Econometric Theory, **23**, 2007, 899–929. Printed in the United States of America.
DOI: 10.1017/S0266466607070387

LOCAL WHITTLE ESTIMATION OF FRACTIONAL INTEGRATION FOR NONLINEAR PROCESSES

Xiaofeng Shao
University of Illinois at Urbana-Champaign

Wei Biao Wu
University of Chicago

We study asymptotic properties of the local Whittle estimator of the long memory parameter for a wide class of fractionally integrated nonlinear time series models. In particular, we solve the conjecture posed by Phillips and Shimotsu (2004, *Annals of Statistics* 32, 656–692) for Type I processes under our framework, which requires a global smoothness condition on the spectral density of the short memory component. The formulation allows the widely used fractional autoregressive integrated moving average (FARIMA) models with generalized autoregressive conditionally heteroskedastic (GARCH) innovations of various forms, and our asymptotic results provide a theoretical justification of the findings in simulations that the local Whittle estimator is robust to conditional heteroskedasticity. Additionally, our conditions are easily verifiable and are satisfied for many nonlinear time series models.

1. INTRODUCTION

Since the seminal work of Robinson (1995a, 1995b), semiparametric estimation of the long memory parameter of time series has been an active area of research. Let $d \in (-\frac{1}{2}, \frac{1}{2})$ be the order of integration. Consider the I(d) process $\{X_t\}$ defined by

$$(1 - B)^d (X_t - \mu) = u_t, \qquad t \in \mathbb{Z}, \tag{1}$$

where μ is an unknown mean, B is the backward shift operator, and $\{u_t\}_{t\in\mathbb{Z}}$ is a mean zero, covariance stationary, short memory process. Let $f_X(\cdot)$ and $f_u(\cdot)$ be the spectral density functions of $\{X_t\}$ and $\{u_t\}$, respectively. Then (1) implies

$$f_X(\lambda) = |1 - e^{i\lambda}|^{-2d} f_u(\lambda). \tag{2}$$

We thank Liudas Giraitis for providing the manuscript by Dalla, Giraitis, and Hidalgo (2006). We are grateful to the two referees and the editor for their detailed comments, which led to substantial improvements. We also thank Michael Stein for helpful comments on an earlier version. The work is supported in part by NSF grant DMS-0478704. Address correspondence to Xiaofeng Shao, Department of Statistics, University of Illinois, at Urbana-Champaign, 725 South Wright St., Champaign, IL, 61820, USA; e-mail: xshao@uiuc.edu.

The process $\{X_t\}$ has long memory if $d \in (0,\frac{1}{2})$, has short memory if $d = 0$, and is antipersistent if $d \in (-\frac{1}{2},0)$. A nonstationary process $\{X_t\}$ can be defined when $d \geq \frac{1}{2}$. For example, if $d \in [\frac{1}{2},\frac{3}{2})$, we can write X_t as the sum of an $\mathrm{I}(d-1)$ process, i.e.,

$$X_t = X_0 + \sum_{j=1}^{t} Y_j, \qquad (1-B)^{d-1} Y_t = u_t, \tag{3}$$

where X_0 is a random variable whose distribution does not depend on t. The case for $d \geq \frac{3}{2}$ can be similarly defined by the repeated use of partial summation (Velasco 1999a, 1999b). An alternative definition of an $\mathrm{I}(d)$ process is given by

$$(1-B)^d (X_t - X_0) = u_t \mathbf{1}(t \geq 1). \tag{4}$$

Equivalently, letting $\phi_k(d) = \prod_{j=1}^{k}((d+j-1)/j)$, $k \geq 1$, $\phi_0(d) = 1$, we have

$$X_t = X_0 + (1-B)^{-d} u_t \mathbf{1}(t \geq 1) = X_0 + \sum_{k=0}^{t-1} \phi_k(d) u_{t-k}. \tag{5}$$

Clearly (5) is well defined for any $d \in \mathbb{R}$. Under the formulation (4), the process $\{X_t\}$ is nonstationary when $d \geq \frac{1}{2}$ and only asymptotically stationary when $d \in (-\frac{1}{2},\frac{1}{2})$. The main distinction between Type I processes (expressions (1) and (3)) and Type II processes (expression (4)) lies in the presample treatment. These two definitions of $\mathrm{I}(d)$ processes could lead to different asymptotic behaviors for various statistics. For example, the normalized partial sum of X_t when $d > \frac{1}{2}$ converges to two distinct fractional Brownian motions (Marinucci and Robinson, 1999a, 1999b). A detailed discussion of their differences can be found in Robinson (2005) and Shimotsu and Phillips (SP hereafter) (2006).

Two popular semiparametric frequency domain approaches to estimate d have been extensively studied in the literature: log periodogram (LP) regression (Geweke and Porter-Hudak, 1983) and local Whittle (LW) estimation (Künsch, 1987). LP regression is easy to compute because it only involves least squares regression, but it is less efficient than LW estimation, which is constructed based on the likelihood principle (Robinson, 1995b). The asymptotic properties of LP regression have been investigated by Robinson (1995a) and Velasco (1999a, 2000) among others for Type I processes and by Kim and Phillips (1999) and Phillips (1999b) for Type II processes. Regarding the asymptotic theory of the LW estimator, Robinson (1995b) and Velasco (1999b) dealt with Type I processes. For Type II processes, see Phillips and Shimotsu (PS hereafter) (2004) and SP (2006). PS (2004) and SP (2006) nearly completed the study of asymptotic properties of the LW estimator for Type II processes in that they characterized the asymptotic distribution for $d \in [-1,-\frac{1}{2}) \cup (-\frac{1}{2},\frac{1}{2}) \cup (\frac{1}{2},M]$, where

M is a fixed constant. Their asymptotic analysis was facilitated by an exact representation of the Fourier transform of Type II fractional processes (Phillips, 1999a). PS (2004) conjectured that their asymptotic results are still valid for Type I processes with different constants in the asymptotic distributions. See the review paper by Moulines and Soulier (2003) for other semiparametric methods of estimating d in the frequency domain.

So far, it seems that most asymptotic analysis of the LW estimator has been limited to linear processes, i.e., $X_t = \mu + \sum_{k=0}^{\infty} a_k \zeta_{t-k}$ (cf. Robinson, 1995b; Velasco, 1999b) or $u_t = \sum_{k=0}^{\infty} a_k \zeta_{t-k}$ (cf. PS, 2004; SP, 2006), where the ζ_t are martingale differences with constant conditional variance, i.e.,

$$\sigma_t^2 := \mathbb{E}(\zeta_t^2 | \mathcal{F}_{t-1}^{\zeta}) = \text{a positive constant.} \tag{6}$$

Here $\mathcal{F}_t^{\zeta} = \sigma(\ldots, \zeta_{t-1}, \zeta_t)$. To obtain the asymptotic distribution of the LW estimator, it is often additionally assumed that $\mathbb{E}(\zeta_t^k | \mathcal{F}_{t-1}^{\zeta}), k = 3,4$, are also constants (Velasco, 1999b; PS, 2004; SP, 2006). As mentioned in Robinson and Henry (1999), it is a drawback not to allow conditional heteroskedasticity, which is an intrinsic feature for autoregressive conditional heteroskedasticity (ARCH) and generalized autoregressive conditional heteroskedasticity (GARCH) models in financial time series. Robinson and Henry (1999) attempted to relax the constant conditional variance condition (6) for the LW estimator in the case of Type I processes when $d \in (-\frac{1}{2}, \frac{1}{2})$. For general nonlinear processes, Dalla et al. (2006) proved the consistency of LW estimates and also discussed the convergence rate. Their results indicate that the LW estimator might have a non-Gaussian limit distribution. No distributional theory was provided in their paper.

This paper has two goals. First, we attempt to solve the conjecture posed in PS (2004) for Type I processes when $d \in [\frac{3}{4}, \frac{3}{2})$. Second, we consider the asymptotic distribution of the LW estimator for a wide class of nonlinear processes

$$u_t = F(\ldots, \varepsilon_{t-1}, \varepsilon_t), \tag{7}$$

where ε_t are independent and identically distributed (i.i.d.) random variables and F is a measurable function such that u_t is well defined. Then u_t is a stationary causal ergodic process. The class of processes that (7) represents is huge; see Rosenblatt (1959, 1971), Kallianpur (1981), and Tong (1990, p. 204). As in Wiener (1958), Priestley (1988), and Wu (2005), (7) can be interpreted as a physical system with $(\ldots, \varepsilon_{t-1}, \varepsilon_t)$ being the input, F being a filter, and u_t being the output. Our dependence measures on the process $\{u_t\}$ basically quantify the degree of dependence of outputs on inputs (cf. Remark 2.3 in Section 2). The class (7) includes the linear process $u_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}$ as a special case. It also includes *general* GARCH (Bollerslev, 1986) and autoregressive moving average–GARCH (ARMA-GARCH) (Ling and Li, 1997) processes. Together with (1), the widely used fractional autoregressive integrated moving average–GARCH

(FARIMA-GARCH) model is also within this framework; see Section 4. Our theoretical results confirm the findings from finite-sample simulations in Robinson and Henry (1999), Henry (2001), and Nielsen and Frederiksen (2005) that the LW estimator is robust to conditional heteroskedastic innovations if the bandwidth is appropriately chosen.

The following notation will be used throughout the paper. For a random variable ξ, write $\xi \in \mathcal{L}^p$ $(p > 0)$ if $\|\xi\|_p := [\mathbb{E}(|\xi|^p)]^{1/p} < \infty$ and let $\|\cdot\| = \|\cdot\|_2$. For $\xi \in \mathcal{L}^1$ define projection operators $\mathcal{P}_k\xi = \mathbb{E}(\xi|\mathcal{F}_k) - \mathbb{E}(\xi|\mathcal{F}_{k-1})$, $\mathcal{F}_k = (\ldots, \varepsilon_{k-1}, \varepsilon_k)$. Let $C > 0$ denote a generic constant that may vary from line to line. Denote by $\Rightarrow$ and $\xrightarrow{\mathbb{P}}$ convergence in distribution and in probability, respectively. The symbols $O_{\mathbb{P}}(1)$, $o_{\mathbb{P}}(1)$, and $o_{a.s.}(1)$ signify being bounded in probability, convergence to zero in probability, and convergence in the almost sure sense, respectively. Let $N(\mu, \sigma^2)$ be a normal distribution with mean μ and variance σ^2.

The paper is organized as follows. Section 2 introduces the LW estimator and some notation and assumptions. Section 3 provides a distributional theory of LW estimates for Type I processes when $d \in (-\frac{1}{2}, \frac{1}{2}) \cup (\frac{1}{2}, \frac{3}{2})$. Applications are given in Section 4, and conclusions are made in Section 5. We leave the technical details to the Appendix.

2. LOCAL WHITTLE ESTIMATION

Let $i = \sqrt{-1}$ be the imaginary unit. For a process $\{Z_t\}_{t\in\mathbb{Z}}$, define the periodogram

$$I_Z(\lambda) = |w_Z(\lambda)|^2, \quad \text{where } w_Z(\lambda) = w_{Z,n}(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} Z_t e^{it\lambda}.$$

Let $\lambda_j = 2\pi j/n, j = 1, \ldots, n$, be the Fourier frequencies. Denote the true value of d by d_0. Given the observation $X_1, \ldots, X_n$, the LW estimate for d_0 is defined as the minimizer of the local objective function $R(d)$, i.e.,

$$\hat{d} = \operatorname{argmin}_{d\in[\Delta_1,\Delta_2]} R(d),$$

where

$$R(d) = \log\left(m^{-1} \sum_{j=1}^{m} \lambda_j^{2d} I_X(\lambda_j)\right) - \frac{2d}{m} \sum_{j=1}^{m} \log \lambda_j. \tag{8}$$

Throughout the paper, we assume that $m = m(n)$ satisfies $m^{-1} + m/n \to 0$, Δ_1 and Δ_2 satisfy $-\frac{1}{2} < \Delta_1 < \Delta_2 < \infty$, and $d_0 \in [\Delta_1, \Delta_2]$.

Let $\mathbb{E}(u_t) = 0$ and $\gamma_u(k) = \mathbb{E}(u_t u_{t+k})$ be the covariance function of (u_t); let

$$f_u(\lambda) = \frac{1}{2\pi} \sum_{k\in\mathbb{Z}} \gamma_u(k)\exp(ik\lambda)$$

be the spectral density. Assume throughout the paper that $G_0 = f_u(0) > 0$. We make the following assumptions.

Assumption 2.1. $C_U := \sum_{k\in\mathbb{Z}} |\gamma_u(k)| < \infty$.

Remark 2.1. Assumption 2.1 indicates that u_t has short memory. Under this assumption, the spectral density f_u is continuous and bounded.

Assumption 2.2. $\sum_{k=0}^{\infty} k\|\mathcal{P}_0 u_k\| < \infty$.

Remark 2.2. The quantity $\|\mathcal{P}_0 u_k\|$ is closely related to the predictive dependence measure introduced by Wu (2005); see Remark 2.3 for more discussion. Following Wu and Min (2005), Assumption 2.2 is called $\mathcal{L}^2$ dependent with order 1. Assumption 2.2 implies $\sum_{k\in\mathbb{Z}} |k\gamma_u(k)| < \infty$ (cf Lemma A.1 in the Appendix); thus $f_u(\cdot)$ is continuously differentiable over $[-\pi, \pi]$.

Remark 2.3. Let $\{\varepsilon_t'\}_{t\in\mathbb{Z}}$ be an i.i.d. copy of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ and, for $k \geq 0$, $u_k' = F(\mathcal{F}_k')$, where $\mathcal{F}_k' = (\mathcal{F}_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_k)$. Interpreting (7) as a physical system, Wu (2005) introduced the physical dependence measure

$$\delta_q(k) := \|u_k - u_k'\|_q$$

and the predictive dependence measure $\omega_q(k) := \|g_k(\mathcal{F}_0) - g_k(\mathcal{F}_0')\|_q$, where $g_k(\mathcal{F}_0) = \mathbb{E}(u_k|\mathcal{F}_0)$. Intuitively, $\delta_q(\cdot)$ quantifies the dependence of u_k on ε_0 by measuring the distance between u_k and its coupled version u_k'. For $\omega_q(k)$, because $g_k(\mathcal{F}_0) = \mathbb{E}(u_k|\mathcal{F}_0)$ is the k-step-ahead predicted mean, $\omega_q(k)$ measures the contribution of ε_0 in predicting future expected values. By Theorem 1 in Wu (2005), $\|\mathcal{P}_0 u_k\| \leq \omega_2(k) \leq 2\|\mathcal{P}_0 u_k\|$. So Assumption 2.2 is equivalent to $\sum_{k=0}^{\infty} k\omega_2(k) < \infty$. In many applications physical and predictive dependence measures are easy to use because they are directly related to data-generating mechanisms.

Assumption 2.3. $\sum_{k_1,k_2,k_3\in\mathbb{Z}} |\text{cum}(u_0, u_{k_1}, u_{k_2}, u_{k_3})| < \infty$.

Remark 2.4. Summability conditions on joint cumulants are widely used in spectral analysis. It is an important problem to verify such conditions. For a linear process $u_t = \sum_{j\in\mathbb{Z}} a_j \varepsilon_{t-j}$ with ε_j being i.i.d., Assumption 2.3. holds if $\sum_{j\in\mathbb{Z}} |a_j| < \infty$ and $\varepsilon_1 \in \mathcal{L}^4$. For nonlinear processes (7), it is satisfied under a geometric moment contraction (GMC) condition with order 4 (Wu and Shao, 2004). The process $\{u_t\}$ satisfies GMC with order α, $\alpha > 0$, if there exists a $C = C(\alpha)$ and $\rho = \rho(\alpha) \in (0,1)$ such that

$$\mathbb{E}(|u_n^* - u_n|^\alpha) \leq C\rho^n, \qquad n \in \mathbb{N}, \tag{9}$$

where $u_n^* = F(\ldots, \varepsilon'_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_n)$ is a coupled version of u_n. Here $\{\varepsilon'_t\}_{t\in\mathbb{Z}}$ is an i.i.d. copy of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. The property (9) indicates that the process $\{u_n\}$ forgets its past exponentially fast, and it can be verified for many nonlinear time series models (Wu and Min, 2005; Shao and Wu, 2007). Define the fourth cumulant spectral density

$$f_4(w_1, w_2, w_3) = \frac{1}{(2\pi)^3} \sum_{k_1, k_2, k_3 \in \mathbb{Z}} \operatorname{cum}(u_0, u_{k_1}, u_{k_2}, u_{k_3}) \exp\left(-i \sum_{j=1}^{3} w_j k_j\right).$$

Under Assumption 2.3, $f_4(\cdot,\cdot,\cdot)$ is continuous and bounded. Section 4 provides another sufficient condition for the summability of joint cumulants (see Theorem 4.1).

Assumption 2.4. Assume $u_t \in \mathcal{L}^q$ and $\sum_{k=1}^{\infty} k\delta_q(k) < \infty$, $q > 4$.

Remark 2.5. For the linear causal process $u_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, $\delta_q(k) = |a_k| \|\varepsilon'_0 - \varepsilon_0\|_q$. So Assumption 2.4 holds if $\sum_{k=0}^{\infty} k|a_k| < \infty$ and $\varepsilon_1 \in \mathcal{L}^q$. Assumption 2.4 implies Assumption 2.2 because $\mathbb{E}[(u_k - u'_k)|\mathcal{F}_0] = \mathcal{P}_0 u_k$ and $\|\mathcal{P}_0 u_k\| \le \|\mathcal{P}_0 u_k\|_q \le \|u_k - u'_k\|_q$.

Assumption 2.5. For some $\beta \in (0,2]$, $f_u(\lambda) = (1 + O(\lambda^\beta))G_0$ as $\lambda \downarrow 0$.

Remark 2.6. Assumption 2.5 is commonly made in the study of local Whittle estimation (PS, 2004; SP, 2006). For the popular FARIMA(p,d,q) or FARIMA-GARCH processes (see Section 4), $\beta = 2$. Our Assumptions 2.2 and 2.4 imply the continuous differentiability of $f_u(\cdot)$; thus $\beta \ge 1$.

Assumption 2.6. Assume for $\beta > 1$,

$$(\log n)^3 = o(m) \quad \text{and} \quad m = o(n^{2/3}), \tag{10}$$

and for $\beta = 1$, $(\log n)^3 = o(m)$, $m^3(\log m)^2/n^2 \to 0$.

Remark 2.7. Previous work by Robinson (1995b) and PS (2004) requires

$$\frac{1}{m} + \frac{m^{2\beta+1}(\log m)^2}{n^{2\beta}} \to 0. \tag{11}$$

Note that for $\beta \ge 1$, Assumption 2.6 is stronger than (11). We are unable to relax Assumption 2.6, which is needed for us to apply Corollary 2 in the companion paper in this issue (Wu and Shao, 2007) to obtain asymptotic distributions of $\hat{d}$. Under (11), the variance of $\hat{d}$ dominates its bias. In terms of choosing the bandwidth that minimizes the mean square error, the optimal order for m is $n^{2\beta/(2\beta+1)}$ (cf. Henry and Robinson, 1996). On the other hand, Assumption 2.6 does not seem to be overly restrictive: Assumption A4′ in Robinson and Henry (1999) requires $m = o(n^{1/2}/\log n)$ in an early attempt to establish a central limit theorem for the LW estimator without imposing (6).

3. MAIN RESULTS

Four cases, $d_0 \in (-\frac{1}{2},\frac{1}{2})$, $d_0 \in (\frac{1}{2},1)$, $d_0 \in (1,\frac{3}{2})$, and $d_0 = 1$ are dealt with in Sections 3.1, 3.2, 3.3, and 3.4, respectively.

3.1. $d_0 \in (-\frac{1}{2},\frac{1}{2})$

Under the Type I formulation (1), we write

$$X_t = \mu + (1-B)^{-d_0}u_t = \mu + \sum_{j=0}^{\infty} \phi_j(d_0)u_{t-j}. \tag{12}$$

Because $I_X(\lambda_j), j = 1, \dots, m$, which is used in the LW estimation, is invariant to the mean, we can and do assume $\mu = 0$.

PROPOSITION 3.1. *For the process (12), let $u_t \in \mathcal{L}^4$. Then under Assumption 2.2, we have $\hat{d} \xrightarrow{\mathbb{P}} d_0$ and*

$$2(\hat{d} - d_0) = -F_m(1 + o_{\mathbb{P}}(1)) + O_{\mathbb{P}}(\log m/m),$$

where $F_m = m^{-1}\sum_{j=1}^{m} s_j \lambda_j^{2d_0} G_0^{-1} I_X(\lambda_j)$ and $s_j = \log(j/m) + 1$.

THEOREM 3.1. *For the process (12), under Assumptions 2.3–2.6, we have*

$$\sqrt{m}(\hat{d} - d_0) \Rightarrow N(0,\tfrac{1}{4}). \tag{13}$$

Robinson (1995b) first obtained the asymptotic distribution of the LW estimator in the setting of linear processes under the assumption that the innovations are martingale differences with constant conditional variance. Robinson and Henry (1999) attempted to relax the latter restriction (see (6)) under the following framework:

$$X_t = \mu + \sum_{j=0}^{\infty} a_j \zeta_{t-j}, \qquad \sum_{j=0}^{\infty} a_j^2 < \infty,$$

where

$$\mathbb{E}(\zeta_t|\mathcal{F}_{t-1}^{\zeta}) = 0, \qquad \sigma_t^2 = \psi_0 + \sum_{j=1}^{\infty} \psi_j \zeta_{t-j}^2, \qquad \psi_0 \geq 0, \qquad \psi_j \geq 0, j \in \mathbb{N}. \tag{14}$$

The widely used ARCH and GARCH models are included in this framework; compare (22) in Section 4. The asymptotic normality was obtained under quite restrictive conditions. For example, they require $\max_{t\in\mathbb{Z}} \mathbb{E}(\zeta_t^8) < \infty$, $\mathbb{E}(\zeta_t^3|\mathcal{F}_{t-1}^{\zeta}) = \mathbb{E}(\zeta_t^3)$ almost surely and for $t \geq u \geq v$,

$$\mathbb{E}(\zeta_t^2 \zeta_u \zeta_{v-1}) = 0, \qquad \mathbb{E}(\zeta_t^4 \zeta_u|\mathcal{F}_{u-1}^{\zeta}) = \mathbb{E}(\zeta_t^4 \zeta_u^2 \zeta_v|\mathcal{F}_{v-1}^{\zeta}) = 0 \quad \text{almost surely.}$$

In the GARCH case (22), we only need to assume a qth moment condition ($q > 4$) and allow various forms of conditional heteroskedasticity; see Section 4 for more discussion. Note that the existence of higher order moments requires more restriction on the parameter space in the GARCH type models. On the other hand, our formulation excludes possible long memory in conditional variance, which is included in their framework.

Remark 3.1. The short memory conditions on u_t (such as Assumptions 2.2 and 2.4) imply the continuous differentiability of $f_u(\cdot)$ over the full band $[-\pi, \pi]$, whereas in previous work (see Robinson, 1995b; Robinson and Henry, 1999; Phillips and Shimotsu, 2004), only local smoothness of $f_u(\cdot)$ or $f_X(\cdot)$ around a neighborhood of zero frequency is imposed. In particular, our assumptions on u_t exclude the so-called Gegenbauer process (Gray, Zhang, and Woodward, 1989), in which the spectral density function has a pole at a nonzero frequency (for more discussion, see Phillips and Shimotsu, 2004). These global smoothness assumptions together with the stronger rate condition on m are the prices we pay for allowing nonlinear processes.

3.2. $d_0 \in (\frac{1}{2}, 1)$

Recall the definition of X_t in (3). Write

$$X_n - X_0 = \sum_{t=1}^{n} Y_t = \sum_{j=0}^{\infty} A_{j,n} u_{n-j}, \tag{15}$$

where $A_{j,n} = \Phi_j - \Phi_{j-n}$, $\Phi_j = \sum_{i=0}^{j} \phi_i(d_0 - 1) = \phi_j(d_0)$ if $j \geq 0$ and $\Phi_j = 0$ if $j < 0$. Further write

$$\begin{aligned} w_X(\lambda_s) &= (2\pi n)^{-1/2} \sum_{t=1}^{n} \sum_{j=1}^{t} Y_j e^{it\lambda_s} = (2\pi n)^{-1/2} \sum_{j=1}^{n} Y_j \sum_{t=j}^{n} e^{it\lambda_s} \\ &= -\frac{e^{i\lambda_s}}{1 - e^{i\lambda_s}} \frac{(X_n - X_0)}{\sqrt{2\pi n}} + \frac{w_Y(\lambda_s)}{1 - e^{i\lambda_s}}. \end{aligned} \tag{16}$$

For a complex number z let $\mathrm{Re}(z)$ be its real part and $\bar{z}$ its conjugate. Then

$$I_X(\lambda_s) = \frac{(X_n - X_0)^2}{2\pi n |1 - e^{i\lambda_s}|^2} + \frac{I_Y(\lambda_s)}{|1 - e^{i\lambda_s}|^2} - \frac{2(X_n - X_0)}{\sqrt{2\pi n} |1 - e^{i\lambda_s}|^2} \mathrm{Re}(w_Y(\lambda_s) e^{-i\lambda_s}). \tag{17}$$

PROPOSITION 3.2. *Suppose that X_t is generated by (3) with $d_0 \in (\frac{1}{2}, 1)$. Assume $u_t \in \mathcal{L}^4$ and Assumption 2.2; we have $\hat{d} \xrightarrow{\mathbb{P}} d_0$ and $2(\hat{d} - d_0) = -F_m(1 + o_{\mathbb{P}}(1)) + O_{\mathbb{P}}(\log m/m)$.*

Remark 3.2. Under Type I formulation, Velasco (1999b) showed that $\hat{d}$ is consistent for linear processes when $d_0 \in [\frac{1}{2}, 1)$. For Type II processes, the consistency was established by PS (2004). At $d_0 = \frac{1}{2}$, we conjecture that the consistency of $\hat{d}$ still holds for nonlinear processes.

THEOREM 3.2. *Let X_t be defined in (3) with $d_0 \in (\frac{1}{2}, 1)$.*

(i) Under Assumptions 2.3–2.6, we have

$$m^{1/2}(\hat{d} - d_0) \Rightarrow U_1/2, \qquad d_0 \in (\tfrac{1}{2}, \tfrac{3}{4}),$$

$$m^{2-2d_0}(\hat{d} - d_0) \Rightarrow J(d_0)U_2^2, \qquad d_0 \in (\tfrac{3}{4}, 1),$$

where

$$J(d_0) = \frac{(2\pi)^{2d_0-2}(1-d_0)}{\Gamma(d_0)^2(2d_0-1)^2} \times \left\{ \frac{1}{(2d_0-1)} + \int_1^\infty (y^{d_0-1} - (y-1)^{d_0-1})^2\, dy \right\} \tag{18}$$

and U_1 and U_2 are i.i.d. $N(0,1)$ random variables.

(ii) Let $u_t = \sum_{i=0}^\infty a_i \varepsilon_{t-i}$, $\sum_{k=0}^\infty k|a_k| < \infty$, $\sum_{i=0}^\infty a_i \neq 0$, $\mathbb{E}\varepsilon_1^2 = 1$, and $\varepsilon_1 \in \mathcal{L}^4$. Then under Assumption 2.5 and (11), for $d_0 = \frac{3}{4}$, we have

$$m^{1/2}(\hat{d} - d_0) \Rightarrow U_1/2 + J(d_0)U_2^2.$$

PS (2004) dealt with Type II processes and conjectured that if $d_0 \in [\frac{3}{4}, 1)$, then their asymptotic results still hold for Type I processes possibly with different $J(d_0)$. Here, we prove the conjecture in the framework of fractionally integrated nonlinear processes. In comparison with the asymptotic distribution in the Type II case, the extra term $\int_1^\infty (y^{d_0-1} - (y-1)^{d_0-1})^2\, dy$ in (18) is due to the so-called prehistorical influence. At $d_0 = \frac{3}{4}$, the first two terms in (17) are of the same stochastic magnitude. In this particular case, it seems very hard to obtain asymptotic results for nonlinear processes.

3.3. $d_0 \in (1, \frac{3}{2})$

THEOREM 3.3. *Suppose that $\{X_t\}$ is generated from (3) with $d_0 \in [1, \frac{3}{2})$ and $u_t \in \mathcal{L}^4$. Then under Assumption 2.2, we have $\hat{d} \xrightarrow{\mathbb{P}} 1$.*

Remark 3.3. Theorem 3.3 confirms the conjecture posed by PS (2004) (see Remark 3.3 therein), which states that $\hat{d} \xrightarrow{\mathbb{P}} 1$ for Type I processes if $d_0 \in (1, \frac{3}{2})$. More refined structures exist if $d_0 = 1$; see Section 3.4.

3.4. $d_0 = 1$

THEOREM 3.4. *Suppose that* $\{X_t\}$ *is generated from (3) with* $d_0 = 1$. *Assume that* $\sum_{k=1}^{\infty} k\delta_4(k) < \infty$ *and* $m = m_n \to \infty$ *satisfies*

$$(m^{3/2}\log m)\chi(n) = O(1), \quad \text{where } \chi(n) = n^{-1/4}\log(n). \tag{19}$$

Then under Assumptions 2.3. and 2.5, we have

$$\sqrt{m}(\hat{d} - d_0) \Rightarrow \frac{-W_1 + \sqrt{2}W_2W_3}{2(1 + W_3^2)}, \tag{20}$$

where W_1, W_2, W_3 *are i.i.d.* $N(0,1)$.

The condition $\sum_{k=1}^{\infty} k\delta_4(k) < \infty$ in Theorem 3.4 is related to Assumption 2.4 with $q = 4$. Assumption 2.4 is stronger.

The limit distribution in (20) is equivalent to the form stated in Theorem 4.2 of PS (2004), who obtained the asymptotic distribution of $\hat{d}$ for Type II processes with $d_0 = 1$. It suggests the interesting dichotomous phenomenon: the asymptotic behaviors in the three cases $d_0 < 1$, $d_0 = 1$, and $d_0 > 1$ are very different. In addition, the condition on the bandwidth m is quite stringent here. Basically we require $m = O(n^{1/6}(\log n)^{-4/3})$.

In summary, for fractional nonlinear processes ((1) and (3)), the LW estimator is consistent when $d_0 \in (-\frac{1}{2}, \frac{1}{2}) \cup (\frac{1}{2}, 1]$ and is inconsistent when $d_0 > 1$. When $d_0 \in (-\frac{1}{2}, \frac{1}{2}) \cup (\frac{1}{2}, \frac{3}{4})$, the asymptotic distribution is normal with asymptotic variance independent of the true value d_0. For Type II processes, SP (2000) proposed a modified LW estimator, which basically replaces $I_X(\lambda_j)$ in $R(d)$ (see (8)) by $I_X^*(\lambda_j)$, where

$$I_X^*(\lambda_j) = |w_X^*(\lambda_j)|^2, \qquad w_X^*(\lambda_j) = w_X(\lambda_s) + \frac{e^{i\lambda_s}}{1 - e^{i\lambda_s}} \frac{X_n - X_0}{\sqrt{2\pi n}}.$$

The modified LW estimator is shown to be consistent for $d_0 \in (0,2)$ and is asymptotically normally distributed with variance $\frac{1}{4}$ for $d_0 \in (\frac{1}{2}, \frac{7}{4})$. We would expect that this result still holds for Type I processes in our setting in view of (16), although the boundary case $d_0 = \frac{3}{2}$ is hard to handle. We shall not pursue this generalization in this paper.

4. APPLICATIONS

In this section, we shall show that our main technical assumptions, Assumptions 2.3 and 2.4, are satisfied by a number of widely used models in financial time series analysis. The so-called FARIMA(p,d,q)-GARCH(r,s) model (Ling and Li, 1997; Li, Ling, and McAleer, 2002) has been used by Baillie, Chung, and Tieslau (1996), Hauser and Kunst (1998a, 1998b), and Lien and Tse (1999) among others to model both long memory and conditional heteroskedasticity. It admits the following form:

$$\phi(B)(1-B)^d(X_t-\mu)=\psi(B)\zeta_t, \tag{21}$$

$$\zeta_t=\varepsilon_t\sigma_t, \qquad \sigma_t^2=\alpha_0+\sum_{i=1}^{r}\alpha_i\zeta_{t-i}^2+\sum_{i=1}^{s}\beta_i\sigma_{t-i}^2, \tag{22}$$

where $\{\varepsilon_t\}$ are i.i.d. with zero mean and unit variance, $\phi(B)=1-\sum_{i=1}^{p}\phi_i B^i$, and $\psi(B)=1+\sum_{i=1}^{q}\psi_i B^i$. Assume that all the roots of $\phi(z)=1-\sum_{i=1}^{p}\phi_i z^i$ and $\psi(z)=1+\sum_{i=1}^{q}\psi_i z^i$ are outside the unit circle, $\phi_p\neq 0$, $\psi_q\neq 0$, and $\phi(z)$ and $\psi(z)$ have no common root. Rewrite (21) into the form of (1) with $u_t=\phi(B)^{-1}\psi(B)\zeta_t$, where u_t is an ARMA-GARCH process. For a GARCH process, the necessary and sufficient conditions for the existence of fourth moments have been investigated by Ling and McAleer (2002a, 2002b). Wu and Min (2005) showed that ζ_t is GMC(4) (cf. eqn. (9) with $\alpha=4$) provided that $\zeta_t\in\mathcal{L}^4$. Because an ARMA process with GMC(4) innovations is still GMC(4) (cf. Shao and Wu, 2007, Thm. 5), the process u_t is GMC(4) if $\zeta_t\in\mathcal{L}^4$. Therefore, our Assumption 2.3 is satisfied because GMC(4) implies the summability of fourth cumulants (cf. Wu and Shao, 2004, Prop. 2).

In the literature, various forms of GARCH have been proposed to model conditional heteroskedasticity. An important class is the asymmetric GARCH(r,s) models (Ding, Granger, and Engle, 1993). Interestingly, for general asymmetric GARCH(r,s) models (which include (22) as a special case), Shao and Wu (2007) showed that they satisfy GMC(q), $q\in\mathbb{N}$ under the qth moment condition. Another popular asymmetric GARCH model is the so-called EGARCH(p,q) (exponential GARCH(p,q)) model proposed by Nelson (1991), which admits the following form:

$$\zeta_t=\sigma_t\varepsilon_t, \qquad \log(\sigma_t^2)=\alpha_0+\frac{\tilde{\psi}(B)}{\tilde{\phi}(B)}g(\varepsilon_{t-1}),$$

$$g(\varepsilon_t)=\theta\varepsilon_t+\gamma[|\varepsilon_t|-\mathbb{E}(|\varepsilon_t|)],$$

where α_0, θ, and γ are constants and $\tilde{\psi}(B)=1+\beta_1 B+\cdots+\beta_q B^q$ and $\tilde{\phi}(B)=1-\alpha_1 B-\cdots-\alpha_p B^p$ are polynomials with zeros outside the unit circle having no common factors. Robinson and Henry (1999) also considered the preceding model in their finite-sample simulation. However, they mentioned that the EGARCH model is not included in their theoretical framework (14). If ε_t are i.i.d. $N(0,1)$, Min (2004) showed that ζ_t is GMC(q) for any $q\in\mathbb{N}$.

For other types of nonlinear time series models such as bilinear models (Subba Rao and Gabr, 1984), threshold autoregressive models (Tong, 1990), and signed volatility models (Yao, 2004), the GMC property has been verified by Wu and Min (2005) and Shao and Wu (2007) under certain contraction conditions.

Regarding our Assumption 2.4, it holds if the process u_t is $\mathrm{GMC}(q)$ for some $q > 4$ (for a rigorous proof see Wu, 2007).

A common assumption in spectral analysis is the summability of joint cumulants up to certain orders (Brillinger, 1975; Rosenblatt, 1985). The following theorem gives a sufficient condition under which the summability of joint cumulants is true. It generalizes Proposition 2 in Wu and Shao (2004).

THEOREM 4.1. *Assume* $u_t = F(\ldots, \varepsilon_{t-1}, \varepsilon_t) \in \mathcal{L}^k$, $k \geq 2$, $k \in \mathbb{N}$. *Then*

$$\sum_{m_1,\ldots,m_{k-1}\in\mathbb{Z}} |\mathrm{cum}(u_0, u_{m_1}, \ldots, u_{m_{k-1}})| < \infty \tag{23}$$

provided that $\sum_{n=0}^{\infty} n^{k-2} (\sum_{m=n}^{\infty} \delta_k(m)^2)^{1/2} < \infty$, *where* $\delta_k(m) = \|u_m - u'_m\|_k$.

Remark 4.1. Because $(\sum_{m=n}^{\infty} \delta_k(m)^2)^{1/2} \leq \sum_{m=n}^{\infty} \delta_k(m)$, (23) holds if

$$\sum_{n=0}^{\infty} n^{k-1} \delta_k(n) < \infty, \qquad k \geq 2. \tag{24}$$

Therefore $\sum_{n=0}^{\infty} n^3 \delta_q(n) < \infty$, $q > 4$ implies our Assumptions 2.3, 2.4, and $\beta = 2$ in Assumption 2.5.

Example 4.1. ARCH(∞) (Robinson, 1991)

$$\zeta_t = \varepsilon_t \sigma_t, \qquad \sigma_t^2 = \psi_0 + \sum_{j=1}^{\infty} \psi_j \zeta_{t-j}^2, \qquad \psi_0 \geq 0, \qquad \psi_j \geq 0, \qquad j \in \mathbb{N}, \tag{25}$$

where ε_t are i.i.d. mean zero random variables having suitable moments. Note that (25) is a special form of (14). Both general $\mathrm{ARCH}(p)$ and $\mathrm{GARCH}(r,s)$ models fall into this framework when the weights ψ_j either vanish for $j > p$ or decay exponentially to zero. This property implies the exponential decay rate of the autocorrelation of ζ_t^2. Giraitis, Kokoszka, and Leipus (2000) gave a sufficient condition for the existence of a stationary solution and found that the autocorrelation of ζ_t^2 can decay slowly like a power function, but no long memory structure of ζ_t^2 is allowed. Further development can be found in Zaffaroni (2004). We shall show that our condition can be satisfied with hyperbolically decaying coefficients ψ_j.

PROPOSITION 4.1. *For (25), let* $\varepsilon_1 \in \mathcal{L}^6$. *Assume that* $\|\varepsilon_1^2\|_3^{1/2} \sum_{j=1}^{\infty} \psi_j^{1/2} < 1$ *and* $\sum_{j=1}^{\infty} \psi_j^{1/2} j^3 < \infty$, *then we have*

$$\sum_{k_1,k_2,k_3\in\mathbb{Z}} |\mathrm{cum}(\zeta_0, \zeta_{k_1}, \zeta_{k_2}, \zeta_{k_3})| < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} k \|\zeta_k - \zeta'_k\|_6 < \infty, \tag{26}$$

where ζ'_k *is the coupled version of* ζ_k *as in Remark 2.3.*

Example 4.2. LARCH (linear ARCH) (Robinson, 1991)

$$\zeta_t = \varepsilon_t \sigma_t, \qquad \sigma_t = a + \sum_{j=1}^{\infty} b_j \zeta_{t-j}, \qquad t \in \mathbb{Z}, \tag{27}$$

where a and $b_j, j \in \mathbb{N}$ are not constrained to be nonnegative. Giraitis, Robinson, and Surgailis (2000) provided a sufficient and necessary condition for the existence of a stationary solution and demonstrated that ζ_t^2 could have long memory autocorrelation unlike the ARCH(∞) case. The following proposition covers part of the short memory case where b_j is allowed to have a hyperbolic decay.

PROPOSITION 4.2. *For (27), let $\varepsilon_1 \in \mathcal{L}^5$. Assume that $\|\varepsilon_1\|_5 \sum_{j=1}^{\infty} |b_j| < 1$ and $\sum_{j=1}^{\infty} |b_j| j^3 < \infty$. Then*

$$\sum_{k_1,k_2,k_3 \in \mathbb{Z}} |\mathrm{cum}(\zeta_0, \zeta_{k_1}, \zeta_{k_2}, \zeta_{k_3})| < \infty \quad \textit{and} \quad \sum_{k=1}^{\infty} k \|\zeta_k - \zeta_k'\|_5 < \infty.$$

5. CONCLUSIONS

This paper presents an asymptotic theory for the LW estimator of a class of fractionally integrated nonlinear processes. The theory we develop here matches the empirical evidence found in finite-sample simulations by Robinson and Henry (1999), Henry (2001), and Nielsen and Frederiksen (2005) that suggest that the LW estimator is robust to conditional heteroskedasticity. Recently, long memory in volatility has received a great deal of attention in the literature (cf. Hurvich, Moulines, and Soulier, 2005, and references therein). Robinson and Henry (1999) showed that the LW estimator of the long memory parameter in the level is unaffected by long memory in volatility. Our framework excludes this interesting case in that the conditional heteroskedastic models included are all of short memory type. However, our framework is general enough to allow various kinds of short memory GARCH models.

Under the current framework, it is certainly worth investigating the asymptotic properties of the LW estimator for Type II processes and also the exact LW estimator (SP, 2005). These topics are beyond the scope of this paper and will be studied in the future.

REFERENCES

Baillie, R.T., C.F. Chung, & M.A. Tieslau (1996) Analyzing inflation by the fractionally integrated ARFIMA-GARCH model. *Journal of Applied Econometrics* 11, 23–40.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31, 307–327.

Brillinger, D.R. (1975) *Time Series: Data Analysis and Theory*. Holden-Day.

Brockwell, P.J. & R.A. Davis (1991) *Time Series: Theory and Methods*. 2nd ed. Springer-Verlag.

Dalla, V., L. Giraitis, & J. Hidalgo (2006) Consistent estimation of the memory parameter for nonlinear time series. *Journal of Time Series Analysis* 27, 211–251.

Ding, Z., C. Granger, & R. Engle (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.

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

Giraitis, L., P. Kokoszka, & R. Leipus (2000) Stationary ARCH models: Dependent structure and central limit theorem. *Econometric Theory* 16, 3–22.

Giraitis, L., P.M. Robinson, & D. Surgailis (2000) A model for long memory conditional heteroscedasticity. *Annals of Applied Probability* 10, 1002–1024.

Gray, H.L., N.-F. Zhang, & W.A. Woodward (1989) On generalized fractional processes. *Journal of Time Series Analysis* 10, 233–257.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Applications.* Academic Press.

Hannan, E.J. (1973) Central limit theorems for time series regression. *Zeitschrift für Wahrscheinlichskeitstheorie und Verwandte Gebiete* 26, 157–170.

Hauser, M.A. & R.M. Kunst (1998a) Fractionally integrated models with ARCH errors: With an application to the Swiss 1-month euromarket interest rate. *Review of Quantitative Finance and Accounting* 10, 95–113.

Hauser, M.A. & R.M. Kunst (1998b) Forecasting high-frequency financial data with the ARFIMA-ARCH model. *Journal of Forecasting* 20, 501–518.

Henry, M. (2001) Robust automatic bandwidth for long memory. *Journal of Time Series Analysis* 22, 293–316.

Henry, M. & P.M. Robinson (1996) Bandwidth choice in Gaussian semiparametric estimation of long range dependence. In P.M. Robinson & M. Rosenblatt (eds.), *Athens Conference on Applied Probability and Time Series Analysis*, vol. II: *Time Series Analysis, In Memory of E.J. Hannan*, Lecture Notes in Statistics, 115, 220–232. Springer-Verlag.

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

Kallianpur, G. (1981) Some ramifications of Wiener's ideas on nonlinear prediction. In N. Wiener & P. Masani (eds.), *Norbert Wiener, Collected Works with Commentaries*, pp. 402–424. MIT Press.

Kim, C.S. & P.C.B. Phillips (1999) Log Periodogram Regression: The Nonstationary Case. Technical report, Yale University.

Künsch, H. (1987) Statistical aspects of self-similar processes. In Yu.A. Prohorov & V.V. Sazonov (eds.), *Proceedings of the 1st World Congress of the Bernoulli Society*, vol. 1, pp. 67–74. Science Press.

Li, W.K., S. Ling, & M. McAleer (2002) A survey of recent theoretical results for time series models with GARCH errors. *Journal of Economic Survey* 16, 245–269.

Lien, D. & Y.K. Tse (1999) Forecasting the Nikkei spot index with fractional cointegration. *Journal of Forecasting* 18, 259–273.

Ling, S. & W.K. Li (1997) On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of the American Statistical Association* 92, 1184–1194.

Ling, S. & M. McAleer (2002a) Necessary and sufficient moment conditions for the GARCH(p,q) and asymmetric power GARCH(p,q) models. *Econometric Theory* 18, 722–729.

Ling, S. & M. McAleer (2002b) Stationarity and the existence of moments of a family of GARCH processes. *Journal of Econometrics* 106, 109–117.

Marinucci, D. & P.M. Robinson (1999a) Weak convergence of multivariate fractional processes. *Stochastic Processes and Their Applications* 86, 103–120.

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

Min, W. (2004) Inference on time series driven by dependent innovations. Ph.D. Dissertation, University of Chicago.

Moulines, E. & P. Soulier (2003) Semiparametric spectral estimation for fractional processes. In P. Doukhan, G. Oppenheim, & M. Taqqu (eds.), *Theory and Applications of Long-Range Dependence*, pp. 252–301. Springer Verlag.

Nelson, D.B. (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59, 347–370.

Nielsen, M.O. & P.H. Frederiksen (2005) Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews* 24, 405–443.

Phillips, P.C.B. (1999a) Discrete Fourier Transforms of Fractional Processes. Technical report, Yale University.

Phillips, P.C.B. (1999b) Unit Root Log Periodogram Regression. Technical report, Yale University.

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

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

Priestley, M.B. (1988) *Nonlinear and Nonstationary Time Series Analysis*. Academic Press.

Robinson, P.M. (1991) Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* 47, 67–84.

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.

Robinson, P.M. & M. Henry (1999) Long and short memory conditional heteroskedasticity in estimating the memory parameter of levels. *Econometric Theory* 15, 299–336.

Rosenblatt, M. (1959) Stationary processes as shifts of functions of independent random variables. *Journal of Mathematical Mechanics* 8, 665–681.

Rosenblatt, M. (1971) *Markov Processes. Structure and Asymptotic Behavior*. Springer-Verlag.

Rosenblatt, M. (1985) *Stationary Sequences and Random Fields.* Birkhäuser.

Shao, X. & W.B. Wu (2005) Local Whittle Estimation of Fractional Integration for Nonlinear Processes. Technical report 560, Department of Statistics, University of Chicago. Available at http://galton.uchicago.edu/techreports/tr560r1.pdf.

Shao, X. & W.B. Wu (2007) Asymptotic spectral theory for nonlinear time series. *Annals of Statistics*, forthcoming.

Shimotsu, K. & P.C.B. Phillips (2000) Modified Local Whittle Estimation of the Memory Parameter in the Nonstationary Case. Technical report, Yale University.

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* 103, 209–233.

Subba Rao, T. & M.M. Gabr (1984) *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Lecture Notes in Statistics 24. Springer-Verlag.

Tong, H. (1990) *Non-linear Time Series: A Dynamical System Approach.* Oxford University Press.

Velasco, C. (1999a) Non-stationary log-periodogram regression. *Journal of Econometrics* 91, 325–371.

Velasco, C. (1999b) Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* 20, 87–127.

Velasco, C. (2000) Non-Gaussian log-periodogram regression. *Econometric Theory* 16, 44–79.

Wiener, N. (1958) *Nonlinear Problems in Random Theory.* MIT Press.

Wu, W.B. (2005) Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences* 102, 14150–14154.

Wu, W.B. (2007) Strong invariance principles for dependent random variables. *Annals of Probability*, forthcoming.

Wu, W.B. & W. Min (2005) On linear processes with dependent innovations. *Stochastic Processes and Their Applications* 115, 939–958.
Wu, W.B. & X. Shao (2004) Limit theorems for iterated random functions. *Journal of Applied Probability* 41, 425–436.
Wu, W.B. & X. Shao (2006) Invariance principles for fractionally integrated nonlinear processes. In J. Sun, A. DasGupta, V. Melfi, & C. Page (eds.), *Recent Developments in Nonparametric Inference and Probability*. IMS Lecture Notes—Monograph Series 50, 20–30. Institute of Mathematical Statistics.
Wu, W.B. & X. Shao (2007) A limit theorem for quadratic forms and its applications. *Econometric Theory* 23 (this issue).
Yao, J.-F. (2004) Stationarity, Tails and Moments of Univariate GARCH Processes. Preprint, University of Rennes, France.
Zaffaroni, P. (2004) Stationarity and memory of ARCH(∞) models. *Econometric Theory* 20, 147–160.

TECHNICAL APPENDIX AND PROOFS

For convenience of notation, write $I_{Xj} = I_X(\lambda_j)$, $I_{uj} = I_u(\lambda_j)$, $I_{Yj} = I_Y(\lambda_j)$, $f_{Xj} = f_X(\lambda_j)$, $f_{uj} = f_u(\lambda_j)$, $f_{Yj} = |1 - e^{i\lambda_j}|^{2-2d_0} f_{uj}$, $\tilde{I}_{Xj} = I_{Xj} f_{Xj}^{-1}$, $\tilde{I}_{uj} = I_{uj} f_{uj}^{-1}$, $\tilde{I}_{Yj} = I_{Yj} f_{Yj}^{-1}$, $w_{Xj} = w_X(\lambda_j)$, $w_{uj} = w_u(\lambda_j)$, and $w_{Yj} = w_Y(\lambda_j)$, $j = 1, \ldots, m$.

Let $g_j = w_{Xj} / \sqrt{f_{Xj}}$ and $h_j = w_{uj} / \sqrt{f_{uj}}$. Denote $D(w) = D_n(w) = \sum_{t=1}^n e^{itw}$, $\alpha(\lambda) = \sum_{l=0}^\infty \phi_l(d_0) e^{il\lambda} = (1 - e^{i\lambda})^{-d_0}$, and $\alpha_j = \alpha(\lambda_j)$. When $d_0 \in (-\frac{1}{2}, \frac{1}{2})$, it is easy to see that $\alpha(\cdot)$ satisfies the following condition: $\alpha(\lambda)$ is differentiable in a neighborhood of the origin $(0, \epsilon)$ and also $\alpha'(\lambda) = O(|\alpha(\lambda)|\lambda^{-1})$ as $\lambda \downarrow 0$. Let $K(w) = (2\pi n)^{-1} |D(w)|^2$ be the Fejér kernel. We introduce the following working assumption.

Assumption A.1. $f_u(\cdot)$ is differentiable at $(0, \epsilon)$ for some $\epsilon > 0$ and

$$|f_u'(\lambda)| = O(\lambda^{-1}) \quad \text{as } \lambda \downarrow 0. \tag{A.1}$$

The following lemma describes the relationship between $\gamma_u(k)$ and $\|\mathcal{P}_0 u_k\|$.

LEMMA A.1. *For u_t in (7), $\sum_{k=0}^\infty k^q \|\mathcal{P}_0 u_k\| < \infty$ implies $\sum_{k\in\mathbb{Z}} |k^q \gamma_u(k)| < \infty$ for $q > 0$.*

Proof of Lemma A.1. Because $\mathcal{P}_j$ and $\mathcal{P}_{j'}$ are orthogonal if $j \neq j'$,

$$\gamma_u(k) = \mathbb{E}(u_0 u_k) = \mathbb{E}\left(\sum_{j\in\mathbb{Z}} \mathcal{P}_j u_0 \sum_{j'\in\mathbb{Z}} \mathcal{P}_{j'} u_k \right) = \sum_{j\in\mathbb{Z}} \mathbb{E}(\mathcal{P}_j u_0 \mathcal{P}_j u_k).$$

Therefore, by the Cauchy–Schwarz inequality,

$$\sum_{k\in\mathbb{Z}} |k^q \gamma_u(k)| \leq \sum_{k\in\mathbb{Z}} |k|^q \sum_{j\in\mathbb{Z}} \|\mathcal{P}_j u_0\| \|\mathcal{P}_j u_k\| = \sum_{j\in\mathbb{Z}} \|\mathcal{P}_j u_0\| \sum_{k\in\mathbb{Z}} |k|^q \|\mathcal{P}_0 u_{k-j}\|$$
$$\leq \max(2^q, 2) \left(\sum_{j=0}^\infty |j|^q \|\mathcal{P}_0 u_j\| \right)^2 < \infty.$$

■

Remark A.1. By Lemma A.1, Assumption 2.2 leads to $\sum_{k\in\mathbb{Z}} |\gamma_u(k) k| < \infty$, which implies that $f_u'(\cdot)$ is continuous on $[-\pi, \pi]$. So Assumption A.1 is satisfied. Because $\|\mathcal{P}_0 u_k\| \leq \delta_2(k)$, we have that $\sum_{k=0}^\infty k^q \delta_2(k) < \infty$ implies $\sum_{k\in\mathbb{Z}} |k^q \gamma_u(k)| < \infty$ for $q > 0$.

LEMMA A.2. *Suppose that* $\{X_t\}$ *is from (1) with* $d_0 \in (-\frac{1}{2}, \frac{1}{2})$. *Under Assumptions 2.1 and A.1, the following expressions hold uniformly over* $1 \le k < j \le m = o(n)$:

$$|\mathbb{E}\{g_j \bar{g}_j\} - 1| + |\mathbb{E}\{h_j \bar{h}_j\} - 1| + |\mathbb{E}\{g_j \bar{h}_j\} - \alpha(\lambda_j)/|\alpha_j|\,| = O(\log j/j); \tag{A.2}$$

$$\mathbb{E}\{g_j g_j\} = O(\log j/j), \qquad \mathbb{E}\{g_j g_k\} = O(\log j/k), \qquad \mathbb{E}\{g_j \bar{g}_k\} = O(\log j/k);$$

$$\mathbb{E}\{h_j h_j\} = O(\log j/j), \qquad \mathbb{E}\{h_j h_k\} = O(\log j/k), \qquad \mathbb{E}\{h_j \bar{h}_k\} = O(\log j/k);$$

$$\mathbb{E}\{g_j h_j\} = O(\log j/j), \qquad \mathbb{E}\{g_j h_k\} = O(\log j/k), \qquad \mathbb{E}\{g_j \bar{h}_k\} = O(\log j/k).$$

Furthermore, if Assumption 2.3 is satisfied, then we have

$$\operatorname{cov}(I_{uj}, I_{uk}) = f_{uj}^2 \mathbf{1}(j = k) + O(\log(j \vee k)(j \wedge k)^{-1}) \tag{A.3}$$

uniformly over $j, k = 1, \ldots, m$. *Here* $a \vee b = \max\{a, b\}$, $a \wedge b = \min\{a, b\}$.

Proof of Lemma A.2. Because the cross spectral density $f_{Xu}(\lambda) = (1 - e^{i\lambda})^{-d_0} f_u(\lambda)$ when $d_0 \in (-\frac{1}{2}, \frac{1}{2})$, it is easily seen that (A.1) implies $|f'_{Xu}(\lambda)| = O(\lambda^{-1-d_0})$ and $|f'_X(\lambda)| = O(\lambda^{-1-2d_0})$ as $\lambda \downarrow 0$. Then the lemma is a direct consequence of Robinson (1995a); see Theorem 2 and its proof therein. Regarding (A.3), we have

$$\begin{aligned}
\operatorname{cov}(I_{uj}, I_{uk}) &= \operatorname{cum}(w_{uj}, \overline{w}_{uj}, w_{uk}, \overline{w}_{uk}) + \operatorname{cov}(w_{uj}, w_{uk})\operatorname{cov}(\overline{w}_{uj}, \overline{w}_{uk}) \\
&\quad + \operatorname{cov}(w_{uj}, \overline{w}_{uk})\operatorname{cov}(\overline{w}_{uj}, w_{uk}) \\
&= O(1/n) + O(\log(j \vee k)(j \wedge k)^{-1}) + f_{uj}^2 \mathbf{1}(j = k)
\end{aligned}$$

uniformly over $j = 1, \ldots, m$ in view of Assumption 2.3. and the first assertion. ∎

Remark A.2. To show (A.2), we only need to use equation (4.1) in the paper by Robinson (1995a), whose proof does not require the rate of convergence of $f_u(\lambda)$ as $\lambda \downarrow 0$. For the remaining nine statements, our assumptions suffice. See Velasco (1999b) for a similar result when $d_0 \in [\frac{1}{2}, 1)$.

The next two lemmas (Lemmas A.3 and A.4) are useful for the consistency and the asymptotic normality of $\hat{d}$, respectively.

LEMMA A.3. *For the process (12) with* $d_0 \in (-\frac{1}{2}, \frac{1}{2})$, *suppose Assumption 2.1 holds. Then*

$$\mathbb{E}|\tilde{I}_{Xj} - \tilde{I}_{uj}| = O(j^{-1/2}) \quad \text{uniformly over } j = 1, \ldots, m. \tag{A.4}$$

Proof of Lemma A.3. Write $\phi_k = \phi_k(d_0)$. By (12),

$$\begin{aligned}
R_{nj} := w_{Xj} - w_{uj}(1 - e^{i\lambda_j})^{-d_0} &= \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t e^{it\lambda_j} - w_{uj}(1 - e^{i\lambda_j})^{-d_0} \\
&= \frac{1}{\sqrt{2\pi n}} \sum_{k=0}^{\infty} \phi_k e^{ik\lambda_j} \left\{ \sum_{t=1}^{n} u_t e^{it\lambda_j} + V_{k,j} \right\} - w_{uj}(1 - e^{i\lambda_j})^{-d_0} \\
&= \frac{1}{\sqrt{2\pi n}} \sum_{k=0}^{\infty} \phi_k e^{ik\lambda_j} V_{k,j},
\end{aligned}$$

where $V_{k,j} = \sum_{t=1-k}^{n-k} u_t e^{it\lambda_j} - \sum_{t=1}^{n} u_t e^{it\lambda_j}$. After straightforward calculations, we get

$$|1 - e^{i\lambda_j}|^{2d_0}\mathbb{E}|R_{nj}|^2 = |1 - e^{i\lambda_j}|^{2d_0}(2\pi n)^{-1}\mathbb{E}\left|\sum_{k=0}^{\infty}\phi_k e^{ik\lambda_j}V_{k,j}\right|^2$$

$$= \int_{-\pi}^{\pi} f_u(\lambda)K(\lambda + \lambda_j)\left|\frac{\alpha(-\lambda)}{\alpha(\lambda_j)} - 1\right|^2 d\lambda$$

$$\leq C_U \int_{-\pi}^{\pi} K(\lambda - \lambda_j)\left|\frac{\alpha(\lambda)}{\alpha(\lambda_j)} - 1\right|^2 d\lambda = O(1/j)$$

uniformly over $j = 1, \ldots, m$. The last equality in the preceding expression is due to Lemma 3 of Robinson (1995b) in view of the properties of $\alpha(\lambda)$. Finally, (A.4) follows from the Cauchy–Schwarz inequality and the fact that $\mathbb{E}I_{uj} = f_{uj} + o(1)$ uniformly over $j = 1, \ldots, m$ under Assumption 2.1 (cf. Brockwell and Davis, 1991, Prop. 10.3.1). ■

LEMMA A.4. *For the process (12) with $d_0 \in (-\frac{1}{2}, \frac{1}{2})$, suppose Assumptions 2.1, 2.3, and A.1 hold. Then*

$$\mathbb{E}\left|\sum_{j=1}^{r}(\tilde{I}_{Xj} - \tilde{I}_{uj})\right| \leq C(r^{1/4}(1 + \log r)^{1/2} + r^{1/2}n^{-1/4}), \qquad r \leq m = o(n),$$

where C is a generic constant independent of r, m, and n.

Proof of Lemma A.4. The proof is a generalization of the argument in Robinson (1995b, pp. 1648–1651) to the nonlinear case. Let $l = 1 + \lfloor r^{1/2}\log r\rfloor$. By Lemma A.3, $\mathbb{E}|\sum_1^l(\tilde{I}_{Xj} - \tilde{I}_{uj})| \leq Cl^{1/2}$. It then suffices to consider $l + 1 \leq j \leq r$. Write $\mathbb{E}\{\sum_{l+1}^r(\tilde{I}_{Xj} - \tilde{I}_{uj})\}^2 = (2\pi)^2(a_1 + a_2 + b_1 + b_2)$, where

$$a_1 = \sum_{l+1}^{r}\{2(\mathbb{E}|g_j|^2)^2 + |\mathbb{E}(g_j^2)|^2 - 2|\mathbb{E}(g_j h_j)|^2 - 2|\mathbb{E}(g_j\bar{h}_j)|^2$$
$$- 2\mathbb{E}|g_j|^2\mathbb{E}|h_j|^2 + 2(\mathbb{E}|h_j|^2)^2 + |\mathbb{E}(h_j^2)|^2\},$$

$$a_2 = \sum_{l+1}^{r}\{\mathrm{cum}(g_j, g_j, \bar{g}_j, \bar{g}_j) - 2\,\mathrm{cum}(g_j, h_j, \bar{g}_j, \bar{h}_j) + \mathrm{cum}(h_j, h_j, \bar{h}_j, \bar{h}_j)\},$$

$$b_1 = 2\sum_{j=l+1}^{r}\sum_{k=j+1}^{r}\{(\mathbb{E}|g_j|^2 - \mathbb{E}|h_j|^2)(\mathbb{E}|g_k|^2 - \mathbb{E}|h_k|^2)$$
$$+ |\mathbb{E}(g_j g_k)|^2 + |\mathbb{E}(g_j\bar{g}_k)|^2 - |\mathbb{E}(g_j h_k)|^2 - |\mathbb{E}(g_j\bar{h}_k)|^2$$
$$- |\mathbb{E}(g_k h_j)|^2 - |\mathbb{E}(g_k\bar{h}_j)|^2 + |\mathbb{E}(h_j h_k)|^2 + |\mathbb{E}(h_j\bar{h}_k)|^2\},$$

$$b_2 = 2\sum_{j=l+1}^{r}\sum_{k=j+1}^{r}\{\mathrm{cum}(g_j, g_k, \bar{g}_j, \bar{g}_k) - \mathrm{cum}(g_j, h_k, \bar{g}_j, \bar{h}_k)$$
$$- \mathrm{cum}(h_j, g_k, \bar{h}_j, \bar{g}_k) + \mathrm{cum}(h_j, h_k, \bar{h}_j, \bar{h}_k)\} =: 2\sum_{j=l+1}^{r}\sum_{k=j+1}^{r} s(j,k).$$

By Lemma A.2, we have $|a_1| \leq C\sum_{l+1}^{r}(\log j)/j \leq C(\log r)^2$ and

$$|b_1| \leq C \sum_{j=l+1}^{r} \sum_{k=j+1}^{r} \left\{ \frac{(\log j)(\log k)}{jk} + \frac{(\log k)^2}{j^2} \right\} \leq C \frac{r(\log r)^2}{l}.$$

Because the summand in a_2 is the summand in b_2 with $j = k$, we shall derive the order of the summand in b_2 first. After a straightforward calculation (cf. Brillinger, 1975), we can write $(2\pi n)^2 f_{uj} f_{uk} s(j,k)$ as

$$\int_{\Pi_3} \left\{ \frac{\alpha(w_1 + w_2 + w_3)\alpha(-w_2)}{|\alpha_j|^2} - 1 \right\} \left\{ \frac{\alpha(-w_1)\alpha(-w_3)}{|\alpha_k|^2} - 1 \right\} dF_{jk}(w_1, w_2, w_3), \quad \textbf{(A.5)}$$

where $\Pi_3 = [-\pi, \pi]^3$ and

$$dF_{jk}(w_1,w_2,w_3) = f_4(w_1,w_2,w_3)E_{jk}(w_1,w_2,w_3)\, dw_1\, dw_2\, dw_3,$$

$$E_{jk}(w_1,w_2,w_3) = D(\lambda_j - w_1 - w_2 - w_3)D(\lambda_k + w_1)D(w_2 - \lambda_j)D(w_3 - \lambda_k).$$

Under Assumption 2.3, $f_4(\cdot,\cdot,\cdot)$ is bounded. Following Robinson (1995b, p. 1649), (A.5) is a sum of three types of components. The first type is

$$\int_{\Pi_3} \left\{ \frac{\alpha(w_1 + w_2 + w_3)}{\alpha_j} - 1 \right\} \left\{ \frac{\alpha(-w_2)}{\bar{\alpha}_j} - 1 \right\}$$
$$\times \left\{ \frac{\alpha(-w_1)}{\alpha_k} - 1 \right\} \left\{ \frac{\alpha(-w_3)}{\bar{\alpha}_k} - 1 \right\} dF_{jk}(w_1, w_2, w_3). \quad \textbf{(A.6)}$$

By the Cauchy–Schwarz inequality, Lemma 3 of Robinson (1995b), and the periodicity, (A.6) is bounded in absolute value by $Cn^2 P_j P_k$, where

$$P_j = \int_{-\pi}^{\pi} \left| \frac{\alpha(\lambda)}{\alpha_j} - 1 \right|^2 K(\lambda - \lambda_j)\, d\lambda = O(j^{-1}) \quad \text{uniformly over } 1 \leq j \leq m.$$

So $|(\text{A.6})| \leq Cn^2 j^{-1} k^{-1}$. A typical second type of component is

$$\int_{\Pi_3} \left\{ \frac{\alpha(w_1 + w_2 + w_3)}{\alpha_j} - 1 \right\} \left\{ \frac{\alpha(-w_1)}{\alpha_k} - 1 \right\} \left\{ \frac{\alpha(-w_3)}{\bar{\alpha}_k} - 1 \right\} dF_{jk}(w_1, w_2, w_3). \quad \textbf{(A.7)}$$

Again, by the Cauchy–Schwarz inequality and the periodicity, we have $|(\text{A.7})| \leq Cn^2 P_j^{1/2} P_k \leq Cn^2 j^{-1/2} k^{-1}$ because $\int_{-\pi}^{\pi} K(\lambda)\, d\lambda = 1$. It is easily seen that other components of the second type can be bounded either by $Cn^2 j^{-1/2} k^{-1}$ or $Cn^2 j^{-1} k^{-1/2}$.

We proceed to show that the third type of component is bounded in magnitude by $Cn^{3/2} j^{-1/2} k^{-1/2}$. An example of the third type of component is

$$
\begin{aligned}
&\int_{\Pi_3}\left\{\frac{\alpha(w_1+w_2+w_3)}{\alpha_j}-1\right\}\left\{\frac{\alpha(-w_1)}{\alpha_k}-1\right\}dF_{jk}(w_1,w_2,w_3)\\
&\quad=\int_{\Pi_3}\left\{\frac{\alpha(\theta)}{\alpha_j}-1\right\}\left\{\frac{\alpha(-w_1)}{\alpha_k}-1\right\}E_{jk}(w_1,\theta-w_1-w_3,w_3)\\
&\qquad\times f_4(w_1,\theta-w_1-w_3,w_3)\,dw_1\,d\theta dw_3\\
&\quad=\int_{\Pi_2}\left\{\frac{\alpha(\theta)}{\alpha_j}-1\right\}\left\{\frac{\alpha(-w_1)}{\alpha_k}-1\right\}D(\lambda_j-\theta)D(\lambda_k+w_1)G_{jk}(\theta,w_1)\,d\theta dw_1,
\end{aligned}
\tag{A.8}
$$

where $\Pi_2=[-\pi,\pi]^2$ and

$$G_{jk}(\theta,w_1)=\int_{-\pi}^{\pi}D(\theta-w_1-w_3-\lambda_j)D(w_3-\lambda_k)f_4(w_1,\theta-w_1-w_3,w_3)\,dw_3.$$

Then we have $|(\text{A.8})|\le Cn^{3/2}P_j^{1/2}P_k^{1/2}\le Cn^{3/2}j^{-1/2}k^{-1/2}$ by the Cauchy–Schwarz inequality if the following relation holds:

$$\Omega:=\int_{\Pi_2}|G_{jk}(\theta,w_1)|^2\,d\theta dw_1\le Cn. \tag{A.9}$$

To show (A.9), let $c(k_1,k_2,k_3)=\operatorname{cum}(u_0,u_{k_1},u_{k_2},u_{k_3})$ and rewrite $G_{jk}(\theta,w_1)$ as

$$
\begin{aligned}
G_{jk}(\theta,w_1)&=\frac{1}{(2\pi)^3}\sum_{k_1,k_2,k_3\in\mathbb{Z}}\sum_{t_1,t_2=1}^{n}c(k_1,k_2,k_3)\int_{-\pi}^{\pi}e^{iw_3(k_2-k_3-t_1+t_2)}\,dw_3\\
&\qquad\times\exp\{i[-w_1k_1-k_2(\theta-w_1)+t_1(\theta-w_1-\lambda_j)-t_2\lambda_k]\}\\
&=\frac{1}{(2\pi)^2}\sum_{k_1,k_2,k_3\in\mathbb{Z}}\sum_{t_1=1}^{n}c(k_1,k_2,k_3)\mathbf{1}(1\le t_1-k_2+k_3\le n)\\
&\qquad\times\exp\{i[-w_1k_1-(\theta-w_1-\lambda_k)k_2-\lambda_kk_3+t_1(\theta-w_1-\lambda_j-\lambda_k)]\}.
\end{aligned}
$$

By a similar argument as previously, we have

$$
\begin{aligned}
(2\pi)^2\Omega&=\Bigg|\sum_{k_1,k_2,k_3\in\mathbb{Z}}\sum_{k_1',k_2',k_3'\in\mathbb{Z}}\sum_{t_1,t_1'=1}^{n}c(k_1,k_2,k_3)c(k_1',k_2',k_3')\\
&\qquad\times\mathbf{1}(1\le t_1-k_2+k_3\le n)\mathbf{1}(1\le t_1'-k_2'+k_3'\le n)\\
&\qquad\times\mathbf{1}(k_1-k_1'-k_2+k_2'+t_1-t_1'=0)\mathbf{1}(k_2'-k_2+t_1-t_1'=0)\\
&\qquad\times\exp\{i[(t_1'-t_1)\lambda_j+(t_1'-t_1+k_2-k_2'-k_3+k_3')\lambda_k]\}\Bigg|\\
&\le n\sum_{k_1,k_2,k_3\in\mathbb{Z}}\sum_{k_1',k_2',k_3'\in\mathbb{Z}}|c(k_1,k_2,k_3)|\,|c(k_1',k_2',k_3')|,
\end{aligned}
$$

which entails (A.9) under Assumption 2.3.. Finally, we deduce that

$$a_2 \le C \sum_1^r (j^{-2} + j^{-3/2} + n^{-1/2} j^{-1}) \le C,$$

$$b_2 \le C \sum_{j=l+1}^{r} \sum_{k=j+1}^{r} (j^{-1}k^{-1} + j^{-1}k^{-1/2} + j^{-1/2}k^{-1} + n^{-1/2} j^{-1/2} k^{-1/2})$$

$$\le C\{(1 + \log r)^2 + r^{1/2} \log r + rn^{-1/2}\}.$$

The conclusion follows because $l = 1 + \lfloor r^{1/2} \log r \rfloor$. ■

LEMMA A.5. *For the process X_t in (3) with $d_0 \in (\frac{1}{2}, \frac{3}{2})$, under Assumption 2.1, $\mathbb{E}(X_n - X_0)^2 \le Cn^{2d_0 - 1}$, $n \in \mathbb{N}$, where C only depends on d_0 and C_U.*

Proof of Lemma A.5. Recall that $D(w) = \sum_{t=1}^n e^{itw}$. By (3), we have

$$\mathbb{E}(X_n - X_0)^2 = \int_{-\pi}^{\pi} |D(\lambda)|^2 |1 - e^{i\lambda}|^{2-2d_0} f_u(\lambda)\, d\lambda$$

$$\le C_U \int_{-\pi}^{\pi} |1 - e^{in\lambda}|^2 |1 - e^{i\lambda}|^{-2d_0}\, d\lambda =: C_U \int_{-\pi}^{\pi} G_n(\lambda, d_0)\, d\lambda.$$

Fix a $\delta \in (0,1)$; let $h_n = \delta/n$. Then on $[0, h_n]$, $G_n(\lambda, d_0) \le C(n\lambda)^2 \lambda^{-2d_0}$. On (h_n, δ), $G_n(\lambda, d_0) \le C\lambda^{-2d_0}$ and $G_n(\lambda, d_0) \le C$ on $[\delta, \pi]$. Hence

$$\int_0^{\pi} G_n(\lambda, d_0)\, d\lambda \le \int_0^{h_n} Cn^2 \lambda^{2-2d_0}\, d\lambda + \int_{h_n}^{\delta} C\lambda^{-2d_0}\, d\lambda + \int_{\delta}^{\pi} C d\lambda \le Cn^{2d_0 - 1}.$$

The constant C only depends on d_0 and C_U once we fix δ. Thus the conclusion follows because $G_n(\lambda, d_0) = G_n(-\lambda, d_0)$. ■

LEMMA A.6. *Let $u_t = F(\dots, \varepsilon_{t-1}, \varepsilon_t)$ and $S_{nk} = n^{-1/2} \sum_{j=1}^k u_j$, $k = 1, \dots, n$. Suppose that $\mathbb{E}u_t = 0$ and $\sum_{k=1}^\infty k\delta_4(k) < \infty$; then on a richer probability space, there exists a standard Brownian motion $\mathbb{B}$ such that*

$$\max_{0 \le k \le n} |S_{nk} - \sqrt{2\pi G_0}\, \mathbb{B}(k/n)| = o_{a.s.}(\chi(n)), \tag{A.10}$$

where $\chi(n) = n^{-1/4} \log(n)$. Consequently,

$$\max_{s=1,\dots,m} |\omega_u(\lambda_s) - \xi_s| = o_{a.s.}(m\chi(n)), \tag{A.11}$$

where $\xi_s = G_0^{1/2} \sum_{k=1}^n \{\mathbb{B}(k/n) - \mathbb{B}((k-1)/n)\} e^{ik\lambda_s}$.

Proof of Lemma A.6. By Theorem 3 and Corollary 5 in Wu (2007), we have (A.10). Because

$$w_u(\lambda_s) - \xi_s = \sum_{k=1}^{n-1} (S_{nk} - \sqrt{2\pi G_0}\mathbb{B}(k/n)) \frac{e^{ik\lambda_s} - e^{i(k+1)\lambda_s}}{\sqrt{2\pi}} + \frac{S_{nn} - \sqrt{2\pi G_0}\mathbb{B}(1)}{\sqrt{2\pi}}$$

$$= o_{a.s.}(\chi(n)) \sum_{k=1}^{n-1} |1 - e^{i\lambda_s}| + o_{a.s.}(\chi(n)) = o_{a.s.}(m\chi(n))$$

uniformly over $s = 1, \ldots, m$, (A.11) follows. ■

Remark A.3. The strong approximation (A.11) is used to obtain the asymptotic distribution of $\hat{d}$ when $d_0 = 1$. A similar result for linear processes has been used by Phillips (1999b) to derive the asymptotic distribution of the estimator from LP regression at $d_0 = 1$.

LEMMA A.7. *For (3) with $d_0 \in (\frac{3}{4}, \frac{3}{2})$, assume $\sum_{k=0}^{\infty} \|\mathcal{P}_0 u_k\|_q < \infty$, where $q > 2/(2d_0 - 1)$ for $d_0 \in (\frac{3}{4}, 1)$ and $q = 2$ for $d_0 \in [1, \frac{3}{2})$. Then*

$$\frac{X_n - X_0}{n^{d_0 - 1/2} K(d_0)} \Rightarrow N(0,1), \tag{A.12}$$

where

$$K(d_0)^2 = 2\pi G_0 \Gamma(d_0)^{-2} \left\{ \frac{1}{2d_0 - 1} + \int_1^\infty (y^{d_0 - 1} - (y-1)^{d_0 - 1})^2 \, dy \right\}$$

$$= J(d_0)(2\pi)^{3 - 2d_0}(2d_0 - 1)^2 G_0 (1 - d_0)^{-1} \quad \text{for } d_0 \in (\tfrac{3}{4}, 1).$$

Proof of Lemma A.7. When $d_0 \in (\frac{3}{4}, 1) \cup (1, \frac{3}{2})$, the result follows from Theorem 2.1 in the paper by Wu and Shao (2006), who proved a functional central limit theorem. When $d_0 = 1$, it follows from Theorem 1 in Hannan (1973). ■

Proof of Proposition 3.1. Let $\eta_j = \lambda_j^{2d_0} I_{Xj}/G_0$. By Theorem 2.1 in Dalla et al. (2006), it suffices to prove that

$$\mathbb{E}(\eta_j) \le C, \qquad j = 1, \ldots, m, \quad \text{and} \quad \frac{1}{\lfloor \tau m \rfloor} \sum_{j=1}^{\lfloor \tau m \rfloor} \eta_j \xrightarrow{\mathbb{P}} 1, \quad \text{for any } \tau \in (0,1]. \tag{A.13}$$

The former assertion is a direct consequence of Lemma A.3 and the fact that $\mathbb{E} I_{uj} = f_{uj} + o(1)$ uniformly over $j = 1, \ldots, m$ under Assumption 2.1 (cf. Brockwell and Davis, 1991, Prop. 10.3.1). For the latter, let $\tau_m = \lfloor \tau m \rfloor$. By Lemma A.3,

$$\tau_m^{-1} \sum_{j=1}^{\tau_m} \eta_j = \frac{1}{G_0 \tau_m} \sum_{j=1}^{\tau_m} \tilde{I}_{Xj} f_{Xj} \lambda_j^{2d_0} = \frac{1}{G_0 \tau_m} \sum_{j=1}^{\tau_m} \tilde{I}_{uj} f_{Xj} \lambda_j^{2d_0} + o_{\mathbb{P}}(1)$$

$$= \frac{1}{G_0 \tau_m} \sum_{j=1}^{\tau_m} I_{uj} + o_{\mathbb{P}}(1) := J_n + o_{\mathbb{P}}(1).$$

We shall adopt a martingale approximation approach to prove $J_n = 1 + o_{\mathbb{P}}(1)$. Let $d_k = \sum_{t=k}^{\infty} \mathcal{P}_k u_t$ be stationary martingale differences, and we approximate $\sum_{t=1}^{n} u_t e^{it\lambda}$ by $\sum_{t=1}^{n} d_t e^{it\lambda}$. By Lemma 4 in Wu and Shao (2007), $\|\sum_{t=1}^{n}(u_t - d_t)e^{it\lambda_j}\| \le C(\sqrt{n}|1-e^{-i\lambda_j}|+1)$, where C is independent of n and λ_j. Therefore,

$$
\begin{aligned}
J_n &= \frac{1}{2\pi n G_0 \tau_m} \sum_{j=1}^{\tau_m} \left| \sum_{t=1}^{n} d_t e^{it\lambda_j} \right|^2 + o_{\mathbb{P}}(1) \\
&= \frac{1}{2\pi n G_0} \sum_{t=1}^{n} d_t^2 + \frac{1}{\pi n G_0} \sum_{k=2}^{n} \sum_{k'=1}^{k-1} d_k d_{k'} a_n(k-k') + o_{\mathbb{P}}(1) \\
&=: J_{1n} + J_{2n} + o_{\mathbb{P}}(1),
\end{aligned}
$$

where $a_n(l) = \tau_m^{-1} \sum_{j=1}^{\tau_m} \cos(l\lambda_j)$. Note that $z_k = d_k \sum_{k'=1}^{k-1} d_{k'} a_n(k-k')$ forms martingale differences with respect to $\mathcal{F}_k$ and $d_k \in \mathcal{L}^4$ under $u_t \in \mathcal{L}^4$ (see Wu and Shao, 2007, Lem. 4); we have $\mathrm{var}(J_{2n}) = O(n^{-2}) \sum_{k=2}^{n} \mathbb{E}(z_k^2) = O(1/m)$ by Burkholder's inequality (Hall and Heyde, 1980, p. 23) and Lemma 5 in Wu and Shao (2007). Thus the conclusion follows because $\|d_0\|^2 = 2\pi G_0$ (Wu and Shao, 2007) and, by the ergodic theorem, $J_{1n} - 1 = o_{a.s.}(1)$. ■

Remark A.4. In Dalla et al. (2006), their d is twice the d used in our paper. Also they set $\Delta_1 = -\frac{1}{2}$ and $\Delta_2 = \frac{1}{2}$ because they only consider the case $d_0 \in (-\frac{1}{2}, \frac{1}{2})$. A detailed check of their argument shows that their Theorem 2.1 is still applicable to our case as long as $d_0 \in [\Delta_1, \Delta_2]$.

Proof of Theorem 3.1. By Proposition 3.1, it suffices to show $\sqrt{m}F_m \Rightarrow N(0,1)$. Because $f_{Xj} G_0^{-1} \lambda_j^{2d_0} = 1 + O(\lambda_j^{\beta})$ holds uniformly over $j = 1, \ldots, m$ under Assumption 2.5 and $\sum_{j=1}^{m} |s_j| \mathbb{E}(\tilde{I}_{Xj}) \lambda_j^{\beta} = o(m^{1/2})$ under (11), we can write $\sqrt{m}F_m = I_1 + I_2 + o_{\mathbb{P}}(1)$, where

$$
I_1 = m^{-1/2} \sum_{j=1}^{m} s_j(\tilde{I}_{Xj} - \tilde{I}_{uj}) \quad \text{and} \quad I_2 = m^{-1/2} \sum_{j=1}^{m} s_j \tilde{I}_{uj}.
$$

For I_1, by summation by parts, we get from Lemma A.4 that

$$
\begin{aligned}
\mathbb{E}(|I_1|) &\le m^{-1/2} \sum_{r=1}^{m-1} \log(1+1/r) \mathbb{E} \left| \sum_{j=1}^{r} (\tilde{I}_{Xj} - \tilde{I}_{uj}) \right| + m^{-1/2} \mathbb{E} \left| \sum_{j=1}^{m} (\tilde{I}_{Xj} - \tilde{I}_{uj}) \right| \\
&\le C m^{-1/2} \sum_{r=1}^{m-1} (r^{-3/4}(1+\log r)^{1/2} + r^{-1/2} n^{-1/4}) + o(1) = o(1).
\end{aligned}
$$

Let $\hat{I}_2 = m^{-1/2} \sum_{j=1}^{m} s_j I_{uj} G_0^{-1}$, and we have $\mathbb{E}|I_2 - \hat{I}_2| = m^{-1/2} \sum_{j=1}^{m} s_j O(\lambda_j^{\beta}) = o(1)$ under (11). By Corollary 2 in Wu and Shao (2007), under Assumption 2.4 and (10), $\hat{I}_2 - \mathbb{E}(\hat{I}_2) \Rightarrow N(0,1)$. The conclusion then follows because $\mathbb{E}(\hat{I}_2) = o(1)$. ■

Proof of Proposition 3.2. A careful investigation of Theorem 2.1 of Dalla et al. (2006) and its proof suggests that their argument still applies if $d_0 \in (\frac{1}{2}, 1)$. Let $\eta_j = \lambda_j^{2d_0} I_{Xj} G_0^{-1}$. It suffices to show (A.13). By Lemma A.5, $\lambda_j^{2d_0} \|X_n - X_0\|^2 (2\pi n)^{-1} |1 - e^{i\lambda_j}|^{-2} \le C n^{2d_0 - 2} \lambda_j^{2d_0 - 2} \le C$, $1 \le j \le m$, where we have applied the fact that $|1-e^{i\lambda}|^{-2} = \lambda^{-2}(1+O(\lambda^2))$ over a neighborhood of $\lambda = 0$. Lemma A.3 and the fact that $\mathbb{E}I_{uj} = f_{uj} + o(1)$ uniformly over $j = 1, \ldots, m$ (cf. Brockwell and Davis, 1991, Prop. 10.3.1) imply that $\lambda_j^{2d_0}\mathbb{E}\{I_Y(\lambda_j)\}|1-e^{i\lambda_j}|^{-2} \le C$ for $j = 1, \ldots, m$. By the Cauchy–Schwarz inequality, the third term in (17) multiplied by $\lambda_j^{2d_0}$ is bounded uniformly over j. So $\mathbb{E}\eta_j \le C, j = 1, \ldots, m$. Next, for any $\tau \in (0,1]$, let $\tau_m = \lfloor \tau m \rfloor$,

$$\frac{1}{\tau_m}\sum_{j=1}^{\tau_m} \frac{\lambda_j^{2d_0}(X_n - X_0)^2}{2\pi n G_0|1-e^{i\lambda_j}|^2} = O_{\mathbb{P}}(n^{2d_0-2})\frac{1}{\tau_m}\sum_{j=1}^{\tau_m}\lambda_j^{2d_0-2} = o_{\mathbb{P}}(1).$$

By Lemma A.3 and the argument in the proof of Proposition 3.1,

$$\begin{aligned}\frac{1}{\tau_m}\sum_{j=1}^{\tau_m}\frac{\lambda_j^{2d_0}I_{Yj}}{G_0|1-e^{i\lambda_j}|^2} &= \frac{1}{\tau_m}\sum_{j=1}^{\tau_m}\tilde{I}_{Yj} + o_{\mathbb{P}}(1)\\ &= \frac{1}{\tau_m}\sum_{j=1}^{\tau_m}\tilde{I}_{uj} + o_{\mathbb{P}}(1) = 1 + o_{\mathbb{P}}(1).\end{aligned}$$

Finally we need to show that

$$\frac{1}{\tau_m}\sum_{j=1}^{\tau_m} G_0^{-1}\frac{2(X_n - X_0)\lambda_j^{2d_0}}{\sqrt{2\pi n}|1-e^{i\lambda_j}|^2}\operatorname{Re}(w_{Yj}e^{-i\lambda_j}) = o_{\mathbb{P}}(1). \quad \textbf{(A.14)}$$

Applying Lemma A.2 to $\{w_{Yj}\}$, we get

$$\begin{aligned}&\frac{n^{2d_0-2}}{\tau_m^2}\sum_{j,k=1}^{\tau_m}\lambda_j^{2d_0}\lambda_k^{2d_0}|1-e^{i\lambda_j}|^{-2}|1-e^{i\lambda_k}|^{-2}|\mathbb{E}\{\operatorname{Re}(w_{Yj}e^{-i\lambda_j})\operatorname{Re}(w_{Yk}e^{-i\lambda_k})\}|\\ &\quad \le \frac{C}{\tau_m^2}\sum_{j=1}^{\tau_m} j^{2d_0-2} + \frac{C}{\tau_m^2}\sum_{k=2}^{\tau_m}\sum_{j=1}^{k-1} j^{d_0-1}k^{d_0-1}j^{-1}\log k = o(1).\end{aligned}$$

Thus (A.14) holds because $X_n = O_{\mathbb{P}}(n^{d_0-1/2})$ (see Lemma A.5). The conclusion follows. ∎

Proof of Theorem 3.2. (i). By Proposition 3.2, $2(\hat{d} - d_0) = -F_m(1 + o_{\mathbb{P}}(1)) + O_{\mathbb{P}}(\log m/m)$. In view of (17), we first show that

$$m^{-1/2}\sum_{j=1}^{m} s_j\lambda_j^{2d_0}\frac{(X_n - X_0)}{\sqrt{2\pi n}|1-e^{i\lambda_j}|^2}(\overline{w}_{Yj}e^{i\lambda_j} + w_{Yj}e^{-i\lambda_j}) = o_{\mathbb{P}}(1). \quad \textbf{(A.15)}$$

Because $(X_n - X_0)/\sqrt{2\pi n} = O_{\mathbb{P}}(n^{d_0-1})$ (see Lemma A.5), it suffices to show that

$$T_m := n^{d_0-1}m^{-1/2}\sum_{j=1}^{m} s_j\lambda_j^{2d_0}\frac{(\overline{w}_{Yj}e^{i\lambda_j} + w_{Yj}e^{-i\lambda_j})}{|1-e^{i\lambda_j}|^2} = o_{\mathbb{P}}(1). \quad \textbf{(A.16)}$$

Summation by parts yields $T_m = n^{d_0-1}m^{-1/2}\{\sum_{r=1}^{m-1}(s_r - s_{r+1})D(r) + D(m)\}$, where $D(r) := \sum_{j=1}^{r}\lambda_j^{2d_0}(\overline{w}_{Yj}e^{i\lambda_j} + w_{Yj}e^{-i\lambda_j})|1-e^{i\lambda_j}|^{-2}, 1 \le r \le m$. Note that $|1-e^{i\lambda_j}|^{-2} = \lambda_j^{-2}(1+O(\lambda_j^2))$ uniformly over $j = 1, \ldots, m$. By Lemma A.2, we have

$$\|D(r)\|^2 \le \sum_{j,k=1}^{r} \lambda_j^{2d_0-2}\lambda_k^{2d_0-2}(1+O(\lambda_j^2))(1+O(\lambda_k^2))$$

$$\times |\mathbb{E}(\overline{w}_{Yj}e^{i\lambda_j} + w_{Yj}e^{-i\lambda_j})(\overline{w}_{Yk}e^{i\lambda_k} + w_{Yk}e^{-i\lambda_k})|$$

$$\le C\sum_{j=1}^{r} \lambda_j^{2(d_0-1)} + C\sum_{k=2}^{r}\sum_{j=1}^{k-1} \lambda_j^{d_0-1}\lambda_k^{d_0-1}\frac{\log k}{j} \le Cn^{2-2d_0}r^{d_0}\log r.$$

Because $d_0 < 1$,

$$\mathbb{E}(|T_m|) \le n^{d_0-1}m^{-1/2}\left\{\sum_{r=1}^{m-1} |s_r - s_{r+1}|\,\|D(r)\| + \|D(m)\|\right\}$$

$$\le Cm^{-1/2}\left\{\sum_{r=1}^{m-1} r^{-1}r^{d_0/2}\sqrt{\log r} + m^{d_0/2}\sqrt{\log m}\right\} = o(1);$$

hence (A.16) holds. When $d_0 \in (\frac{1}{2},\frac{3}{4})$, by the argument in the proof of Theorem 3.1. (where the role of $\lambda_j^{2d_0}I_{Xj}$ is replaced by $\lambda_j^{2d_0-2}I_{Yj}$),

$$\sqrt{m}(\hat{d} - d_0) = -m^{-1/2}\sum_{j=1}^{m} s_j\lambda_j^{2d_0-2}I_{Yj}(2G_0)^{-1} + o_{\mathbb{P}}(1) \Rightarrow N(0,\tfrac{1}{4}).$$

When $d_0 \in (\frac{3}{4},1)$,

$$m^{2-2d_0}(\hat{d} - d_0) = -m^{1-2d_0}\sum_{1}^{m} s_j\lambda_j^{2d_0-2}\frac{(X_n - X_0)^2}{4\pi nG_0} + o_{\mathbb{P}}(1) \Rightarrow J(d_0)U_2^2$$

in view of Lemma A.7 and the fact that $m^{1-2d_0}\sum_1^m s_j j^{2d_0-2} = (2d_0-1)^{-2}(2d_0-2) + o(1)$.

(ii). Denote $a(e^{i\lambda}) = \sum_{j=0}^{\infty} a_j e^{ij\lambda}$. Recall (15) for the expression of $X_n - X_0$. We first claim that

$$X_n - X_0 - a(1)\sum_{j=0}^{\infty} A_{j,n}\varepsilon_{n-j} = o_{\mathbb{P}}(n^{d_0-1/2}). \qquad \textbf{(A.17)}$$

Let $\Omega_1 = (2\pi)^{2d_0-2}(2d_0-1)^{-2}(1-d_0)$, $y_t = n^{1/2-d_0}A_{n-t,n}\varepsilon_t = n^{1/2-d_0}\phi_{n-t}(d_0)\varepsilon_t$, $1 \le t \le n$, and $P_n = n^{1/2-d_0}\sum_{k=n}^{\infty} A_{k,n}\varepsilon_{n-k}$. By (17) and (A.15), we have $\sqrt{m}F_m = I_1 + I_2 + o_{\mathbb{P}}(1)$, where

$$I_1 = m^{-1/2}\sum_{1}^{m} G_0^{-1}s_j\lambda_j^{2d_0}|1-e^{i\lambda_j}|^{-2}\frac{(X_n - X_0)^2}{2\pi n}$$

$$= (2\pi n)^{-1}m^{-1/2}\sum_{1}^{m} G_0^{-1}s_j\lambda_j^{2d_0-2}a(1)^2\left(\sum_{k=0}^{\infty} A_{k,n}\varepsilon_{n-k}\right)^2(1+o_{\mathbb{P}}(1))$$

$$= -2\Omega_1\left(\sum_{1}^{n} y_t + P_n\right)^2(1+o_{\mathbb{P}}(1))$$

and

$$I_2 = m^{-1/2} \sum_1^m s_j \lambda_j^{2d_0} I_Y(\lambda_j) G_0^{-1} |1 - e^{i\lambda_j}|^{-2}$$

$$= m^{-1/2} \sum_1^m s_j \lambda_j^{2d_0-2} I_Y(\lambda_j) G_0^{-1} + o_{\mathbb{P}}(1) = m^{-1/2} \sum_1^m s_j I_{uj} G_0^{-1} + o_{\mathbb{P}}(1).$$

The validity of the last equality in the preceding expression follows from Lemma A.4 and the argument in the proof of Theorem 3.1. Under the assumption $\sum_{j=0}^{\infty} j|a_j| < \infty$, it is easy to see that $a(e^{i\lambda})$ is differentiable in a neighborhood of zero and $(d/d\lambda)a(e^{i\lambda}) = O(\lambda^{-1})$ as $\lambda \to 0+$. Based on the preceding properties of $\alpha(\cdot)$, Robinson (1995b, p. 1644) showed that $\sum_1^m s_j I_{uj} G_0^{-1} = \sum_1^m 2\pi s_j I_\varepsilon(\lambda_j) + o_{\mathbb{P}}(m^{1/2})$. Therefore $I_2 = m^{-1/2} \sum_1^m s_j 2\pi I_\varepsilon(\lambda_j) + o_{\mathbb{P}}(1) = \sum_{t=1}^n z_t + o_{\mathbb{P}}(1)$, where $z_1 = 0$, $z_t = \varepsilon_t \sum_{s=1}^{t-1} \varepsilon_s c_{t-s}$ for $t \geq 2$. Here $c_s = 2n^{-1} m^{-1/2} \sum_1^m s_j \cos(s\lambda_j)$. Letting $\Omega_2 = \Gamma(d_0)^{-2}(2d_0 - 1)^{-1}$ and $\Omega_3 = \Gamma(d_0)^{-2}\{\int_1^\infty (y^{d_0-1} - (y-1)^{d_0-1})^2\, dy\}$, we shall show that

$$\left(\sum_1^n z_t, \sum_1^n y_t, P_n \right) \Rightarrow (U_1, \Omega_2^{1/2} W_1, \Omega_3^{1/2} W_2), \tag{A.18}$$

where U_1, W_1, and W_2 are i.i.d. $N(0,1)$. Notice that P_n is independent of $(\sum_1^n y_t, \sum_1^n z_t)$; it suffices to show that $P_n \Rightarrow \Omega_3^{1/2} W_2$ because PS (2004, p. 687) have shown that $(\sum_1^n z_t, \sum_1^n y_t) \Rightarrow (U_1, \Omega_2^{1/2} W_1)$. Write

$$P_n = n^{1/2-d_0} \sum_{k=n}^{n^2-1} A_{k,n} \varepsilon_{n-k} + n^{1/2-d_0} \sum_{k=n^2}^{\infty} A_{k,n} \varepsilon_{n-k} =: K_{1n} + K_{2n}.$$

Note that $|A_{k,n}| \leq \kappa := \sum_{j=0}^\infty |\phi_j(d_0 - 1)| < \infty$ for all $k \geq n$. Then for any $\epsilon > 0$, the following Lindeberg condition holds:

$$\sum_{k=n}^{n^2-1} \mathbb{E}\{A_{k,n}^2 n^{1-2d_0} \varepsilon_{n-k}^2 \mathbf{1}(n^{1/2-d_0} |A_{k,n} \varepsilon_{n-k}| > \epsilon)\}$$

$$= O(n^{1-2d_0}) \sum_{k=n+1}^{n^2-1} [k^{d_0-1} - (k-n)^{d_0-1}]^2 \times \mathbb{E}\{\varepsilon_1^2 \mathbf{1}(|\varepsilon_1| > n^{d_0-1/2}\epsilon/\kappa)\} \to 0,$$

which entails that $K_{1n}[\mathrm{var}(K_{1n})]^{-1/2} \Rightarrow N(0,1)$. Because K_{1n} and K_{2n} are independent of each other, the convergence of P_n follows from the fact that $\mathrm{var}(K_{2n}) = O(\int_n^\infty (y^{d_0-1} - (y-1)^{d_0-1})^2\, dy) = o(1)$.

Combining the preceding results, we apply the continuous mapping theorem and get $\sqrt{m}(\hat{d} - d_0) \Rightarrow -U_1/2 + \Omega_1(\Omega_2^{1/2} W_1 + \Omega_3^{1/2} W_2)^2$, where the limit has the same distribution as $U_1/2 + J(d_0) U_2^2$.

It remains to show (A.17). By Beveridge-Nelson decomposition (Phillips and Solo, 1992),

$$u_j = a(1)\varepsilon_j + \tilde{\varepsilon}_{j-1} - \tilde{\varepsilon}_j, \qquad \tilde{\varepsilon}_j = \sum_{i=0}^\infty \tilde{a}_i \varepsilon_{j-i}, \qquad \tilde{a}_i = \sum_{k=i+1}^\infty a_k.$$

Letting $B_{j,n} = A_{j,n} - A_{j+1,n}$, then we have

$$X_n - X_0 - a(1)\sum_{j=0}^{\infty} A_{j,n}\varepsilon_{n-j} = \sum_{j=0}^{\infty} A_{j,n}(\tilde{\varepsilon}_{n-j-1} - \tilde{\varepsilon}_{n-j})$$

$$= \sum_{j=0}^{\infty} B_{j,n}\tilde{\varepsilon}_{n-j-1} - A_{0,n}\tilde{\varepsilon}_n,$$

where $\|A_{0,n}\tilde{\varepsilon}_n\| < \infty$ and

$$\left\|\sum_{j=0}^{\infty} B_{j,n}\tilde{\varepsilon}_{n-j-1}\right\|^2 \le \int_{-\pi}^{\pi}\left|\sum_{j=0}^{\infty} B_{j,n}e^{ij\lambda}\right|^2\left|\sum_{j=0}^{\infty}\tilde{a}_j e^{ij\lambda}\right|^2 \frac{1}{2\pi}\,d\lambda$$

$$\le 4\left(\sum_{i=0}^{\infty}|\phi_i(d_0-1)|\sum_{j=0}^{\infty}|\tilde{a}_j|\right)^2 < \infty.$$

Thus (A.17) holds, and the proof is completed. ■

Remark A.5. As we can see from the proof of Theorem 3.2, the third term of (17) is negligible for all $d_0 \in (\frac{1}{2}, 1)$. The first term of (17) becomes dominant when $d_0 \in (\frac{3}{4}, 1)$, whereas the second term is dominant when $d_0 \in (\frac{1}{2}, \frac{3}{4})$. At $d_0 = \frac{3}{4}$, the first two terms have the same stochastic order, and the asymptotic distribution is a mixture. This phenomenon has been observed by PS (2004) in the Type II case, and our result for the Type I case is consistent with their observation. When $d_0 \in (\frac{3}{4}, 1)$, Assumption 2.4 can be replaced by $\sum_{k=1}^{\infty} k\|\mathcal{P}_0 u_k\|_q < \infty$, $q > 2/(2d_0 - 1)$; compare Lemma A.7 and Assumption 2.2.

Proof of Theorem 3.3. The proof follows the argument in PS (2004) for the Type II case, Lemmas A.2, A.3, A.5, and A.7. For the details, see Shao and Wu (2005). ■

The following lemma is used in proving the asymptotic distribution of $\hat{d}$ when $d_0 = 1$.

LEMMA A.8. *Let the nonnegative random variables* $\eta_j = \eta_{j,m}$, $j = 1, \ldots, m$ *satisfy* $\mathbb{E}\eta_j \le C, 1 \le j \le m$, *where C is a finite constant. Suppose that the random variables* $Q_m(x,d), m \ge 1$ *satisfy that for any* $b \in (0,1)$, $d \in [d_1, d_2] \subset \mathbb{R}$,

$$\sup_{b\le x\le 1}\sup_{d\in[d_1,d_2]}|Q_m(x,d)| = o_{\mathbb{P}}(1), \quad \text{as } m\to\infty \qquad \textbf{(A.19)}$$

and for some $\gamma \in (0,1)$,

$$\sup_{d\in[d_1,d_2]}|Q_m(x,d)| \le Cx^{-\gamma}|\log(x)[\log(x)+1]|, \qquad x \in [0,b]. \qquad \textbf{(A.20)}$$

Then as $m \to \infty$,

$$\sup_{d\in[d_1,d_2]}\left|\frac{1}{m}\sum_{j=1}^{m} Q_m(j/m,d)\eta_j\right| = o_{\mathbb{P}}(1).$$

Proof of Lemma A.8. Let $b_m = \lfloor bm \rfloor$; then we have

$$\left| \frac{1}{m} \sum_{j=1}^{m} Q_m(j/m, d)\eta_j \right| \le \left(\sum_{j=1}^{b_m} + \sum_{j=b_m+1}^{m} \right) \left| \frac{1}{m} Q_m(j/m, d)\eta_j \right| =: I_{1m}(d) + I_{2m}(d).$$

By (A.19), $\sup_{d\in[d_1,d_2]} I_{2m}(d) \le \sup_{b\le x\le 1} \sup_{d\in[d_1,d_2]} |Q_m(x,d)| m^{-1} \sum_{j=b_m+1}^{m} \eta_j = o_{\mathbb{P}}(1)$. Under (A.20), $\mathbb{E} \sup_{d\in[d_1,d_2]} I_{1m}(d) \le Cm^{-1} \sum_{j=1}^{b_m} (j/m)^{-\gamma} |\log(j/m)(\log(j/m)+1)| \to 0$ as $b \downarrow 0$. The conclusion follows. ■

Proof of Theorem 3.4. The following argument is a variant of the one used by Dalla et al. (2006) in their proof of Theorem 2.1. Again, we assume $X_0 = 0$ for the convenience of presentation. Let $v_j = \log j - m^{-1} \sum_{k=1}^{m} \log k$ and $\eta_j = \lambda_j^{2d_0} G_0^{-1} I_{Xj}$. Recall (8) for the objective function $R(d)$. Note that $R'(d) = 2T_n(d)/S_n(d)$, where

$$T_n(d) = \frac{1}{m} \sum_{j=1}^{m} (j/m)^{2(d-d_0)} v_j \eta_j \quad \text{and} \quad S_n(d) = \frac{1}{m} \sum_{j=1}^{m} (j/m)^{2(d-d_0)} \eta_j.$$

Fix an $\epsilon \in (0, \frac{1}{4})$. On $\Omega_{1n} := \{|\hat{d} - d_0| \le \epsilon\}$, we have

$$S_n(\hat{d}) \ge \frac{1}{m} \sum_{1}^{m} \frac{j^{2\epsilon}\eta_j}{m^{2\epsilon}} = \frac{1}{m} \sum_{1}^{m} \frac{j^{2\epsilon}}{m^{2\epsilon}} \left(1 + \frac{X_n^2}{2\pi n G_0}\right) + o_{\mathbb{P}}(1) \ge \frac{1}{2\epsilon + 1} + o_{\mathbb{P}}(1),$$

which yields $T_n(\hat{d}) = R'(\hat{d}) S_n(\hat{d})/2 = 0$ on $\Omega_{1n} \cap \Omega_{2n}$, where $\Omega_{2n} = \{S_n(\hat{d}) \ge 1/(2\epsilon + 2)\}$ and $\mathbb{P}(\Omega_{2n}) \to 1$ as $n \to \infty$. By the mean-value theorem, there exists a $\bar{d}$ that lies in between $\hat{d}$ and d_0 such that $T_n(\hat{d}) - T_n(d_0) = (\hat{d} - d_0) T_n'(\bar{d})$ on $\Omega_{1n} \cap \Omega_{2n}$, where $T_n'(\bar{d}) = 2m^{-1} \sum_{1}^{m} (j/m)^{2(\bar{d}-d_0)} \log(j/m) v_j \eta_j$. Assume that

$$T_n'(\bar{d}) - 2[1 + X_n^2/(2\pi n G_0)] = o_{\mathbb{P}}(1) \quad \text{on } \Omega_{1n} \cap \Omega_{2n}. \tag{A.21}$$

The proof of (A.21) is given at the end of this proof. Recall (A.11) for the form of $\xi_j, 1 \le j \le m$. We apply Lemma A.5 and (17) and get

$$\begin{aligned}
\sqrt{m} T_n(d_0) &= \frac{1}{\sqrt{m}} \sum_{1}^{m} v_j \eta_j \\
&= \frac{1}{\sqrt{m}} \sum_{1}^{m} \frac{v_j}{G_0} \left[(1 + O(\lambda_j^2)) \left\{ I_{uj} + \frac{X_n^2}{2\pi n} - \frac{2X_n \operatorname{Re}(w_{uj} e^{-i\lambda_j})}{\sqrt{2\pi n}} \right\} \right] \\
&= \frac{1}{\sqrt{m}} \sum_{1}^{m} \frac{v_j}{G_0} \left\{ I_{uj} - \frac{2X_n \operatorname{Re}(w_{uj} e^{-i\lambda_j})}{\sqrt{2\pi n}} \right\} + o_{\mathbb{P}}(m^{5/2} n^{-2} \log m) \\
&= \frac{1}{\sqrt{m}} \sum_{1}^{m} \frac{v_j |\xi_j|^2}{G_0} - \frac{I\!B(1)}{\sqrt{m}} \sum_{1}^{m} \frac{v_j}{\sqrt{G_0}} (\xi_j e^{-i\lambda_j} + \bar{\xi}_j e^{i\lambda_j}) + o_{\mathbb{P}}(1).
\end{aligned}$$

The last equality in the preceding expression is due to Lemma A.6 and (19). By Theorem 3.3, $(\hat{d} - d_0)(1 + o_{\mathbb{P}}(1)) = (\hat{d} - d_0)\mathbf{1}(\Omega_{1n} \cap \Omega_{2n})$. Because $X_n^2/(2\pi n) = G_0 I\!B(1)^2 + o_{\mathbb{P}}(1)$,

$$\sqrt{m}(\hat{d} - d_0)(1 + o_{\mathbb{P}}(1)) = \frac{-Z_n^{(1)} + I\!B(1)Z_n^{(2)} + o_{\mathbb{P}}(1)}{2\{1 + I\!B(1)^2\} + o_{\mathbb{P}}(1)},$$

where $Z_n^{(1)} = m^{-1/2}\sum_1^m v_j|\xi_j|^2 G_0^{-1}$ and $Z_n^{(2)} = m^{-1/2}\sum_1^m v_j(\xi_j e^{-i\lambda_j} + \bar{\xi}_j e^{i\lambda_j})G_0^{-1/2}$. We shall show that $(Z_n^{(1)}, Z_n^{(2)}) \Rightarrow (W_1, \sqrt{2}W_2)$. By the Cramer–Wold device, it suffices to show that, for any $\alpha \in (0,1)$,

$$m^{-1/2}\sum_1^m v_j z_j \Rightarrow \alpha W_1 + \sqrt{2}(1-\alpha)W_2, \tag{A.22}$$

where $z_j = \alpha G_0^{-1}|\xi_j|^2 + (1-\alpha)G_0^{-1/2}(\xi_j e^{-i\lambda_j} + \bar{\xi}_j e^{i\lambda_j})$, $j = 1,\ldots,m$ are independent random variables. Note that $\sum_1^m v_j = 0$, $m^{-1}\sum_1^m v_j^2 = 1 + o(1)$, and $\operatorname{var}(m^{-1/2}\sum_1^m v_j z_j) \to \alpha^2 + 2(1-\alpha)^2$. Then the convergence (A.22) follows from the obvious Lindeberg condition, i.e., for any $\delta > 0$,

$$\frac{1}{m}\sum_{j=1}^m v_j^2 \mathbb{E}\{z_j^2 \mathbf{1}(|v_j z_j| \geq \delta\sqrt{m})\} \leq \frac{1}{m}\sum_{j=1}^m v_j^2 \mathbb{E}\{z_1^2 \mathbf{1}(|z_1| \geq \delta\sqrt{m}/(2\log m))\} \to 0.$$

It is easily seen that $I\!B(1)$ is independent of ξ_j, $j = 1,\ldots,m$, and thus independent of $Z_n^{(1)}$ and $Z_n^{(2)}$. Applying the continuous mapping theorem, we get

$$\sqrt{m}(\hat{d} - d_0) \Rightarrow \frac{-W_1 + \sqrt{2}W_2 W_3}{2(1 + W_3^2)}. \tag{A.23}$$

It remains to show (A.21). To that end, write

$$\begin{aligned} T_n'(\bar{d})/2 &= H_{1n} + H_{2n} \\ &= \frac{1}{m}\sum_1^m \log(j/m) v_j \eta_j + \frac{1}{m}\sum_1^m [(j/m)^{2(\bar{d}-d_0)} - 1]\log(j/m) v_j \eta_j. \end{aligned}$$

Because $v_j = \log(j/m) + 1 + O(\log m/m)$ uniformly over $j = 1,\ldots,m$ (see Robinson, 1995b, Lem. 2) and $X_n^2/n = O_{\mathbb{P}}(1)$ (see Lemma A.5), we have

$$\begin{aligned} H_{1n} - 1 - \frac{X_n^2}{2\pi n G_0} &= \frac{1}{m}\sum_1^m \log\left(\frac{j}{m}\right) v_j(\eta_j - 1 - X_n^2(2\pi n G_0)^{-1}) + o_{\mathbb{P}}(1) \\ &= H_{1n}^{(1)} + H_{1n}^{(2)} + H_{1n}^{(3)} + o_{\mathbb{P}}(1), \end{aligned}$$

where by (17),

$$H_{1n}^{(1)} = \frac{1}{m}\sum_1^m \log\left(\frac{j}{m}\right) v_j X_n^2 (2\pi n G_0)^{-1}(\lambda_j^2|1 - e^{i\lambda_j}|^{-2} - 1),$$

$$H_{1n}^{(2)} = \frac{1}{m}\sum_1^m \log\left(\frac{j}{m}\right) v_j\left(\frac{I_{uj}\lambda_j^2}{G_0|1 - e^{i\lambda_j}|^2} - 1\right),$$

$$H_{1n}^{(3)} = -\frac{1}{m}\sum_1^m \log\left(\frac{j}{m}\right)\frac{2v_j\lambda_j^2 \operatorname{Re}(w_{uj}e^{-i\lambda_j})}{|1 - e^{i\lambda_j}|^2 G_0}\frac{X_n}{\sqrt{2\pi n}}.$$

By Lemma A.5, $H_{1n}^{(1)} = o_{\mathbb{P}}(1)$. With regard to $H_{1n}^{(2)}$, we have

$$H_{1n}^{(2)} = o_{\mathbb{P}}(1) + m^{-1}\sum_{1}^{m}\log(j/m)v_j(I_{uj} - \mathbb{E}I_{uj})G_0^{-1},$$

where the variance of the latter term is $o(1)$ by (A.3) of Lemma A.2. That $H_{1n}^{(3)} = o_{\mathbb{P}}(1)$ follows from Lemma A.2 and a similar argument as before; compare the proof of Theorem 3.3.

By Lemmas A.2 and A.5, there exists a finite constant C such that $\mathbb{E}\eta_j \le C$, $j = 1, \ldots, m$. Therefore, on $\Omega_{1n} \cap \Omega_{2n}$, $H_{2n} = (1/m)\sum_1^m[(j/m)^{2(\bar{d}-d_0)} - 1]\log(j/m)\{\log(j/m) + 1\}\eta_j + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)$ follows from Theorem 3.3 and Lemma A.8 by setting $Q_m(x,d) = [x^{2(d-d_0)} - 1]\log(x)\{\log(x) + 1\}$, $d_1 = d_0 - \epsilon$, and $d_2 = d_0 + \epsilon$. Thus the conclusion holds. ■

Proof of Theorem 4.1. Let $J = \text{cum}(u_0, u_{m_1}, \ldots, u_{m_{k-1}})$, where $0 = m_0 \le m_1 \le \cdots \le m_{k-1}$. Denote by $n_l = m_l - m_{l-1}$, $1 \le l \le k-1$. Define the random vector $Y_0 = Y_{0,l} = (u_{m_0 - m_{l-1}}, \ldots, u_{m_{l-2}-m_{l-1}}, u_0)$. Further let $(\varepsilon_n')_{n\in\mathbb{Z}}$ be an i.i.d. copy of $(\varepsilon_n)_{n\in\mathbb{Z}}$, $\Omega_n = (\ldots, \varepsilon_{n-1}', \varepsilon_n')$ and define $u_t^* = F(\Omega_0, \varepsilon_1, \ldots, \varepsilon_t)$, $t \ge 0$. Following Proposition 2 in Wu and Shao (2004), by the stationarity and the additivity of cumulants,

$$\begin{aligned} J &= \text{cum}(Y_0, u_{m_l - m_{l-1}}, u_{m_{l+1}-m_{l-1}}, \ldots, u_{m_{k-1}-m_{l-1}}) \\ &= \sum_{j=0}^{k-l-1} \text{cum}(Y_0, u^*_{m_l - m_{l-1}}, \ldots, u^*_{m_{l+j-1}-m_{l-1}}, u_{m_{l+j}-m_{l-1}} - u^*_{m_{l+j}-m_{l-1}}, \\ &\qquad u_{m_{l+j+1}-m_{l-1}}, \ldots, u_{m_{k-1}-m_{l-1}}) =: \sum_{j=0}^{k-l-1} B_j. \end{aligned}$$

Denote by $\zeta_j = \|u_j - u_j^*\|_k$ and $S_j = S_{j,k} := \sum_{i=j}^{\infty}\delta_k(i)^2$. Proposition 2 of Wu and Shao (2004) asserts that $|B_j| \le C_1\zeta_{m_{l+j}-m_{l-1}}$, where C_1 only depends on k and the moments $\mathbb{E}|u_0|^i$, $1 \le i \le k$. Therefore, by Proposition 2 of Wu (2007),

$$J \le C_1 \sum_{j=0}^{k-l-1} \zeta_{m_{l+j}-m_{l-1}} \le C_2 \sum_{j=0}^{k-l-1} S^{1/2}_{m_{l+j}-m_{l-1}} \le C_3 S^{1/2}_{n_l},$$

where $C_2 = 18k^{3/2}(k-1)^{-1/2}C_1$, $C_3 = C_2 k$. Because the preceding expression holds for any l, $1 \le l \le k-1$, we have $J \le C_3 \min_{1\le l\le k-1} S^{1/2}_{n_l}$. Finally,

$$\begin{aligned} \sum_{m_1,\ldots,m_{k-1}\in\mathbb{Z}} |\text{cum}(u_0, u_{m_1}, \ldots, u_{m_{k-1}})| &\le k! \sum_{0\le m_1\le\cdots\le m_{k-1}} |\text{cum}(u_0, u_{m_1}, \ldots, u_{m_{k-1}})| \\ &\le k! \sum_{n_1=0}^{\infty}\cdots\sum_{n_{k-1}=0}^{\infty} C_3 \min_{1\le l\le k-1} S^{1/2}_{n_l} \\ &\le C_3 k! \sum_{n=0}^{\infty} n^{k-2} S_n^{1/2} < \infty \end{aligned}$$

by our assumption. ■

Proof of Proposition 4.1. Write (25) into the following form:

$$\rho_t := \zeta_t^2 = \sigma_t^2 \varepsilon_t^2, \qquad \sigma_t^2 = \psi_0 + \sum_{j=1}^{\infty} \psi_j \zeta_{t-j}^2.$$

Note that $\|\varepsilon_1^2\|_3^{1/2} \sum_{j=1}^{\infty} \psi_j^{1/2} < 1$ implies $\|\varepsilon_1^2\| \sum_{j=1}^{\infty} \psi_j < 1$. By Theorem 2.1 of Giraitis, Kokoszka, and Leipus (2000), there exists a unique strictly stationary solution, and $\rho_t \in \mathcal{L}^2$. Denote it by $\zeta_t = G(\ldots, \varepsilon_{t-1}, \varepsilon_t)$ and let $\zeta_t' = G(\ldots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_t)$ for $t \in \mathbb{N}$. Note that

$$\begin{aligned}\rho_t^3 &= \varepsilon_t^6 \left(\psi_0 + \sum_{j=1}^{\infty} \psi_j \rho_{t-j} \right)^3 \\ &= \varepsilon_t^6 \left\{ \psi_0^3 + 3\psi_0^2 \sum_{j=1}^{\infty} \psi_j \rho_{t-j} + 3\psi_0 \left(\sum_{j=1}^{\infty} \psi_j \rho_{t-j} \right)^2 + \left(\sum_{j=1}^{\infty} \psi_j \rho_{t-j} \right)^3 \right\}.\end{aligned}$$

By the Cauchy–Schwarz inequality, we have $\mathbb{E}\rho_t^3 \le [1 - \mathbb{E}\varepsilon_t^6 (\sum_{j=1}^{\infty} \psi_j)^3]^{-1} \mathbb{E}\varepsilon_t^6 [\psi_0^3 + 3\psi_0^2 \sum_{j=1}^{\infty} \psi_j \mathbb{E}\rho_t + 3\psi_0 (\sum_{j=1}^{\infty} \psi_j)^2 \mathbb{E}\rho_t^2] < \infty$, i.e., $\rho_t \in \mathcal{L}^3$. For $k \in \mathbb{N}$, let $\rho_k' := (\zeta_k')^2 = \varepsilon_k^2 \{\psi_0 + \sum_{j=1}^{k} \psi_j (\zeta_{k-j}')^2 + \sum_{j=k+1}^{\infty} \psi_j \zeta_{k-j}^2\}$. Thus we get

$$\rho_k - \rho_k' = \varepsilon_k^2 \sum_{j=1}^{k} \psi_j [\rho_{k-j} - \rho_{k-j}'].$$

Let $s_k := \|\rho_k - \rho_k'\|_3$, $\alpha_j = \psi_j \|\varepsilon_1^2\|_3$; then $s_k \le \sum_{j=1}^{k} \alpha_j s_{k-j}$, which implies $\sqrt{s_k} \le \sum_{j=1}^{k} \sqrt{\alpha_j} \sqrt{s_{k-j}}$. It suffices to show that $\sum_{k=1}^{\infty} k^3 s_k^{1/2} < \infty$ because it implies

$$\sum_{k=1}^{\infty} k \|\zeta_k - \zeta_k'\|_6 \le \sum_{k=1}^{\infty} k s_k^{1/2} < \infty$$

and

$$\sum_{k=1}^{\infty} k^3 \|\zeta_k - \zeta_k'\|_4 \le \sum_{k=1}^{\infty} k^3 \|\rho_k - \rho_k'\|^{1/2} \le \sum_{k=1}^{\infty} k^3 s_k^{1/2} < \infty, \tag{A.24}$$

where the latter implies the first assertion of (26) in view of Remark 4.1. In the preceding discussion, we have applied the fact that $(\zeta_k - \zeta_k')^2 = \varepsilon_k^2 (\sigma_k - \sigma_k')^2 \le \varepsilon_k^2 |\sigma_k^2 - (\sigma_k')^2| = |\rho_k - \rho_k'|$, where σ_k' is the coupled version of σ_k with ε_0 replaced by ε_0'. Let $\tilde{s}_0 = s_0$ and $\sqrt{\tilde{s}_k} = \sum_{j=1}^{k} \sqrt{\alpha_j} \sqrt{\tilde{s}_{k-j}}$; then $s_k \le \tilde{s}_k$. Define $g(z) = \sum_{j=1}^{\infty} \sqrt{\alpha_j} z^j$ and $h(z) = \sum_{j=0}^{\infty} \sqrt{\tilde{s}_j} z^j$; then $h(z) = s_0/(1 - g(z))$. Note that $g(1) < 1$. Let $h^{(p)}(z)$ be the pth derivative of $h(z)$. A simple calculation shows that $h^{(p)}(1)$, $p = 1,2,3$ are all finite under $\sum_{j=1}^{\infty} j^3 \psi_j^{1/2} < \infty$. Thus $\sum_{k=1}^{\infty} k^3 \sqrt{s_k} \le \sum_{k=1}^{\infty} k^3 \sqrt{\tilde{s}_k} < \infty$. The conclusion follows. ■

Proof of Proposition 4.2. Note that $\|\varepsilon_1\|_5 \sum_{j=1}^{\infty} |b_j| < 1$ implies $\|\varepsilon_1\| \{\sum_{j=1}^{\infty} b_j^2\}^{1/2} < 1$. It follows from Theorem 2.1 of Giraitis, Robinson, and Surgailis (2000) that there exists a strictly stationary solution $\zeta_t = G(\ldots, \varepsilon_{t-1}, \varepsilon_t)$ and $\zeta_t \in \mathcal{L}^2$. By the same argument as in the proof of Proposition 4.1, $\zeta_t \in \mathcal{L}^5$. For $k \in \mathbb{N}$, $\zeta_k - \zeta_k' = \varepsilon_k \sum_{j=1}^{k} b_j (\zeta_{k-j} - \zeta_{k-j}')$. The conclusion follows from a similar argument as in the proof of Proposition 4.1, and we omit the details. ■