-\alpha+\theta} m^{-1-\theta} \cdot O_p(n^{-1}m + m^{-\varepsilon} \log n), & \alpha \leq -\varepsilon. \end{cases} \tag{B.15}$$

Proof. Define $A^+ = [\varepsilon, 1-\varepsilon] \cup [1+\varepsilon, 2-\varepsilon]$ and $A^- = [-C, -\varepsilon]$, so that $A^+ \cup A^-$ covers the admissible value of α. For the first term on the left of (B.15), from (B.14) and Lemma B.2, we obtain

$$D_{nj}(\theta) w_{\Delta^\alpha vj} = \begin{cases} C_n(\theta) n^{1/2-\alpha} j^{\alpha-1} [1 + O(\lambda_j) + O(j^{-\varepsilon})], & \alpha \in A^+, \\ C_n(\theta) n^{1/2-\alpha} j^{-1} [1 + O(\lambda_j) + O(j^{-\varepsilon})], & \alpha \in A^-, \end{cases} \tag{B.16}$$

where $C_n(\theta)$ is a nonrandom function of θ such that $0 < |C_n(\theta)| < \infty$ uniformly in θ. The required result follows from Lemmas 5.4 and 5.6 of SP, (B.16), and $\mathrm{E}|w_u(\lambda_j)|^2 < \infty$ (e.g., SP, eqn. (19)). For the second term on the left of (B.15), the stated bound follows straightforwardly from Lemma B.2, (B.14), and Lemma 5.4 of SP. ■

LEMMA B.5. *Let $\eta > 0$ be a fixed number and $p \sim m/e$ as $m \to \infty$. There exist $\varepsilon \in (0, 0.1)$ and $\bar{\kappa} \in \left(0, \frac{1}{4}\right)$ such that, for all fixed $\kappa \in (0, \bar{\kappa}]$ and sufficiently large m,*

$$\inf_{\gamma \in [-C, -\eta] \cup [\eta, C]} \frac{1}{m} \sum_{j=[\kappa m]}^{m} \left(\frac{j}{p}\right)^{\gamma} \geq 1 + 2\varepsilon.$$

Proof. Lemma 5.5 of SP establishes the stated result for $\gamma \in [-C, -1+2\Delta] \cup [1, C]$ with $\Delta \in (0, 1/(2e))$. Hence, we only need to show the stated result for $\gamma \in [-1+2\Delta, -\eta] \cup [\eta, 1]$. From Lemma 5.4 of SP, we have, uniformly in $\gamma \in [-1+2\Delta, 1]$,

$$\frac{1}{m} \sum_{j=[\kappa m]}^{m} \left(\frac{j}{p}\right)^{\gamma} = \left(\frac{m}{p}\right)^{\gamma} \frac{1}{m} \sum_{j=[\kappa m]}^{m} \left(\frac{j}{m}\right)^{\gamma} = (e + o(1))^{\gamma} \left(\int_{\kappa}^{1} x^{\gamma}\, dx + o(1) \right)$$

$$= (\gamma + 1)^{-1} e^{\gamma} (1 - \kappa^{\gamma}) + o(1), \quad \text{as } m \to \infty. \tag{B.17}$$

Note that $g(\gamma) = (\gamma + 1)^{-1} e^{\gamma}$ takes the value 1 when $\gamma = 0$, $g'(\gamma) > 0$ when $\gamma \geq \eta$, and $g'(\gamma) < 0$ when $\gamma \in [-1+2\Delta, -\eta]$. Therefore, choosing κ sufficiently small makes (B.17) larger than $1 + 2\varepsilon$ for $\gamma \in [-1+2\Delta, -\eta] \cup [\eta, 1]$ and sufficiently large m. ■

The rest of this Appendix discusses the asymptotic properties of the tapered estimator under Type II processes.

First we discuss the taper used by Vel. Let h_t denote a taper of order p generated by Kolmogorov's proposal. Then h_t satisfies the regularity conditions in Vel and in Robinson (2005), and the tapered estimator is invariant to a polynomial time trend of order $p-1$. Define the tapered DFT and periodogram as $w_{xp}^T(\lambda_j) = (2\pi n)^{-1/2}\sum_{t=1}^n h_t X_t e^{it\lambda_j}$ and $I_{xp}^T(\lambda_j) = |w_{xp}^T(\lambda_j)|^2$. As in Vel (p. 99), define the tapered local Whittle estimator as $\widehat{d}_p = \arg\min_{d\in[\Delta_1,\Delta_2]} R_p(d)$, where $-\frac{1}{2} < \Delta_1 < \Delta_2 < \infty$, $R_p(d) = \log\widehat{G}_p(d) - 2dpm^{-1}\sum_{j=p,2p,\ldots,m}\log\lambda_j$, and $\widehat{G}_p(d) = pm^{-1}\sum_{j=p,2p,\ldots,m}\lambda_j^{2d} I_{xp}^T(\lambda_j)$.

PROPOSITION B.1. *Suppose X_t is generated by (11) with $d_0 \in (\Delta_1,\Delta_2)$ and $\beta_{p0} = \cdots = \beta_{k0} = 0$. Suppose $p \geq \max\left\{\left[\Delta_2 + \frac{1}{2}\right] + 1, 2\right\}$ and Assumptions 1′–5′ and 7′ hold. Then $m^{1/2}(\widehat{d}_p - d_0) \to_d N(0, p\Phi/4)$ as $n \to \infty$, where Φ is defined in equation (10) in Vel (p. 101).*

Proof. Let s be the integer part of $d_0 + \frac{1}{2}$ and Y_t be a Type I $I(d_0)$ process with the sth-order polynomial time trend

$$Y_t = (1-L)^{-s}U_t^{(s)}\mathbf{1}\{t \geq 1\} + \mu_0^{(0)} + \mu_0^{(1)}t + \cdots + \mu_0^{(s)}t^s, \qquad U_t^{(s)} = (1-L)^{s-d_0}u_t,$$

where u_t satisfies Assumptions 1′–3′. The proof consists of two steps. First, we show that $\widehat{d}_p$ is consistent and has the stated limiting distribution if the objective function is constructed using Y_t. Second, we show that replacing Y_t with X_t in the objective function does not change the limiting behavior of $\widehat{d}_p$.

The first part is proved by checking that Y_t satisfies the assumptions of Theorems 5 and 6 of Vel. As discussed in Lobato and Velasco (2000, p. 414), the asymptotic normality of the tapered estimator still holds even if Assumption 8 of Vel is weakened to

$$f_{U(s)}(\lambda) - G\lambda^{-2(d-s)} = O(\lambda^{-2(d-s)+\beta}) \quad \text{for } \beta \in (0,2]. \tag{B.18}$$

The term $U_t^{(s)}$ satisfies (B.18) by Assumption 1′ and $|1-e^{i\lambda}|^{2s-2d_0} = \lambda^{-2(d_0-s)} + O(\lambda^{-2(d_0-s)+2})$. Therefore, it suffices to check that Y_t and $U_t^{(s)}$ satisfy Assumptions 5, 7, 9, and 10 of Vel.

For Assumption 5 of Vel, define $d(\lambda) = \sum_{k=0}^\infty d_k e^{ik\lambda} = (1-e^{i\lambda})^{-d_0}$ and $\alpha(\lambda) = \sum_{k=0}^\infty \alpha_k e^{ik\lambda}$; then we have $\alpha(\lambda) = c(\lambda)d(\lambda)$. Now $\partial\alpha(\lambda)/\partial\lambda = O(|\alpha(\lambda)|/\lambda)$ follows from Assumption 2′, and Assumption 5 is satisfied. Assumption 7 of Vel follows from (B.18). Assumption 9 is satisfied because $\partial|1-e^{i\lambda}|^{2s-2d_0}/\partial\lambda = O(\lambda^{-1-2(d_0-s)})$ and $\partial f_u(\lambda)/\partial\lambda = O(\lambda^{-1})$ from Assumption 2 and $f_u(\lambda) > 0$ for λ sufficiently small. For Assumption 10 of Vel, note that we can rewrite $U_t^{(s)}$ as a linear process as $U_t^{(s)} = \sum_{l=0}^\infty \alpha_l\varepsilon_{t-l}$; then Assumption 10 is satisfied because $\sum_{k=0}^\infty \alpha_k^2 = \mathrm{E}(U_t^{(s)})^2 < \infty$. Thus, we complete the first part of the proof.

Second, we use the results on the difference between $w_{xp}^T(\lambda_j)$ and $w_{yp}^T(\lambda_j)$ by Robinson (2005) to show that $\widehat{d}_p$ has the stated limit distribution when the objective function is constructed with X_t. Note that $(d+q,d)$ in Robinson (2005) corresponds to our $(d_0, d_0 - s)$ and the statement of the theorem in Robinson has a typo: $d \in \left(-\frac{1}{2},\frac{1}{2}\right]$ should be replaced with $d \in \left[-\frac{1}{2},\frac{1}{2}\right)$. From the theorem of Robinson (2005), we have

$$\mathrm{E}\left\{\lambda_j^{2d_0}\left|w_{xp}^T(\lambda_j) - w_{yp}^T(\lambda_j)\right|^2\right\} \leq Cj^{-2\eta-1}\log n,$$

where $2\eta = 2(d_0 - s) - 1 < 0$. It follows from the triangle inequality, Cauchy–Schwarz inequality, and $\mathrm{E}\lambda_j^{2d_0} I_{yp}^T(\lambda_j) = O(1)$ for $j = p, 2p, \ldots, m$ (Vel, Thm. 4) that

$$\begin{aligned}
&\mathrm{E}\lambda_j^{2d_0} \left| I_{xp}^T(\lambda_j) - I_{yp}^T(\lambda_j) \right| \\
&\quad \leq \mathrm{E}\lambda_j^{2d_0} \left| w_{xp}^T(\lambda_j) - w_{yp}^T(\lambda_j) \right|^2 + 2\mathrm{E}\lambda_j^{2d_0} \left| w_{xp}^T(\lambda_j) - w_{yp}^T(\lambda_j) \right| \left| w_{yp}^T(\lambda_j) \right| \\
&\quad = O(j^{-\eta - 1/2} \log n), \quad \text{for } j = p, 2p, \ldots, m.
\end{aligned}$$

Note that the periodogram I_j in Vel is equal to $I_{yp}(\lambda_j)$ in our notation. Therefore, if we replace I_j in $A_p(d)$ in line 3, page 112, of Vel with $I_{xp}^T(\lambda_j)$, then the right-hand side of (A14) in Vel has an additional term whose order is $O_p(m^{-\xi} \log m \log n)$ for some $\xi > 0$, and the proof of consistency is not affected.

For asymptotic normality (Vel, Thm. 6), if we replace I_j in (A23) on page 116 of Vel with $I_{xp}(\lambda_j)$, then the right-hand side of (A23) has an additional term $O_p(r^{-\eta+1/2} \log n)$. Consequently, the left-hand side of the equation in line 14, page 117 of Vel has an additional term $O_p(m^{-\eta-1/2} \log n)$. Because this is $o_p(1)$, the right-hand side of that equation remains unchanged, and their argument carries through. Finally, the equation in line 18, page 117, of Vel has an additional term $O_p(m^{-\eta} \log n)$, which is $o_p(1)$, and asymptotic normality follows. ■

The tapered estimator by HC takes the difference of the data and applies a complex-valued taper $h_t^{HC} = 0.5[1 - \exp(2\pi i(t - 1/2)/n)]$ to ΔX_t. The objective function is defined in terms of ΔX_t, and the estimator is defined as $\widehat{d}_{HC} = \arg\min_{d \in [\Delta_1', \Delta_2']} R_{HC}(d)$, where $-\frac{3}{2} < \Delta_1' < \Delta_2' < \frac{1}{2}$, $R_{HC}(d) = \log \widehat{G}_{HC}(d) - 2(d-1)m^{-1} \sum_{j=1}^m \log \lambda_{(j+1/2)}$, and $\widehat{G}_{HC}(d) = m^{-1} \sum_{j=1}^m \lambda_{(j+1/2)}^{2(d-1)} I_{\Delta x}^{HC}(\lambda_j)$. HC propose to use the powers of h_t^{HC} as a taper with the higher-order differences of X_t to allow for larger values of d. To save space, we restrict the range of d to be $\left(-\frac{1}{2}, \frac{3}{2}\right)$ and allow only a linear trend. In Proposition B.2, additional assumptions on $f_u(\lambda)$ are necessary to satisfy Assumption A1 in HC.

PROPOSITION B.2. *Suppose X_t is generated by (11) with $d_0 \in (\Delta_1', \Delta_2')$ and $\beta_{20} = \cdots = \beta_{k0} = 0$. Suppose Assumptions 1′–5′ and 7′ hold and $f_u(\lambda) = G_0 + E_\beta \lambda^\beta + o(\lambda^\beta)$ with $\beta \in (1, 2]$ and $E_\beta < \infty$. Then $m^{1/2}(\widehat{d}_{HC} - d_0) \to_d N(0, (1.5)/4)$ as $n \to \infty$.*

Proof. Applying the first part of the proof of Proposition B.1, we can easily show that our Assumptions 1–5 and 1′–5′ imply that m and $\Delta Y_t = (1-L)^{-d_0+1} u_t$ satisfy Assumptions A1′–A4′ of HC. Therefore, both consistency and asymptotic normality of $\widehat{d}_{HC}$ follow if we show

$$\mathrm{E}\left\{\lambda_j^{2(d_0-1)} |w_{\Delta y}^{HC}(\lambda_j) - w_{\Delta x}^{HC}(\lambda_j)|^2\right\} \leq Cj^{-2\eta-1} \log n, \qquad j = 1, \ldots, m \tag{B.19}$$

for some $\eta > 0$, because then the proof of Theorems 1 and 2 of HC carries through if we replace their I_j^T (that corresponds to our $I_{\Delta y}^{HC}(\lambda_j)$) with $I_{\Delta x}^{HC}(\lambda_j)$. Specifically, Lemma 1 and equation (8) of HC still hold, and Lemma 6 of HC has an additional $O_p(r^{-\eta+1/2})$ term that does not affect the validity of their Theorem 2.

We proceed to show (B.19). First, observe that the HC taper satisfies the bounds (2.1)–(2.3) on page 286 of Robinson (2005) with $p = 1$. The other two conditions on $h(t)$

on page 286 do not matter for the theorem of Robinson (2005) to hold. Therefore, for $d_0 \in \left[\frac{1}{2}, \frac{3}{2}\right)$, applying (2.6) in the theorem of Robinson (2005) to $(\Delta X_t, \Delta Y_t)$ gives

$$\mathrm{E}\{\lambda^{2(d_0-1)}|w_{\Delta y}^{HC}(\lambda) - w_{\Delta x}^{HC}(\lambda)|^2\} \leq C|\log\lambda|^{\mathbf{1}\{d_0=1/2\}}(n\lambda)^{2(d_0-1)-2}, \qquad 0 < \lambda \leq \pi.$$

Hence, (B.19) holds with $\eta = (3/2) - d_0 > 0$.

For $d_0 \in \left[-\frac{1}{2}, \frac{1}{2}\right)$, summation by parts gives

$$\sum_{t=1}^{n} h_t^{HC}(\Delta Y_t - \Delta X_t)e^{it\lambda} = \sum_{t=1}^{n-1}(h_t^{HC}e^{it\lambda} - h_{t+1}^{HC}e^{i(t+1)\lambda})(Y_t - X_t) + h_n^{HC}e^{in\lambda}(Y_n - X_n). \tag{B.20}$$

Because $h_{t+1}^{HC} = h_t^{HC}e^{2\pi i/n} + 0.5(1 - e^{2\pi i/n})$, routine algebra gives $h_t^{HC}e^{it\lambda} - h_{t+1}^{HC} e^{i(t+1)\lambda} = h_t^{HC}e^{it\lambda}(1 - e^{i(\lambda+2\pi/n)}) + 0.5(e^{2\pi i/n} - 1)e^{i(t+1)\lambda}$. It follows that $w_{\Delta y}^{HC}(\lambda) - w_{\Delta x}^{HC}(\lambda) = A_\lambda + B_\lambda + R_\lambda$, where

$$A_\lambda = \left(1 - e^{i(\lambda+2\pi/n)}\right)\left[w_y^{HC}(\lambda) - w_x^{HC}(\lambda)\right],$$

$$B_\lambda = 0.5(e^{2\pi i/n} - 1)e^{i\lambda}[w_y(\lambda) - w_x(\lambda)],$$

and $R_\lambda = (2\pi n)^{-1/2}h_n^{HC}e^{in\lambda}(Y_n - X_n)e^{i(\lambda+2\pi/n)} - (2\pi n)^{-1/2}0.5(e^{2\pi i/n} - 1)e^{i\lambda}e^{in\lambda} (Y_n - X_n)$. For A_{λ_j}, it follows from (2.7) in the theorem of Robinson (2005) and $\lambda_j^{-2}|1 - e^{i(\lambda_j+2\pi/n)}|^2 < C$ that, for $1/n \leq \lambda_j \leq \pi$,

$$\mathrm{E}\left\{\lambda_j^{2(d_0-1)}|A_{\lambda_j}|^2\right\} \leq C(n\lambda_j)^{2d_0-2}\log n \leq Cj^{-2(1/2-d_0)-1}\log n,$$

with $\frac{1}{2} - d_0 > 0$. For B_{λ_j}, using (2.6) of the theorem of Robinson (2005) and $e^{2\pi i/n} - 1 = O(n^{-1})$, we have $\mathrm{E}\{\lambda_j^{2(d_0-1)}|B_{\lambda_j}|^2\} \leq C(n\lambda_j)^{-3}\log n = Cj^{-3}\log n$ for $1/n \leq \lambda_j \leq \pi$. Finally, for R_{λ_j}, it follows from $|h_n^{HC}| \leq Cn^{-1}$ and $\mathrm{E}(Y_n - X_n)^2 = O(n^{2d_0-1})$ (Marinucci and Robinson, 1999, p. 119) that $\mathrm{E}\{\lambda_j^{2(d_0-1)}|R_{\lambda_j}|^2\} \leq C(n\lambda_j)^{2d_0-2}n^{-2} \leq Cj^{-2(1/2-d_0)-1}n^{-2}$ for $1/n \leq \lambda_j \leq \pi$. Therefore, (B.19) holds with $\frac{1}{2} - d_0 > 0$, and the proof is completed. ■