Local Polynomial Whittle Estimation of Long-range Dependence

By

Donald W.K. Andrews and Yixiao Sun

January 2001

COWLES FOUNDATION DISCUSSION PAPER NO. 1293

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

http://cowles.econ.yale.edu/

Local Polynomial Whittle Estimation of Long-range Dependence[1]

Donald W. K. Andrews
Cowles Foundation for Research in Economics
Yale University

Yixiao Sun
Department of Economics
Yale University

November 1999
Revised: January 2001

Abstract

The local Whittle (or Gaussian semiparametric) estimator of long range dependence, proposed by Künsch (1987) and analyzed by Robinson (1995a), has a relatively slow rate of convergence and a finite sample bias that can be large. In this paper, we generalize the local Whittle estimator to circumvent these problems. Instead of approximating the short-run component of the spectrum, $\varphi(\lambda)$, by a constant in a shrinking neighborhood of frequency zero, we approximate its logarithm by a polynomial. This leads to a "local polynomial Whittle" (LPW) estimator.

Following the work of Robinson (1995a), we establish the asymptotic bias, variance, mean-squared error (MSE), and normality of the LPW estimator. We determine the asymptotically MSE-optimal bandwidth, and specify a plug-in selection method for its practical implementation. When $\varphi(\lambda)$ is smooth enough near the origin, we find that the bias of the LPW estimator goes to zero at a faster rate than that of the local Whittle estimator, and its variance is only inflated by a multiplicative constant. In consequence, the rate of convergence of the LPW estimator is faster than that of the local Whittle estimator, given an appropriate choice of the bandwidth m.

We show that the LPW estimator attains the optimal rate of convergence for a class of spectra containing those for which $\varphi(\lambda)$ is smooth of order $s \geq 1$ near zero. When $\varphi(\lambda)$ is infinitely smooth near zero, the rate of convergence of the LPW estimator based on a polynomial of high degree is arbitrarily close to $n^{-1/2}$.

Keywords: Asymptotic bias, asymptotic normality, bias reduction, long memory, minimax rate, optimal bandwidth, Whittle likelihood.

AMS 1991 *subject classifications.* Primary 62M15, 62G05; secondary 60G18.
JEL Classification Numbers: C13, C14, C22.

1 Introduction

We consider a stationary long memory process $\{x_t\}$ with spectral density $f(\lambda)$ satisfying:

Assumption 1. $f(\lambda) = |\lambda|^{-2d_0}\varphi(\lambda)$, where $\varphi(\lambda)$ is continuous at 0, $0 < \varphi(0) < \infty$, and $d_0 \in (d_1, d_2)$ with $-1/2 < d_1 < d_2 < 1/2$.

The parameter d_0 determines the long-memory properties of $\{x_t\}$ and $\varphi(\lambda)$ determines its short-run dynamics.

Our objective is to estimate the long memory parameter d_0. In order to maintain generality of the short-run dynamics of $\{x_t\}$, we do not impose a specific functional form on $\varphi(\lambda)$. Instead, we allow $\varphi(\lambda)$ to belong to a family that is characterized by regularity conditions near frequency zero. This is a narrow-band approach to estimating the long memory parameter.

Examples in the literature of the narrow-band approach include the widely used GPH estimator introduced by Geweke and Port-Hudak (1983) and the local Whittle estimator (also known as the Gaussian semiparametric estimator) suggested by Künsch (1987) and analyzed by Robinson (1995a). These methods approximate the logarithm of $\varphi(\lambda)$ by a constant in a shrinking neighborhood of the origin. In consequence, the typical rate of convergence is just $n^{-2/5}$, no matter how regular $\varphi(\lambda)$ is. In addition, these estimators can be quite biased due to contamination from high frequencies (e.g., see Agiakloglou, Newbold, and Wohar (1993)).

To alleviate these problems, we approximate the logarithm of $\varphi(\lambda)$ near zero by a constant plus an even polynomial of degree $2r$, viz., $\log G - \sum_{k=1}^{r} \theta_k \lambda^{2k}$. The choice of an *even* polynomial reflects the symmetry of the spectrum about zero. This approximation is used to specify a *local polynomial Whittle* (LPW) likelihood function. We consider estimators of d_0 that are determined by the LPW likelihood.

Let $(d^*(r), G^*(r), \theta^*(r))$ denote an estimator that minimizes the (negative) LPW likelihood with respect to (d, G, θ) over the parameter space $[d_1, d_2] \times R \times \Theta$, where $\theta = (\theta_1, ..., \theta_r)'$ and Θ is a compact and convex subset of R^r. We show that an LPW–MIN estimator, $d^*(r)$, is consistent for d_0 by extending the argument of Robinson (1995a). To establish asymptotic normality of $d^*(r)$, a typical argument would first establish consistency of $(G^*(r), \theta^*(r))$. But, showing that $(G^*(r), \theta^*(r))$ is consistent is problematic, because the LPW likelihood becomes flat as a function of θ as $n \to \infty$ and the rate at which it flattens differs for each element of θ.

To circumvent this problem, we analyze an LPW estimator that is defined to be a solution to the first-order conditions (FOCs) rather than to the minimization problem. In particular, we consider the FOCs for the concentrated LPW likelihood in which the scalar G has been concentrated out. The FOC approach is effective because one can use different normalizations of the FOCs for the different parameters $d, \theta_1, ..., \theta_r$. By doing so, one can ensure that the FOCs for all parameters and the corresponding Hessian matrix are asymptotically non-degenerate.

The LPW–FOC estimator $(\widehat{d}(r), \widehat{\theta}(r))$ of (d, θ) is defined as follows. $(\widehat{d}(r), \widehat{\theta}(r))$ is a solution in $[d_1, d_2] \times \Theta$ to the FOCs of the concentrated LPW likelihood. If there are multiple solutions, $(\widehat{d}(r), \widehat{\theta}(r))$ is the solution for which $\widehat{d}(r)$ is closest to some LPW-MIN estimator $d^*(r)$. If there is no solution, $(\widehat{d}(r), \widehat{\theta}(r))$ is $(d^*(r), \theta^*(r))$.

We establish consistency and asymptotic normality of the LPW–FOC estimator simultaneously using the following steps. First, we show: (i) there exists a solution $(\widetilde{d}(r), \widetilde{\theta}(r))$ to the FOCs with probability that goes to one as $n \to \infty$ and this solution is consistent and asymptotically normal with the estimator of d_0 equaling $d_0 + O_p(m^{-1/2})$, where m is the number of frequencies near zero that are employed in the LPW likelihood. This implies that $(\widehat{d}(r), \widehat{\theta}(r))$ satisfies the FOCs with probability that goes to one as $n \to \infty$. But, it does not imply that *all* solutions to the FOCs are consistent and asymptotically normal.

We show next: (ii) all solutions $(\overline{d}(r), \overline{\theta}(r))$ to the FOCs for which $\overline{d}(r)$ is $\log^5 m$–consistent for d_0 (meaning $\overline{d}(r) = d_0 + o_p(\log^{-5} m)$) are consistent and asymptotically normal with the same asymptotic distribution. This is an unusual result in that it does not assume that $\overline{\theta}(r)$ is close to θ_0. The result holds because the (normalized) Hessian matrix of the LPW likelihood does not depend on θ, at least up to a term that is $o_p(1)$ uniformly over $\theta \in \Theta$.

We then show: (iii) $d^*(r)$ is $\log^5 m$–consistent for d_0. This result, combined with the proof of the existence of a solution that is $\log^5 m$–consistent in (i), implies that $\widehat{d}(r)$ is $\log^5 m$–consistent for d_0. In consequence, result (ii) implies that $(\widehat{d}(r), \widehat{\theta}(r))$ is consistent and asymptotically normal. These results hold when $\varphi(\lambda)$ is smooth of order s at zero (defined precisely below), where $s > 2r$ and $s \geq 1$.

For example, suppose m is chosen to diverge to infinity at what is found to be the asymptotically MSE-optimal rate, viz., $\lim_{n\to\infty} m^{\phi+1/2}/n^{\phi} = A \in (0, \infty)$, where $\phi = \min\{s, 2 + 2r\}$. Also, suppose that $s \geq 2 + 2r$. Then, the asymptotic normal result is

$$m^{1/2}(\widehat{d}(r) - d_0) \to_d N(Ab_{2+2r}\tau_r,\ c_r/4) \text{ as } n \to \infty, \tag{1.1}$$

where τ_r and c_r are known constants (specified below) for which c_r increases in r and $c_0 = 1$ and b_{2+2r} is the $(2 + 2r)$–th derivative of $\log \varphi(\lambda)$ at $\lambda = 0$. This yields the consistency, asymptotic normality, "asymptotic bias," and "asymptotic mean-squared error" of $\widehat{d}(r)$. In this case, $n^{\phi/(2\phi+1)}(\widehat{d}(r) - d_0) = O_p(1)$. If m is chosen to diverge at a slower rate, then the mean in the asymptotic normal distribution is zero.

Our results show that the effect of including the polynomial $\sum_{k=1}^{r} \theta_k \lambda^{2k}$ in the local Whittle likelihood is to increase the asymptotic variance of $\widehat{d}(r)$ by the multiplicative constant c_r, but to reduce its asymptotic bias by an order of magnitude provided $\varphi(\lambda)$ is smooth of order $s > 2$. The asymptotic bias goes from $O(m^2/n^2)$ when $r = 0$ to $O(m^{\phi}/n^{\phi})$ with $\phi > 2$ when $r > 0$ and $s > 2$. In consequence, the rate of convergence of $\widehat{d}(r)$ is faster when $r > 0$ than when $r = 0$ provided $s > 2$ (and m is chosen appropriately). For example, for $r > 0$, $s \geq 2 + 2r$, and m chosen as in (1.1), the rate of convergence of $\widehat{d}(r)$ is $n^{-(2+2r)/(5+4r)}$, whereas the rate of convergence for $\widehat{d}(0)$ is $n^{-2/5}$. Note that, for r arbitrarily large, the rate of convergence is arbitrarily close to $n^{-1/2}$, the rate that is obtained in a parametric model.

We calculate the asymptotic MSE optimal choice of bandwidth m and provide a plug-in version. The plug-in version is based on the $(r + 1)$-th element of $\widehat{\theta}(r + 1)$. The latter times $-(2 + 2r)!$ is a consistent estimator of the $(2 + 2r)$-th derivative of

$\log \varphi(\lambda)$ at $\lambda = 0$, which appears in the formula for the asymptotic MSE optimal choice of bandwidth.

The results of the paper provide some new results for the local Whittle estimator $d^*(0)$ considered by Robinson (1995a). The results show that this estimator has an asymptotic bias (defined as $m^{-1/2}$ times the mean of its asymptotic normal distribution) that is the same as that of the GPH estimator. Robinson's (1995a, b) results show that the asymptotic variance of the local Whittle estimator is smaller than that of the GPH estimator. Combining these results establishes that the asymptotic mean-squared error of the local Whittle estimator is smaller than that of the GPH estimator (provided m is chosen appropriately).

A limited number of Monte Carlo simulation results show that the asymptotic results of the paper mimic the finite sample properties of LPW estimators quite well. We simulate finite sample biases, standard deviations, root mean squared errors (RMSEs), confidence interval coverage probabilities, and confidence interval average lengths for LPW estimators with $r = 0, 1, 2$. The true process considered is a first-order autoregressive fractionally integrated (ARFIMA(1, 1, 0)) process with sample size $n = 512$. The introduction of polynomial terms is found to significantly reduce the finite sample bias of the local Whittle estimator when the AR parameter is non-zero (i.e., the function $\varphi(\lambda)$ is not flat). At the same time, polynomial terms inflate the local Whittle estimator's standard deviation. Andrews and Sun (2001) provide a more extensive set of simulation results for LPW estimators than is given in this paper.

The results of this paper are similar to those of Andrews and Guggenberger (1999), who consider adding the regressors $\lambda_j^2, ..., \lambda_j^{2r}$ to a log-periodogram regression that is used to estimate d_0. The resulting estimator has the same asymptotic bias as the LPW–FOC estimator $\widehat{d}(r)$, but larger variance. For any r, its variance is larger by the factor $(\pi^2/24) \div (1/4) = 1.645$. The properties of the bias-reduced log-periodogram estimator of Andrews and Guggenberger (1999) are determined under the assumption of Gaussianity of $\{x_t\}$, whereas the properties of the LPW–FOC estimator considered here are determined without requiring $\{x_t\}$ to be Gaussian.

An alternative to the narrow-band approach considered here is a broad-band approach. In this approach, one imposes regularity conditions on $\varphi(\lambda)$ for λ in the whole interval $[0, \pi]$ and one utilizes a nonparametric estimator of $\varphi(\lambda)$ for $\lambda \in [0, \pi]$. For example, Moulines and Soulier (1999, 2000), Hurvich and Brodsky (2000), and Hurvich (2000) approximate $\log \varphi(\lambda)$ by a truncated Fourier series, while Bhansali and Kokoszka (1997) approximate $\varphi(\lambda)$ by the spectrum of an autoregressive model. These papers establish that the broad-band estimators exhibit a faster rate of convergence than the GPH and local Whittle estimators under the given regularity conditions. These estimators exhibit an asymptotic mean-squared error of order $\log(n)/n$ if the number of parameters in the model goes to infinity at a suitable rate.

Some limited simulation results are reported that compare LPW estimators (for $r = 0, 1, 2$) with the bias-reduced GPH estimators of Andrews and Guggenberger (1999) (for $r = 0, 1, 2$) and with the broad-band fractional exponential (FEXP) estimator considered by Moulines and Soulier (1999, 2000), Hurvich and Brodsky (2000), and Hurvich (2000). The estimators are compared based on the minimum RMSE over different values of m for the LPW and GPH estimators and on the minimum RMSE over different numbers of terms in the expansion for the fractional exponential estimator. The true process is taken to be an ARFIMA(1, 1, 0) process with AR parameter $\phi = 0, .3, .6, .9, -.3, -.6$, or $-.9$ and sample size $n = 512$. The results are not sensitive to the value of d_0 (within the stationary region). For all values of ϕ, the estimator with the smallest minimum RMSE is an LPW or local Whittle estimator. When ϕ is non-zero, the best estimator is an LPW estimator with $r = 1$ or 2. When ϕ is zero, the best estimator is the local Whittle estimator (i.e., $r = 0$). Although the number of simulation results reported is small, the results indicate that LPW estimators are competitive with existing semiparametric estimators. Andrews and Sun (2001) provide additional simulation results that corroborate this finding.

We note that the LPW estimator can be viewed as a semiparametric local (to frequency zero) version of an approximate maximum likelihood estimator of a particular parametric FEXP model considered by Diggle (1990) and Beran (1993) that utilizes polynomials, rather than trigonometric polynomials.

Other papers in the literature that are related to this paper include Henry and Robinson (1996) and Hurvich and Deo (1999). These papers approximate $\varphi(\lambda)$ by a more flexible function than a constant in order to obtain a data-driven choice of m. In contrast, the present paper uses a more flexible approximation of $\varphi(\lambda)$ than a constant for the purposes of bias reduction and increased rate of convergence in the estimation of d_0.

The idea of using a local polynomial approximation can be applied to other estimators of d_0, such as the average-periodogram estimator of Robinson (1994). In addition, the results of this paper could be extended, presumably, to nonstationary fractional time series along the line of Shimotsu and Phillips (1999).

The remainder of the paper is organized as follows. Section 2 defines the LPW likelihood function. Section 3 states the assumptions used. Section 4 shows that there exists a sequence of solutions to the FOCs that is consistent and asymptotically normal. Section 5 shows that the LPW–FOC estimator is consistent and asymptotically normal with bias that may be non-negligible. Section 6 establishes that the LPW–FOC estimator attains the optimal rate of convergence for estimation of d_0. Section 7 provides the Monte Carlo simulation results. Section 8 contains proofs.

Throughout the paper, wp$\rightarrow 1$ abbreviates "with probability that goes to one as $n \rightarrow \infty$" and $||\cdot||$ signifies the Euclidean norm.

2 Definition of LPW Likelihood

The j-th fundamental frequency λ_j, the discrete Fourier transform w_j of $\{x_t\}$, and the periodogram I_j of $\{x_t\}$ are defined by

$$\lambda_j = 2\pi j/n, \; w_j = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} x_t \exp(it\lambda_j), \text{ and } I_j = |w_j|^2. \tag{2.1}$$

The local polynomial Whittle log-likelihood is $-m/2$ times

$$Q_r(d,G,\theta) = m^{-1}\sum_{j=1}^{m}\left\{\log\left[G\lambda_j^{-2d}\exp(-p_r(\lambda_j,\theta))\right] + \frac{I_j}{G\lambda_j^{-2d}\exp(-p_r(\lambda_j,\theta))}\right\},$$

where

$$p_r(\lambda_j,\theta) = \sum_{k=1}^{r}\theta_k\lambda_j^{2k} \text{ and } \theta = (\theta_1,...,\theta_r)'. \tag{2.2}$$

The log-likelihood is local to frequency zero, because m is taken such that $1/m + m/n \to 0$ as $n \to \infty$. The log-likelihood is based on approximating $\log\varphi(\lambda)$ by $\log G - p_r(\lambda,\theta)$ for λ near zero. The local Whittle likelihood considered in Robinson (1995a) is obtained by setting $\theta = 0$.

Concentrating $Q_r(d,G,\theta)$ with respect to $G \in (-\infty,\infty)$ yields the concentrated LPW log-likelihood $R_r(d,\theta)$:

$$\begin{aligned} R_r(d,\theta) &= \log\widehat{G}(d,\theta) - m^{-1}\sum_{j=1}^{m}p_r(\lambda_j,\theta) - 2dm^{-1}\sum_{j=1}^{m}\log\lambda_j + 1, \text{ where} \\ \widehat{G}(d,\theta) &= m^{-1}\sum_{j=1}^{m}I_j\exp(p_r(\lambda_j,\theta))\lambda_j^{2d}. \end{aligned} \tag{2.3}$$

An LPW–MIN estimator $(d^*(r),G^*(r),\theta^*(r))$ of (d,G,θ) solves the following minimization problem:

$$\begin{aligned} (d^*(r),\theta^*(r)) &= \underset{d\in[d_1,d_2],\theta\in\Theta}{\arg\min} R_r(d,\theta) \text{ and} \\ G^*(r) &= \widehat{G}(d^*(r),\theta^*(r)), \end{aligned} \tag{2.4}$$

where Θ is a compact and convex set in R^r.

The LPW–FOC estimator $(\widehat{d}(r),\widehat{\theta}(r))$ is a solution in $[d_1,d_2]\times\Theta$ to the FOCs:

$$\frac{\partial}{\partial(d,\theta')'}R_r(d,\theta) = 0. \tag{2.5}$$

If there are multiple solutions, $(\widehat{d}(r),\widehat{\theta}(r))$ is the solution for which $\widehat{d}(r)$ is closest to some LPW-MIN estimator $d^*(r)$. If there is no solution, $(\widehat{d}(r),\widehat{\theta}(r))$ is $(d^*(r),\theta^*(r))$. By definition, $\widehat{G}(r) = \widehat{G}(\widehat{d}(r),\widehat{\theta}(r))$.

3 Assumptions

We now introduce the assumptions that are employed (in conjunction with Assumption 1) to establish the consistency and asymptotic normality of $(\widehat{d}(r),\widehat{\theta}(r))$. These assumptions utilize the following definition. Let $[s]$ denote the integer part of s. We say that a real function h defined on a neighborhood of zero is smooth of order $s > 0$ at zero if h is $[s]$ times continuously differentiable in some neighborhood of zero and its derivative of order $[s]$, denoted $h^{([s])}$, satisfies a Hölder condition of order $s-[s]$ at zero, i.e., $|h^{([s])}(\lambda) - h^{([s])}(0)| \leq C|\lambda|^{s-[s]}$ for some constant $C < \infty$ and all λ in a neighborhood of zero.

Assumption 2. $\varphi(\lambda)$ is smooth of order s at $\lambda = 0$, where $s > 2r$ and $s \geq 1$.

Assumption 2 imposes the regularity on the function $\varphi(\lambda)$ that characterizes the semiparametric nature of the model. Under Assumption 2, $\log\varphi(\lambda)$ has a Taylor expansion of the form:

$$\begin{aligned} \log\varphi(\lambda) &= \log\varphi(0) + \sum_{k=1}^{[s/2]} \frac{b_{2k}}{(2k)!}\lambda^{2k} + O(\lambda^s) \text{ as } \lambda \to 0+, \text{ where} \\ b_k &= \left.\frac{d^k}{d\lambda^k}\log\varphi(\lambda)\right|_{\lambda=0}. \end{aligned} \tag{3.1}$$

The true values for G and θ are $G_0 = \varphi(0)$ and $\theta_0 = (\theta_{0,1}, ..., \theta_{0,r})'$, where

$$\theta_{0,k} = -\frac{b_{2k}}{(2k)!} \text{ for } k = 1, ..., r. \tag{3.2}$$

Assumption 3. (a) The time series $\{x_t : t = 1, ..., n\}$ satisfies

$$x_t - Ex_0 = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j},$$

where

$$\begin{aligned} &\sum_{j=0}^{\infty} \alpha_j^2 < \infty, \; E(\varepsilon_t|F_{t-1}) = 0 \text{ a.s.}, \; E(\varepsilon_t^2|F_{t-1}) = 1 \text{ a.s.}, \\ E(\varepsilon_t^3|F_{t-1}) &= \sigma_3 \text{ a.s.}, \; E(\varepsilon_t^4|F_{t-1}) = \sigma_4 \text{ a.s. for } t = ..., -1, 0, 1, ..., \end{aligned}$$

and F_{t-1} is the σ-field generated by $\{\varepsilon_s : s < t\}$.
(b) There exists a random variable ε with $E\varepsilon^2 < \infty$ such that for all $\nu > 0$ and some $K > 0$, $P(|\varepsilon_t| > \nu) < KP(|\varepsilon| > \nu)$.
(c) In some neighborhood of the origin, $(d/d\lambda)\alpha(\lambda) = O(|\alpha(\lambda)|/\lambda)$ as $\lambda \to 0+$, where $\alpha(\lambda) = \sum_{j=1}^{\infty} \alpha_j e^{-ij\lambda}$.

Assumption 3 states that the time series $\{x_t\}$ is a linear process with martingale difference innovations. Unlike most results for log-periodogram regression estimators, Assumption 3 allows for non-Gaussian processes. Assumption 3(a) and (b) is the same as Assumption A3$'$ of Robinson (1995a). Assumption 3(c) is the same as Assumption A2$'$ of Robinson (1995a). It should be possible to weaken the assumption that $E(\varepsilon_t^4|F_{t-1}) = \sigma_4$ a.s. along the lines of Robinson and Henry (1999).

Assumption 4. $m^{2r+1/2}/n^{2r} \to \infty$ and $m^{\phi+1/2}/n^{\phi} = O(1)$ as $n \to \infty$, where $\phi = \min\{s, 2+2r\}$.

The two conditions in Assumption 4 are always compatible because $s > 2r$ by Assumption 2. The first condition of Assumption 4 is used to ensure that the matrix B_n that is used to normalize the gradient and Hessian of $mR_r(d, \theta)$ satisfies $\lambda_{\min}(B_n) \to \infty$, which is required for consistency of $(\widehat{d}(r), \widehat{\theta}(r))$. The second condition of Assumption 4 is used to guarantee that the normalized gradient of $mR_r(d_0, \theta_0)$ is $O_p(1)$, which is required for asymptotic normality of $(\widehat{d}(r), \widehat{\theta}(r))$.

If $r = 0$ and Assumption 2 holds with $s = 2$, then Assumption A1$'$ of Robinson (1995a) holds with $\beta = 2$. In this case, his Assumption A4$'$ on m is weakest and it requires that $1/m + m^5(\log^2 m)/n^4 \to 0$. In contrast, in this case, Assumption 4 requires $1/m \to 0$ and $m^5/n^4 = O(1)$, which is slightly weaker than Robinson's Assumption A4$'$. (It seems that the $\log^2 m$ term in Robinson's Assumption A4$'$ is superfluous. It is used on p. 1644 of Robinson's proof of Theorem 2 to bound (4.11), but does not appear to be necessary because $\nu_j - \nu_{j+1} = O(j^{-1})$ and $\nu_m = O(1)$, where $\nu_j := \log j - m^{-1} \sum_{k=1}^{m} \log k$.)

Assumption 5. Θ is compact and convex and θ_0 lies in the interior of Θ.

4 Existence of Solutions to the First-order Conditions

We start this section by stating a general Lemma that provides sufficient conditions for the existence of a consistent sequence of solutions to the FOCs of a sequence of stochastic optimization problems. The Lemma also provides an asymptotic representation of the (normalized) solutions. Next, we apply the Lemma to the LPW likelihood. The Lemma has numerous antecedents in the literature, e.g., see Weiss (1971, 1973), Crowder (1976), Heijmans and Magnus (1986), and Wooldridge (1994). The Lemma given here is closest to that of Wooldridge (1994, Theorem 8.1).

Let $\{L_n(\gamma) : n \geq 1\}$ be a sequence of minimands for estimation of the parameter $\gamma_0 \in \Gamma \subset R^k$, where Γ is the parameter space. Denote the gradient and Hessian of $L_n(\gamma)$ by $\nabla L_n(\gamma)$ and $\nabla^2 L_n(\gamma)$ respectively.

Lemma 1 *Suppose γ_0 is in the interior of Γ, $L_n(\gamma)$ is twice continuously differentiable on a neighborhood of γ_0, and there exists a sequence of $k \times k$ non-random nonsingular matrices B_n such that*

(i) $||B_n^{-1}|| \to 0$ *as* $n \to \infty$,

(ii) $(B_n^{-1})' \nabla L_n(\gamma_0) = O_p(1)$ *as* $n \to \infty$,

(iii) *for some* $\eta > 0$, $\lambda_{\min}\left((B_n^{-1})' \nabla^2 L_n(\gamma_0) B_n^{-1}\right) \geq \eta$ *wp* $\to 1$, *and*

(iv) $\sup_{\gamma \in \Gamma : ||B_n(\gamma - \gamma_0)|| \leq K_n} ||(B_n^{-1})'(\nabla^2 L_n(\gamma) - \nabla^2 L_n(\gamma_0)) B_n^{-1}|| = o_p(1)$ *as* $n \to \infty$

for some sequence of scalar constants $\{K_n : n \geq 1\}$ for which $K_n \to \infty$ as $n \to \infty$. Then, there exists a sequence of estimators $\{\widetilde{\gamma}_n : n \geq 1\}$ that satisfy the first-order conditions $\nabla L_n(\widetilde{\gamma}_n) = 0$ wp$\to 1$ and

$$B_n(\widetilde{\gamma}_n - \gamma_0) = -Y_n + o_p(1) = O_p(1), \text{ where}$$
$$Y_n = ((B_n^{-1})' \nabla^2 L_n(\gamma_0) B_n^{-1})^{-1} (B_n^{-1})' \nabla L_n(\gamma_0).$$

The proof of Lemma 1 is given in Section 8.

We apply Lemma 1 with $\gamma = (d, \theta')'$, $L_n(\gamma) = mR_r(d, \theta)$, and B_n equal to the $(r+1) \times (r+1)$ diagonal matrix with j-th diagonal element $[B_n]_{jj}$ defined by

$$[B_n]_{11} = m^{1/2} \text{ and } [B_n]_{jj} = \left(\frac{2\pi m}{n}\right)^{2j-2} m^{1/2} \text{ for } j = 2, ..., r+1. \tag{4.1}$$

The first condition of Assumption 4 guarantees that $||B_n^{-1}|| \to 0$, as required by condition (i) of Lemma 1.

To verify conditions (ii)–(iv) of Lemma 1, we need to establish some properties of the normalized score (i.e., gradient) and Hessian of $mR_r(d, \theta)$. The score vector and Hessian matrix of $mR_r(d, \theta)$ are denoted $S_n(d, \theta) = m\nabla R_r(d, \theta)$ and $H_n(d, \theta) = m\nabla^2 R_r(d, \theta)$ respectively. Some algebra gives

$$\begin{aligned}
S_n(d, \theta) &= \widehat{G}^{-1}(d, \theta) \sum_{j=1}^{m} \left(y_j(d, \theta) - m^{-1} \sum_{k=1}^{m} y_k(d, \theta) \right) X_j \text{ and} \\
H_n(d, \theta) &= \widehat{G}^{-2}(d, \theta) \left(\widehat{G}(d, \theta) \sum_{j=1}^{m} y_j(d, \theta) X_j X_j' \right. \\
&\quad \left. -m \left(m^{-1} \sum_{j=1}^{m} y_j(d, \theta) X_j \right) \left(m^{-1} \sum_{j=1}^{m} y_j(d, \theta) X_j \right)' \right)
\end{aligned} \tag{4.2}$$

where

$$\begin{aligned}
y_j(d, \theta) &= I_j \exp(p_r(\lambda_j, \theta)) \lambda_j^{2d} \text{ and} \\
X_j &= (2 \log j, \lambda_j^2, ..., \lambda_j^{2r})'.
\end{aligned} \tag{4.3}$$

We show below that the normalized Hessian, $B_n^{-1} H_n(d_0, \theta_0) B_n^{-1}$, converges in probability to the $(r+1) \times (r+1)$ matrix Ω_r defined by

$$\Omega_r = \begin{pmatrix} 4 & 2\mu_r' \\ 2\mu_r & \Gamma_r \end{pmatrix}, \tag{4.4}$$

where μ_r is a column r-vector with k-th element $\mu_{r,k}$, Γ_r is an $r \times r$ matrix with (i, k)-th element $[\Gamma_r]_{i,k}$,

$$\begin{aligned}
\mu_{r,k} &= \frac{2k}{(2k+1)^2} \text{ for } k = 1, ..., r, \text{ and} \\
[\Gamma_r]_{i,k} &= \frac{4ik}{(2i+2k+1)(2i+1)(2k+1)} \text{ for } i, k = 1, ..., r.
\end{aligned} \tag{4.5}$$

For $r = 0$, define $\Omega_r = 4$.

We show below that the asymptotic bias of the normalized score, $B_n^{-1} S_n(d_0, \theta_0)$, is $-\nu_n(r, s)$, where

$$\begin{aligned}
\nu_n(r, s) &= m^{\phi+1/2} n^{-\phi} \left(1(s \geq 2+2r) b_{2+2r} \kappa_r \xi_r^+ + 1(2r < s < 2+2r) O(1) \right) \\
&= 1\,(s \geq 2+2r)\, m^{5/2+2r} n^{-(2+2r)} b_{2+2r} \kappa_r \xi_r^+ \\
&\quad +1 \left(2r < s < 2+2r) O(m^{s+1/2} n^{-s} \right),
\end{aligned} \tag{4.6}$$

and

$$\xi_r^+ = \begin{pmatrix} 2 \\ \xi_r \end{pmatrix},$$
$$\xi_r = (\xi_{r,1}, ..., \xi_{r,r})',$$
$$\xi_{r,k} = \frac{2k(3+2r)}{(2r+2k+3)(2k+1)} \text{ for } k = 1, ..., r, \text{ and}$$
$$\kappa_r = -\frac{(2\pi)^{2+2r}(2+2r)}{(3+2r)!(3+2r)}. \tag{4.7}$$

The following Lemma establishes the asymptotic properties of the normalized score and Hessian, which are needed to verify conditions (ii)–(iv) of Lemma 1. The following quantities arise in the Lemma:

$$D_m(\eta) = \{d \in [d_1, d_2] : (\log^5 m)|d - d_0| < \eta\} \text{ for } \eta > 0 \text{ and}$$
$$J_n = \sum_{j=1}^{m} \left(X_j - m^{-1} \sum_{k=1}^{m} X_k \right) \left(X_j - m^{-1} \sum_{k=1}^{m} X_k \right)'. \tag{4.8}$$

Lemma 2 *Under Assumptions* 1–5*, as* $n \to \infty$*, we have*

(a) $\sup_{d \in D_m(\eta_n), \theta \in \Theta} ||B_n^{-1}(H_n(d, \theta) - H_n(d_0, \theta))B_n^{-1}|| = o_p(1)$ *for all sequences of constants* $\{\eta_n : n \geq 1\}$ *for which* $\eta_n = o(1)$*,*

(b) $\sup_{\theta \in \Theta} ||B_n^{-1}(H_n(d_0, \theta) - H_n(d_0, \theta_0))B_n^{-1}|| = o_p(1)$*,*

(c) $||B_n^{-1}(H_n(d_0, \theta_0) - J_n)B_n^{-1}|| = o_p(1)$*,*

(d) $B_n^{-1} J_n B_n^{-1} \to \Omega_r$*, and*

(e) $B_n^{-1} S_n(d_0, \theta_0) + \nu_n(r, s) \to_d N(0, \Omega_r)$*.*

Remarks 1. The result of part (b) of the Lemma is unusual. It states that the normalized Hessian matrix $H_n(d_0, \theta)$ does not depend on θ up to $o_p(1)$ uniformly over $\theta \in \Theta$. In most nonlinear estimation problems, this would not hold. This property of the Hessian is essential for the proof of consistency and asymptotic normality of $(\widehat{d}(r), \widehat{\theta}(r))$ that is used here to work.

2. The proof of Lemma 2 relies heavily on the proof of Theorem 2 of Robinson (1995a). It also uses Theorem 2 of Robinson (1995b) and Theorem 5.2.4 of Brillinger (1975).

We now use the results of Lemma 2 to verify the conditions (ii)–(iv) of Lemma 1. Condition (ii) holds by Lemma 2(e) and the second condition of Assumption 4. Condition (iii) holds by Lemma 2(c) and (d) and the positive definiteness of Ω_r. Condition (iv) holds with $K_n = m^{1/2} \eta_n \log^{-5} m$ for some sequence η_n that goes to zero sufficiently slowly that $K_n \to \infty$, e.g., $\eta_n = \log^{-1} m$, by Lemma 2(a) and (b).

In consequence, the application of Lemma 1 with $L_n(\gamma) = mR_r(d, \theta)$ combined with the convergence results of Lemma 2 gives the following Theorem:

Theorem 1 *Under Assumptions* 1–5, *there exist solutions* $(\widetilde{d}(r), \widetilde{\theta}(r))$ *to the first-order conditions* $(\partial/\partial(d,\theta')')R_r(d,\theta) = 0$ *wp*$\rightarrow 1$ *and*

$$B_n \begin{pmatrix} \widetilde{d}(r) - d_0 \\ \widetilde{\theta}(r) - \theta_0 \end{pmatrix} - \Omega_r^{-1}\nu_n(r,s) \rightarrow_d N(0, \Omega_r^{-1}).$$

An immediate consequence of Theorem 1 and the definition of $(\widehat{d}(r), \widehat{\theta}(r))$ is the following:

Corollary 1 *Under Assumptions* 1–5, *the* LPW–FOC *estimator* $(\widehat{d}(r), \widehat{\theta}(r))$ *solves the FOCs* (2.5) *wp*$\rightarrow 1$.

5 Asymptotic Normality of the LPW–FOC Estimator

In this section, we show: (i) any sequence of solutions to the FOCs for which the estimator of d_0 is $\log^5 m$-consistent (i.e., equals $d_0 + o_p(\log^{-5} m)$) is consistent and asymptotically normal. We also show: (ii) the LPW–MIN, $d^*(r)$, estimator of d_0 is $\log^5 m$-consistent. The latter is used to show: (iii) the LPW–FOC, $\widehat{d}(r)$, estimator of d_0 is $\log^5 m$-consistent. Results (i) and (iii) and Corollary 1 then imply that the LPW–FOC estimator $(\widehat{d}(r), \widehat{\theta}(r))$ is consistent and asymptotically normal.

Suppose a sequence of estimators $(\overline{d}(r), \overline{\theta}(r))$ satisfies the FOCs (2.5) wp$\rightarrow 1$ and $\overline{d}(r) - d_0 = o_p(\log^{-5} m)$. Then, there exists a sequence of constants $\eta_n > 0$ that converges to zero sufficiently slowly that $\overline{d}(r) \in D_m(\eta_n)$ wp$\rightarrow 1$. Element by element mean value expansions of $B_n^{-1} m \nabla R_r(\overline{d}(r), \overline{\theta}(r))$ $(= B_n^{-1} S_n(\overline{d}(r), \overline{\theta}(r)))$ about (d_0, θ_0) gives

$$\begin{aligned} 0 &= B_n^{-1} S_n(\overline{d}(r), \overline{\theta}(r)) \\ &= B_n^{-1} S_n(d_0, \theta_0) + \left(B_n^{-1} H_n(d_1, \theta_1) B_n^{-1}\right) B_n \begin{pmatrix} \overline{d}(r) - d_0 \\ \overline{\theta}(r) - \theta_0 \end{pmatrix} \\ &= B_n^{-1} S_n(d_0, \theta_0) + \Omega_r(1 + o_p(1)) B_n \begin{pmatrix} \overline{d}(r) - d_0 \\ \overline{\theta}(r) - \theta_0 \end{pmatrix}, \end{aligned} \tag{5.1}$$

where (d_1, θ_1) (which may differ across the rows of $H_n(d_1, \theta_1)$) lies between $(\overline{d}(r), \overline{\theta}(r))$ and (d_0, θ_0) and, hence, $d_1 \in D_m(\eta_n)$ wp$\rightarrow 1$ and $\theta_1 \in \Theta$ using the convexity of Θ (by Assumption 5). The third equality holds by Lemma 2(a)–(d). Rearrangement of (5.1) yields

$$\begin{aligned} B_n \begin{pmatrix} \overline{d}(r) - d_0 \\ \overline{\theta}(r) - \theta_0 \end{pmatrix} - \Omega_r^{-1}\nu_n(r,s) &= -\Omega_r^{-1}(1 + o_p(1)) B_n^{-1} S_n(d_0, \theta_0) - \Omega_r^{-1}\nu_n(r,s) \\ &\rightarrow_d N(0, \Omega_r^{-1}) \end{aligned} \tag{5.2}$$

as $n \rightarrow \infty$. The convergence in distribution holds by Lemma 2(e).

Hence, we have established the following result:

Theorem 2 *Suppose Assumptions* 1–5 *hold. Let* $(\overline{d}(r), \overline{\theta}(r))$ *be any sequence of estimators that satisfies the FOCs* (2.5) *wp*$\rightarrow 1$ *and for which* $\overline{d}(r) - d_0 = o_p(\log^{-5} m)$. *Then,*

$$B_n \begin{pmatrix} \overline{d}(r) - d_0 \\ \overline{\theta}(r) - \theta_0 \end{pmatrix} - \Omega_r^{-1} \nu_n(r,s) \rightarrow_d N(0, \Omega_r^{-1}) \text{ as } n \rightarrow \infty.$$

Next we show that the LPW–MIN estimator, $d^*(r)$, of d_0 is $\log^5 m$-consistent. This, in turn, implies that the LPW–FOC estimator, $\widehat{d}(r)$, also is $\log^5 m$-consistent. The reason is that there exists a solution to the first-order conditions, $\widetilde{d}(r)$, that is $\log^5 m$-consistent by Theorem 1. Hence, $d^*(r)$ and $\widetilde{d}(r)$ differ by $o_p(\log^{-5} m)$. By Corollary 1 and the definition of $(\widehat{d}(r), \widehat{\theta}(r))$, wp$\rightarrow 1$, $(\widehat{d}(r), \widehat{\theta}(r))$ is the solution to the FOCs whose estimator of d_0 is closest to $d^*(r)$. Thus, $\widehat{d}(r)$ cannot differ from $d^*(r)$ by more than $\widetilde{d}(r)$ does. That is, $\widehat{d}(r) - d^*(r) = o_p(\log^{-5} m)$ and $\widehat{d}(r) - d_0 = o_p(\log^{-5} m)$.

Lemma 3 *Suppose Assumptions* 1–5 *hold. Then,*
(a) $d^*(r) - d_0 = o_p(\log^{-5} m)$ *and*
(b) $\widehat{d}(r) - d_0 = o_p(\log^{-5} m)$.

Remark. The proof of Lemma 3 shows that an estimator that minimizes $R_r(d, \theta)$ over $[d_1, d_2]$ for any value of θ yields a $\log^5 m$–consistent estimator of d_0. This is not too surprising, because the choice $\theta = 0$, upon which the $\log^5 m$–consistent local Whittle estimator $d^*(0)$ is based, is not necessarily a better choice of θ than any other $\theta \in \Theta$.

Corollary 1 and Lemma 3(b) imply that $(\widehat{d}(r), \widehat{\theta}(r))$ satisfy the conditions of Theorem 2. Hence, Theorem 2 implies the following consistency and asymptotic normality result, which is the main result of the paper.

Corollary 2 *Under Assumptions* 1–5, *the* LPW–FOC *estimator* $(\widehat{d}(r), \widehat{\theta}(r))$ *satisfies*

$$\begin{pmatrix} m^{1/2}(\widehat{d}(r) - d_0) \\ m^{1/2} Diag\left((2\pi m/n)^2, ..., (2\pi m/n)^{2r}\right)\left(\widehat{\theta}(r) - \theta_0\right) \end{pmatrix} - \Omega_r^{-1} \nu_n(r,s) \rightarrow_d N(0, \Omega_r^{-1})$$

as $n \rightarrow \infty$.

Remarks 1. By the formula for a partitioned inverse,

$$\Omega_r^{-1} = \begin{pmatrix} \frac{c_r}{4} & -\frac{c_r}{2}\mu_r' \Gamma_r^{-1} \\ -\frac{c_r}{2}\Gamma_r^{-1}\mu_r & \Gamma_r^{-1} + c_r \Gamma_r^{-1}\mu_r\mu_r'\Gamma_r^{-1} \end{pmatrix}, \text{ where}$$
$$c_r = (1 - \mu_r' \Gamma_r^{-1} \mu_r)^{-1} \text{ for } r > 0 \text{ and } c_0 = 1. \tag{5.3}$$

Hence, the asymptotic variance of $m^{1/2}(\widehat{d}(r) - d_0))$ is $c_r/4$, which is free of nuisance parameters. The use of the polynomial $p_r(\lambda_j, \theta)$ in the specification of the local Whittle likelihood increases the asymptotic variance of $\widehat{d}(r)$ by the multiplicative constant c_r. For example, $c_1 = 9/4$, $c_2 = 3.52$, $c_3 = 4.79$, $c_4 = 6.06$, and $c_5 = 7.33$.

2. The "asymptotic bias" of $\widehat{d}(r)$ equals the first element of $m^{-1/2}\Omega_r^{-1}\nu_n(r,s)$. Using (5.3) and the definition of $\nu_n(r,s)$ in (4.6), the asymptotic bias of $\widehat{d}(r)$ equals

$$1(s \geq 2+2r)\tau_r b_{2+2r} m^{2+2r} n^{-(2+2r)} + 1(2r < s < 2+2r) O(m^s/n^s), \text{ where}$$
$$\tau_r = \frac{\kappa_r c_r}{2}(1 - \mu_r' \Gamma_r^{-1}\xi_r). \tag{5.4}$$

For example, $\tau_0 = -2.19$, $\tau_1 = 2.23$, $\tau_2 = -.793$, $\tau_3 = .146$, $\tau_4 = -.0164$, and $\tau_5 = .00125$.

3. By (5.4), the asymptotic bias of $\widehat{d}(r)$ is of order m^ϕ/n^ϕ, where $\phi = \min\{s, 2+2r\}$. In contrast, the asymptotic bias of $\widehat{d}(0)$ is of order m^2/n^2. The asymptotic bias of $\widehat{d}(r)$ is smaller than that of $\widehat{d}(0)$ by an order of magnitude provided $\varphi(\cdot)$ is smooth of order $s > 2$, because in this case $\phi > 2$.

4. If $s \geq 2+2r$ and $\lim_{n\to\infty} m^{5/2+2r}/n^{2+2r} = A \in (0,\infty)$, then

$$\begin{pmatrix} m^{1/2}(\widehat{d}(r) - d_0) \\ m^{1/2} Diag\left((2\pi m/n)^2, ..., (2\pi m/n)^{2r}\right)\left(\widehat{\theta}(r) - \theta_0\right) \end{pmatrix} \to_d N\left(A b_{2+2r}\kappa_r \Omega_r^{-1}\xi_r^+, \Omega_r^{-1}\right). \tag{5.5}$$

The only unknown quantity in the asymptotic distribution is b_{2+2r}. The asymptotic bias and variance of $m^{1/2}(\widehat{d}(r) - d_0)$ are $A\tau_r b_{2+2r}$ and $c_r/4$ respectively.

5. Assumption 4 allows one to take m much larger for $\widehat{d}(r)$ than for $\widehat{d}(0)$. In consequence, by appropriate choice of m, one has asymptotic normality of $\widehat{d}(r)$ with a faster rate of convergence than is possible with $\widehat{d}(0)$.

6. Inflation of the asymptotic variance by the factor c_r due to the addition of parameters, see Remark 1, also is found in Andrews and Guggenberger (1999) for a bias-reduced log-periodogram regression estimator of d_0. In consequence, the LPW–FOC estimator $\widehat{d}(r)$ maintains exactly the same advantage over the bias-reduced log-periodogram regression estimator, in terms of having a smaller asymptotic variance, as the local Whittle estimator has over the GPH log-periodogram regression estimator. For any $r \geq 0$, the ratio of their asymptotic variances is $(c_r/4) \div (\pi^2 c_r/24) \doteq .608$.

7. The expression for the asymptotic bias in (5.4) is the same as that found in Andrews and Guggenberger (1999) for the bias-reduced log-periodogram estimator of d_0. Hence, the LPW–FOC estimator has the same asymptotic bias, but smaller asymptotic variance, than the bias-reduced log-periodogram estimator of d_0.

Suppose $s \geq 2+2r$. Using Remarks 1 and 2, the "asymptotic mean-squared error" of $\widehat{d}(r)$ is

$$AMSE(\widehat{d}(r)) = \tau_r^2 b_{2+2r}^2 (\frac{m}{n})^{4+4r} + \frac{c_r}{4m}. \tag{5.6}$$

Minimization of AMSE($\widehat{d}(r)$) with respect to m gives the AMSE–optimal choice of m :

$$m_{opt} = \left[\left(\frac{c_r}{16(1+r)\tau_r^2 b_{2+2r}^2}\right)^{1/(5+4r)} n^{(4+4r)/(5+4r)}\right], \tag{5.7}$$

where $[a]$ denotes the integer part of a. When $r = 0$ and $s = 2$, this gives the same formula for m_{opt} as in Henry and Robinson (1996).

The formula for m_{opt} contains only one unknown, b_{2+2r}. A consistent estimator of b_{2+2r} can be obtained from $\widehat{\theta}(r+1)$, the LPW-FOC estimator of θ that uses a polynomial of degree $2r+2$. Let $\overline{m}$ denote the number of frequencies used in the calculation of $\widehat{\theta}(r+1)$. Let $\widehat{\theta}(r+1)_{r+1}$ denote the $(r+1)$-th element of $\widehat{\theta}(r+1)$. Suppose $s > 2+2r$. By Corollary 2 and the definition of θ_0 in (3.2),

$$\widehat{b}_{2+2r} = -(2+2r)!\widehat{\theta}(r+1)_{r+1} \rightarrow_p b_{2+2r} \tag{5.8}$$

provided $\Omega_{r+1}^{-1}\nu_n(r+1,s) \rightarrow 0$ and Assumption 4 holds with (m,r,ϕ) replaced by $(\overline{m}, r+1, \overline{\phi})$, where $\overline{\phi} = \min\{s, 4+2r\}$. These conditions hold if $\overline{m}^{1/2}(\overline{m}/n)^{2+2r} \rightarrow \infty$ and $\overline{m}^{1/2}(\overline{m}/n)^{\overline{\phi}} \rightarrow 0$. For example, they hold if $\overline{m} = Cn^{\gamma}$ for some $\gamma \in ((2+2r)/(5/2+2r), \overline{\phi}/(1/2+\overline{\phi}))$ and $0 < C < \infty$. (The interval for γ is nondegenerate because $\overline{\phi} > 2+2r$ when $s > 2+2r$.)

A data-dependent choice of m for computation of $\widehat{d}(r)$ is obtained by plugging $\widehat{b}_{2+2r}$ into (5.7):

$$\widehat{m}_{opt} = \left[\left(\frac{c_r}{16(1+r)\tau_r^2\widehat{b}_{2+2r}^2}\right)^{1/(5+4r)} n^{(4+4r)/(5+4r)}\right]. \tag{5.9}$$

This specification for $\widehat{m}_{opt}$ differs from those proposed in Henry and Robinson (1996).

Theorem 2 and Lemma 3(a) can be used to obtain new results for the local Whittle estimator $d^*(0)$ that is analyzed in Robinson (1995a). By Lemma 3(a), $d^*(0)$ is in the interior of $[d_1, d_2]$ wp$\rightarrow$ 1. Since there are no parameters θ to consider in this case, $d^*(0)$ satisfies the FOCs (2.5) wp$\rightarrow$ 1. Hence, the conditions of Theorem 2 are satisfied and $d^*(0)$ is asymptotically normal.

Corollary 3 *Under Assumptions* 1–5, $m^{1/2}(d^*(0) - d_0) - \nu_n(0,s)/4 \rightarrow_d N(0, 1/4)$ *as* $n \rightarrow \infty$.

Remarks 1. The "asymptotic bias" of $d^*(0)$ is

$$m^{-1/2}\nu_n(0,s)/4 = -1(s \geq 2)(2\pi^2/9)(m^2/n^2)b_2 + 1(1 \leq s < 2)O(m^s/n^s).$$

2. Corollary 3 shows that the local Whittle estimator of d_0 has the same asymptotic bias as that of the GPH estimator when $s \geq 2$, but smaller asymptotic variance. The latter is well-known, but the former is a new result. This result implies that the local Whittle estimator dominates the GPH estimator in terms of asymptotic mean-squared error (where the latter is defined to be the second moment of the asymptotic distribution of the estimator) provided m is chosen appropriately.

3. Robinson (1995a) does not provide an expression for the asymptotic bias of the local Whittle estimator. His Assumption A4$'$ restricts the growth rate of m such that $\nu_n(0,s) = o_p(1)$.

6 Optimal Rate of Convergence

In this section, we show that the LPW–FOC estimator attains the optimal rate of convergence for estimation of d_0 established in Andrews and Guggenberger (1999) for Gaussian processes. In fact, the LPW–FOC estimator attains this rate whether or not the process is Gaussian. This is an advantage of the LPW–FOC estimator over the bias-reduced estimator considered in Andrews and Guggenberger (1999), which is shown to attain the optimal rate for Gaussian processes. The optimal rate established in Andrews and Guggenberger (1999) is related to, and relies on, results of Giraitis, Robinson, and Samarov (1997).

We consider a minimax risk criterion with 0–1 loss. The class of spectral density functions that are considered includes functions that are smooth of order $s \geq 1$. The optimal rate is $n^{-s/(2s+1)}$, which is arbitrarily close to the parametric rate $n^{-1/2}$ if s is arbitrarily large. We show that the LPW–FOC estimator, $\widehat{d}(r)$, attains this rate when r is the largest integer less than $s/2$ and m is chosen appropriately.

Let s and the elements of $a = (a_0, a_{00}, a_1, ..., a_{[s/2]})'$, $\delta = (\delta_1, \delta_2, \delta_3)'$, and $K = (K_1, K_2, K_3)'$ be positive finite constants with $a_0 < a_{00}$ and $\delta_1 < 1/2$. We consider the following class of spectral densities:

$$\begin{aligned}
\mathcal{F}(s, a, \delta, K) = \{ & f : f(\lambda) = |\lambda|^{-2d_f}\varphi(\lambda),\ |d_f| \leq 1/2 - \delta_1,\ \int_{-\pi}^{\pi} f(\lambda)d\lambda \leq K_1, \text{ and} \\
& \varphi \text{ is an even function on } [-\pi, \pi] \text{ that satisfies (i) } a_0 \leq \varphi(0) \leq a_{00}, \\
& \text{(ii) } \varphi(\lambda) = \varphi(0) + \sum_{k=1}^{[s/2]} \varphi_k \lambda^{2k} + \Delta(\lambda) \text{ for some constants } \varphi_k \text{ with } |\varphi_k| \leq a_k \text{ for} \\
& k = 1, ..., [s/2] \text{ and some function } \Delta(\lambda) \text{ with } |\Delta(\lambda)| \leq K_2\lambda^s \text{ for all } 0 \leq \lambda \leq \delta_2, \\
& \text{(iii) } |\varphi(\lambda_1) - \varphi(\lambda_2)| \leq K_3|\lambda_1 - \lambda_2| \text{ for all } 0 < \lambda_1 < \lambda_2 \leq \delta_3 \}. \qquad (6.1)
\end{aligned}$$

If φ is an even function on $[-\pi, \pi]$ that is smooth of order $s \geq 1$ at zero and $f(\lambda) = |\lambda|^{-2d_f}\varphi(\lambda)$ for some $|d_f| < 1/2$, then f is in $\mathcal{F}(s, a, \delta, K)$ for some a, δ, and K. Condition (ii) of $\mathcal{F}(s, a, \delta, K)$ holds in this case by taking a Taylor expansion of $\varphi(\lambda)$ about $\lambda = 0$. The constants φ_k equal $\varphi^{(2k)}(0)/(2k)!$ for $k = 1, ..., [s/2]$ and $\Delta(\lambda)$ is the remainder in the Taylor expansion. Condition (iii) of $\mathcal{F}(s, a, \delta, K)$ holds in this case by a mean value expansion because φ has a bounded first derivative in a neighborhood of zero.

The optimal rate results are given in the following Theorem. Part (a) is from Theorem 3 of Andrews and Guggenberger (1999).

Theorem 3 *Let s and the elements of $a = (a_0, a_{00}, a_1, ..., a_{[s/2]})'$, $\delta = (\delta_1, \delta_2, \delta_3)'$, and $K = (K_1, K_2, K_3)'$ be any positive real numbers with $s \geq 1$, $a_0 < a_{00}$, $\delta_1 < 1/2$, and $K_1 \geq 2\pi a_{00}$.*
(a) *Suppose $\{x_t\}$ is a sequence of Gaussian random variables with spectral density function $f \in \mathcal{F}(s, a, \delta, K)$. Then, there is a constant $C > 0$ such that*

$$\liminf_{n\to\infty} \inf_{\widehat{d}_n} \sup_{f \in \mathcal{F}(s,a,\delta,K)} P_f(n^{s/(2s+1)}|\widehat{d}_n - d_f| \geq C) > 0,$$

where the inf *is taken over all estimators* $\widehat{d}_n$ *of* d_f *and* P_f *denotes probability when the true spectral density is* f.

(b) *Suppose* $\{x_t\}$ *is a sequence of random variables that has spectral density function* $f \in \mathcal{F}(s,a,\delta,K)$ *and satisfies Assumptions 3 and 5 with* θ_0 *in Assumption 5 defined by* $\theta_{0,k} = -\varphi_k$ *for* $k = 1,...,r$ *and* $\theta_0 = (\theta_{0,1},...,\theta_{0,r})'$. *Let* $m = D_0 n^{2s/(2s+1)}$ *for some constant* $D_0 \in (0,\infty)$ *and let* $r \geq 0$ *be the largest integer (strictly) less than* $s/2$. *Then,*

$$\lim_{C\to\infty} \limsup_{n\to\infty} \sup_{f\in\mathcal{F}(s,a,\delta,K)} P_f(n^{s/(2s+1)}|\widehat{d}(r) - d_f| \geq C) = 0.$$

Remark Part (b) of the Theorem is proved by showing that $\Psi_n := m^{1/2}(\widehat{d}(r)-d_f) - [\Omega_r^{-1}\nu_n(r,s)]_1$ is asymptotically normal *uniformly over* $f \in \mathcal{F}(s,a,\delta,K)$, where $[v]_1$ denotes the first element of the vector v. For a fixed spectral density f, asymptotic normality of Ψ_n is established by showing that the normalized score, $B_n^{-1}S_n(d_0,\theta_0)$, can be written as $\sum_{u=1}^4 T_{u,n}$, where $T_{1,n} = o_p(1)$, $T_{2,n} = o(1)$, $T_{3,n} \to_d N(0,\Omega_r)$, and $T_{4,n} + \nu_n(r,s) \to 0$, see the proof of Lemma 2(e). Hence, asymptotic normality of Ψ_n is driven by the term $T_{3,n}$. The key to the proof of part (b) is that the distribution of $T_{3,n}$ does not depend on f. One obtains asymptotic normality of Ψ_n uniformly over $f \in \mathcal{F}(s,a,\delta,K)$ provided the other terms behave appropriately uniformly over $f \in \mathcal{F}(s,a,\delta,K)$.

7 Monte Carlo Simulations

In this section, we present some simulation results that illustrate that the asymptotic results derived in the previous sections are indicative of finite sample performance, at least in the limited number of cases considered. Andrews and Sun (2001) provide simulation results for a much wider variety of cases. The simulation results also show that the LPW estimators perform well in terms of MSE in comparison to other semiparametric estimators of d_0, including the bias-reduced GPH estimators of Andrews and Guggenberger (1999) and the broad-band FEXP estimator considered by Moulines and Soulier (1999, 2000), Hurvich and Brodsky (2000), and Hurvich (2000).

We consider Gaussian first-order autoregressive fractionally integrated (ARFIMA (1, 1, 0)) processes with autoregressive parameter ϕ, long memory parameter d_0, and sample size $n = 512$. We consider three values of d_0, viz., $-.4$, 0, and .4. None of the results are sensitive to the value of d_0, so we just report results for $d_0 = 0$. We consider seven values of ϕ, viz., 0, .3, .6, .9, $-.3$, $-.6$, and $-.9$. Ten thousand simulation repetitions are employed.

We consider LPW estimators with $r = 0$, 1, and 2. We calculate the biases, standard deviations, and root mean squared errors (RMSEs) of these three estimators as functions of m. We also calculate the coverage probabilities and average lengths of nominal 90% confidence intervals constructed using these three estimators as functions of m. The confidence intervals (CIs) are based on the asymptotic normality result of Corollary 2. The CIs are given by

$$\left[\widehat{d}(r) - z_{.95}\left[(B_n^{-1}J_nB_n^{-1})^{-1}\right]_{11}/m^{1/2},\ \widehat{d}(r) + z_{.95}\left[(B_n^{-1}J_nB_n^{-1})^{-1}\right]_{11}/m^{1/2}\right] \quad (7.1)$$

for $r = 0, 1, 2$, where $z_{.95}$ is the .95 quantile of the standard normal distribution and $[A]_{11}$ denotes the $(1, 1)$ element of the matrix A.[2]

Figure 1 reports results for the case of an ARFIMA(1, 1, 0) process with $d_0 = 0$ and AR parameter $\phi = .6$. Figure 1(a) shows that the bias of $\widehat{d}(0)$ is large and increases rapidly with m. The biases of $\widehat{d}(1)$ and $\widehat{d}(2)$ are substantially smaller than that of $\widehat{d}(0)$ and they increase less rapidly with m. On the other hand, Figure 1(b) shows that the standard deviation of $\widehat{d}(0)$ is smaller than that of $\widehat{d}(1)$ and $\widehat{d}(2)$. As expected, the standard deviations of all three estimators decrease with m. Figure 1(c) shows that the minimum RMSE across different values of m is lowest for $\widehat{d}(2)$ and highest for $\widehat{d}(0)$. The RMSE functions are noticeably flatter for $\widehat{d}(1)$ and $\widehat{d}(2)$ than for $\widehat{d}(0)$, which means that a reasonable choice of m is easier to obtain for the former estimators than the latter.

Figure 1(d) shows that the CI coverage probabilities are fairly close to the nominal value of .9 provided m is not taken too large. The range of values of m that are not "too large" is much wider for the CIs based on $\widehat{d}(1)$ and $\widehat{d}(2)$ than for the CI based on $\widehat{d}(0)$. Figure 1(e) shows that the superior performance of the coverage probabilities of $\widehat{d}(1)$ and $\widehat{d}(2)$ comes at the expense of having longer CIs on average than those based on $\widehat{d}(0)$.

For brevity, we do not provide figures analogous to Figure 1 for other values of the AR parameter ϕ, but we comment on these figures briefly. The bias curves are steeper for $\phi = .9$ and less steep for other values of ϕ. For all values of ϕ except $\phi = 0$, $\widehat{d}(1)$ and $\widehat{d}(2)$ have noticeably smaller biases than $\widehat{d}(0)$. For $\phi = -.9, ..., .3$, $\widehat{d}(1)$ and $\widehat{d}(2)$ essentially eliminate the bias of $\widehat{d}(0)$ over a very wide range of values of m. The standard deviation and average length of CIs figures are essentially the same for all values of ϕ. The RMSE figures differ with ϕ. For $\phi = -.9, ..., .3$ and the estimators $\widehat{d}(1)$ and $\widehat{d}(2)$, the coverage probability figures show that the true and nominal coverage probabilities are fairly close to each other over a wide range of values of m (noticeably wider than in Figure 1(d) for $\phi = .6$). Except when $\phi = 0$, the range of values of m that yield good coverage probabilities for $\widehat{d}(0)$ is much more restrictive. When $\phi = .9$, ranges of values of m that yield good coverage probabilities for $\widehat{d}(r)$ for $r = 0, 1, 2$ are more restrictive than in Figure 1(d) for $\phi = .6$.

Next, we compare the RMSE performance of the three LPW estimators $\widehat{d}(0)$, $\widehat{d}(1)$, and $\widehat{d}(2)$ with three bias-reduced GPH estimators indexed by $r = 0$, 1, and 2 (where $r = 0$ yields the standard GPH estimator) and the broad-band FEXP estimator, as defined in Hurvich and Brodsky (2000). Table 1 reports the minimum RMSE for each estimator over m values in the range $[8, 256]$ for the LPW and GPH estimators and over 0 to 16 terms in the expansion for the FEXP estimator. Table 1 shows that the smallest RMSE is attained by an LPW or local Whittle estimator for every value of ϕ. When ϕ is non-zero, the LPW estimators $\widehat{d}(1)$ and/or $\widehat{d}(2)$ are best and $\widehat{d}(1)$ provides the best overall performance. When ϕ is zero, the local Whittle estimator $\widehat{d}(0)$ is best. For each value of r, the LPW estimator $\widehat{d}(r)$ has smaller minimum

RMSE than the bias-reduced GPH estimator with the same value of r for all values of ϕ. The FEXP estimator has higher minimum RMSE than $\widehat{d}(0)$ for all values of ϕ and higher minimum RMSE than $\widehat{d}(1)$ for all values of ϕ except $\phi = 0$.

The results of Table 1 show that LPW estimators are competitive with existing semiparametric estimators of d_0 in terms of finite sample performance. For simple ARFIMA(1, 1, 0) processes, they have smaller minimum RMSE than bias-reduced GPH estimators and the FEXP estimator.

8 Proofs

Proof of Lemma 1. Let $\Gamma_{n0} = \{\gamma \in \Gamma : ||B_n(\gamma - \gamma_0)|| \leq K_n,\ ||\gamma - \gamma_0|| < \delta\}$ for some $\delta > 0$ such that $L_n(\gamma)$ is twice differentiable on $\{\gamma \in R^k : ||\gamma - \gamma_0|| < \delta\}$ and $\{\gamma \in R^k : ||\gamma - \gamma_0|| < \delta\} \subset \Gamma$. Using condition (iv), a Taylor expansion about γ_0, and some algebra, we obtain: for $\gamma \in \Gamma_{n0}$,

$$\begin{aligned} L_n(\gamma) - L_n(\gamma_0) &= \nabla L_n(\gamma_0)'(\gamma - \gamma_0) + \frac{1}{2}(\gamma - \gamma_0)'\nabla^2 L_n(\gamma_0)(\gamma - \gamma_0) + \rho_n(\gamma) \\ &= \frac{1}{2}(B_n(\gamma - \gamma_0) + Y_n)'[(B_n^{-1})'\nabla^2 L_n(\gamma_0)B_n^{-1}](B_n(\gamma - \gamma_0) + Y_n) \\ &\quad -\frac{1}{2}Y_n'(B_n^{-1})'\nabla L_n(\gamma_0) + \rho_n(\gamma), \end{aligned} \tag{8.1}$$

where for all $\gamma \in \Gamma_{n0}$,

$$\begin{aligned} |\rho_n(\gamma)| &\leq \sup_{\overline{\gamma} \in \Gamma_{n0}} |(\gamma - \gamma_0)'(\nabla^2 L_n(\overline{\gamma}) - \nabla^2 L_n(\gamma_0))(\gamma - \gamma_0)| \\ &\leq ||B_n(\gamma - \gamma_0)||^2 \sup_{\overline{\gamma} \in \Gamma_{n0}} ||(B_n^{-1})'(\nabla^2 L_n(\overline{\gamma}) - \nabla^2 L_n(\gamma_0))B_n^{-1}|| \\ &= ||B_n(\gamma - \gamma_0)||^2 o_p(1). \end{aligned} \tag{8.2}$$

Let $\gamma_n^* = \gamma_0 - B_n^{-1}Y_n$. Conditions (ii) and (iii) imply that $Y_n = O_p(1)$. This and condition (i) imply that $\gamma_n^* \in \Gamma_{n0}$ wp$\rightarrow$ 1. In consequence, by (8.1) and (8.2),

$$\begin{aligned} L_n(\gamma_n^*) - L_n(\gamma_0) &= -\frac{1}{2}Y_n'(B_n^{-1})'\nabla L_n(\gamma_0) + \rho_n(\gamma_n^*) \text{ and} \\ \rho_n(\gamma_n^*) &= o_p(1). \end{aligned} \tag{8.3}$$

For any $\varepsilon > 0$ and $n \geq 1$, let $\Gamma_n(\varepsilon) = \{\gamma \in \Gamma : ||B_n(\gamma - \gamma_0) + Y_n|| \leq \varepsilon\}$. Note that γ_n^* is in the interior of $\Gamma_n(\varepsilon)$ wp$\rightarrow$ 1. We have $\Gamma_n(\varepsilon) \subset \Gamma_{n0}$ wp$\rightarrow$ 1, and so, $\sup_{\gamma \in \Gamma_n(\varepsilon)} |\rho_n(\gamma)| = o_p(1)$ by (8.2). Let $\partial\Gamma_n(\varepsilon)$ denote the boundary of $\Gamma_n(\varepsilon)$. Combining (8.1)–(8.3), for $\gamma \in \partial\Gamma_n(\varepsilon)$,

$$L_n(\gamma) - L_n(\gamma_n^*) = \frac{1}{2}\mu_n'(B_n^{-1})'\nabla^2 L_n(\gamma_0)B_n^{-1}\mu_n + o_p(1) \tag{8.4}$$

for some k-vector μ_n with $||\mu_n|| = \varepsilon > 0$. The right-hand side is bounded away from zero wp$\rightarrow$ 1 uniformly over all k-vectors μ_n with $||\mu_n|| = \varepsilon$ by condition (iii). Hence, the minimum of $L_n(\gamma)$ over $\gamma \in \partial\Gamma_n(\varepsilon)$ is greater than its value at the interior point

γ_n^*. In consequence, the minimum of $L_n(\gamma)$ over $\gamma \in \Gamma_n(\varepsilon)$ is attained at a point, say $\widetilde{\gamma}_n(\varepsilon)$, (not necessarily unique) in the interior of $\Gamma_n(\varepsilon)$ wp$\rightarrow$ 1. This point satisfies the first-order conditions $\nabla L_n(\widetilde{\gamma}_n(\varepsilon)) = 0$ wp$\rightarrow$ 1.

In consequence, for all $J \geq 1$, $P(\nabla L_n(\widetilde{\gamma}_n(1/j)) = 0\ \forall j = 1, ..., J)) \rightarrow 1$ as $n \rightarrow \infty$. Thus, there is a sequence $\{J_n : n \geq 1\}$ such that $J_n \uparrow \infty$ and $P(\nabla L_n(\widetilde{\gamma}_n(1/j)) = 0\ \forall j = 1, ..., J_n)) \rightarrow 1$ as $n \rightarrow \infty$. For example, take $J_1 = 2$, $J_n = J_{n-1} + 1$ if $P(\nabla L_n(\widetilde{\gamma}_n(1/j)) = 0\ \forall j \leq J_{n-1} + 1)) > 1 - 1/J_{n-1}$, and $J_n = J_{n-1}$ otherwise, for $n = 2, 3, ...$. Define $\widetilde{\gamma}_n = \widetilde{\gamma}_n(1/J_n)$ for $n \geq 1$. Then, $P(\nabla L_n(\widetilde{\gamma}_n) = 0) \geq 1 - 1/J_{n-1} \rightarrow 1$ as $n \rightarrow \infty$. In addition, $\widetilde{\gamma}_n \in \Gamma_n(1/J_n)$ for all $n \geq 1$. Hence, $B_n(\widetilde{\gamma}_n - \gamma_0) = -Y_n + o_p(1) = O_p(1)$. $\square$

We prove Lemma 2(a)-2(d) in reverse order.

Proof of Lemma 2(d). Part (d) holds by approximating sums by integrals. See Andrews and Guggenberger (1999, Lemma 2(a), (h), and (i)) for details (noting that $X_j = -2\log\lambda_j$ in Andrews and Guggenberger (1999)). $\square$

Proof of Lemma 2(c). The normalized Hessian can be written as

$$\begin{aligned} B_n^{-1} H_n(d,\theta) B_n^{-1} &= \widehat{G}^{-2}(d,\theta) \left(\widehat{G}(d,\theta) m^{-1} \sum_{j=1}^{m} y_j(d,\theta) \widetilde{X}_j \widetilde{X}_j' \right. \\ &\qquad \left. - \left(m^{-1} \sum_{j=1}^{m} y_j(d,\theta) \widetilde{X}_j \right) \left(m^{-1} \sum_{j=1}^{m} y_j(d,\theta) \widetilde{X}_j \right)' \right), \text{ where} \\ \widetilde{X}_j &= (2\log j, (j/m)^2, ..., (j/m)^{2r})'. \end{aligned} \tag{8.5}$$

Let

$$\widehat{G}_{a,b}(d,\theta) = m^{-1} \sum_{j=1}^{m} I_j \exp(p_r(\lambda_j,\theta)) \lambda_j^{2d} (2\log j)^a (j/m)^{2b} \tag{8.6}$$

for $a = 0, 1, 2$ and $b = 0, ..., r$. The $(1,1)$, $(1,k)$, and (k,ℓ) elements of $B_n^{-1} H_n(d,\theta) B_n^{-1}$ for $k, l = 2, ..., r+1$ are

$$\begin{aligned} &\widehat{G}_{0,0}^{-2}(\widehat{G}_{0,0}\widehat{G}_{2,0} - \widehat{G}_{1,0}^2), \\ &\widehat{G}_{0,0}^{-2}(\widehat{G}_{0,0}\widehat{G}_{1,k-1} - \widehat{G}_{1,0}\widehat{G}_{0,k-1}), \text{ and} \\ &\widehat{G}_{0,0}^{-2}(\widehat{G}_{0,0}\widehat{G}_{0,k+\ell-2} - \widehat{G}_{0,k-1}\widehat{G}_{0,\ell-1}), \end{aligned} \tag{8.7}$$

respectively, where the dependence on (d,θ) has been suppressed for simplicity.

Define $J_{a,b}$ as $\widehat{G}_{a,b}(d,\theta)$ is defined, but with $I_j \exp(p_r(\lambda_j,\theta))\lambda_j^{2d}$ replaced by G_0. That is,

$$J_{a,b} = G_0 m^{-1} \sum_{j=1}^{m} (2\log j)^a (j/m)^{2b} \tag{8.8}$$

for $a = 0, 1, 2$ and $b = 0, ..., r$. The elements of $B_n^{-1} J_n B_n^{-1}$ are given by the formulae in (8.7) with $\widehat{G}_{a,b}(d_0, \theta_0)$ replaced by $J_{a,b}$. Note that $J_{a,b} = O_p(\log^a m)$ and $J_{0,0} = G_0 > 0$. Hence, to prove Lemma 2(c), it suffices to show that

$$\Delta_{a,b} := |\widehat{G}_{a,b}(d_0, \theta_0)/G_0 - J_{a,b}/G_0| = o_p(\log^{-2} m) \tag{8.9}$$

for $a = 0, 1, 2$ and $b = 0, ..., r$.

Let

$$g_j = \lambda_j^{-2d_0} G_0 \exp(-p_r(\lambda_j, \theta_0)). \tag{8.10}$$

By summation by parts, we have

$$\begin{aligned}
\Delta_{a,b} &= |m^{-1} \sum_{j=1}^{m} (\frac{I_j}{g_j} - 1)(2\log j)^a (\frac{j}{m})^{2b}| \\
&\leq \left| m^{-1} \sum_{k=1}^{m-1} [(2\log k)^a (\frac{k}{m})^{2b} - (2\log(k+1))^a (\frac{k+1}{m})^{2b}] \sum_{j=1}^{k} (\frac{I_j}{g_j} - 1) \right| \\
&\quad + \left| (2\log m)^a m^{-1} \sum_{j=1}^{m} (\frac{I_j}{g_j} - 1) \right| \\
&:= \zeta_{1,m} + \zeta_{2,m}.
\end{aligned} \tag{8.11}$$

Using the triangle inequality and then mean-value expansions, we obtain

$$\begin{aligned}
\zeta_{1,m} &\leq m^{-1} \sum_{k=1}^{m-1} \left(\left| (2\log k)^a (\frac{k}{m})^{2b} - (2\log k)^a (\frac{k+1}{m})^{2b} \right| \right. \\
&\quad \left. + \left| (2\log k)^a (\frac{k+1}{m})^{2b} - (2\log(k+1))^a (\frac{k+1}{m})^{2b} \right| \right) \left| \sum_{j=1}^{k} (\frac{I_j}{g_j} - 1) \right| \\
&\leq 2^a m^{-1} \sum_{k=1}^{m-1} \left((\log k)^a 2b (\frac{k+1}{m})^{2b-1} m^{-1} + a(\log(k+1))^{a-1} k^{-1} (\frac{k+1}{m})^{2b} \right) \\
&\quad \times \left| \sum_{j=1}^{k} (\frac{I_j}{g_j} - 1) \right| \\
&\leq 2^a (\log m)^a (2b + a) m^{-1} \sum_{k=1}^{m-1} k^{-1} \left| \sum_{j=1}^{k} (\frac{I_j}{g_j} - 1) \right|.
\end{aligned} \tag{8.12}$$

By altering the statement and proof of (4.8) of Robinson (1995a) and using (4.9) of Robinson (1995a) without change, we obtain:

$$\text{(i)} \ \sum_{j=1}^{k} (\frac{I_j}{g_j} - 2\pi I_{\varepsilon j}) = O_p(k^{1/3} \log^{2/3} k + k^{\phi+1} n^{-\phi} + k^{1/2} n^{-1/4}) \text{ and}$$

(ii) $\sum_{j=1}^{k}(2\pi I_{\varepsilon j}-1)=O_p(k^{1/2})$, where

$$I_{\varepsilon j}=|w_\varepsilon(\lambda_j)|^2 \text{ and } w_\varepsilon(\lambda)=(2\pi n)^{-1/2}\sum_{t=1}^{n}\varepsilon_t e^{it\lambda}, \tag{8.13}$$

as $n\to\infty$ uniformly over $k=1,...,m$. The requisite alteration to (4.8) is that $k^{\phi+1}n^{-\phi}$ replaces Robinson's $k^{\beta+1}n^{-\beta}$ in the remainder of result (i). This occurs because our Assumption 2 assumes that $\varphi(\lambda)$ is smooth of order s at zero, which differs from Robinson's Assumption A2′. In consequence, the second equation on p. 1648 of Robinson's proof of (4.8) needs to be changed. We need to show that $\sum_{j=1}^{k}|1-g_j/f_j|=O(k^{\phi+1}n^{-\phi})$ uniformly over $k=1,...,m$. This holds by a Taylor expansion of $\log\varphi(\lambda)$ to order $2r$ with remainder $O(\lambda_j^{\phi})$ and some calculations.

Combining (8.11)–(8.13), $\zeta_{1,m}$ and $\zeta_{2,m}$ are $O_p((\log^a m)m^{-1/2}+(\log^a m)m^{\phi}n^{-\phi})$ $=o_p(\log^{-2}m)$, where the equality uses Assumption 4. □

Proof of Lemma 2(b). By (8.9) and $J_{a,b}=O_p(\log^a m)$, we obtain $\widehat{G}_{a,b}(d_0,\theta_0)=O_p(\log^a m)$ for $a=0,1,2$ and $b=0,...,r$ and $\widehat{G}_{0,0}(d_0,\theta_0)=G_0+o_p(\log^{-2}m)$, where $G_0>0$. These results and (8.7) imply that it suffices to show that

$$\sup_{\theta\in\Theta}|\widehat{G}_{a,b}(d_0,\theta)-\widehat{G}_{a,b}(d_0,\theta_0)|=o_p(\log^{-2}m) \tag{8.14}$$

for all $a=0,1,2$ and $b=0,...,r$. The left-hand side of (8.14) equals

$$\begin{aligned}
&\sup_{\theta\in\Theta}|m^{-1}\sum_{j=1}^{m}I_j\left[\exp(p_r(\lambda_j,\theta))-\exp(p_r(\lambda_j,\theta_0))\right]\lambda_j^{2d_0}(2\log j)^a(j/m)^{2b}|\\
&\le \sup_{\theta\in\Theta,k=1,...,m}|\exp\{p_r(\lambda_k,\theta)-p_r(\lambda_k,\theta_0)\}-1|\,m^{-1}\sum_{j=1}^{m}I_j\exp(p_r(\lambda_j,\theta_0))\lambda_j^{2d_0}(2\log j)^a\\
&=O(\lambda_m^2)\widehat{G}_{a,0}(d_0,\theta_0),\\
&=O_p((m/n)^2(\log^a m))\\
&=o_p(\log^{-2}m),
\end{aligned} \tag{8.15}$$

where the first equality holds by a mean-value expansion using the compactness of Θ, the second equality holds by (8.9) and $J_{a,b}=O_p(\log^a m)$, and the third equality holds by Assumption 4. □

Proof of Lemma 2(a). We have (i) $\widehat{G}_{a,b}(d_0,\theta)=J_{a,b}+o_p(\log^{-2}m)$ by (8.9) and (8.14), (ii) $J_{a,b}=O(\log^a m)$, (iii) $J_{0,0}J_{2,0}-J_{1,0}^2=O(1)$ by elementary calculations replacing sums by integrals and noting that the part of $J_{0,0}J_{2,0}$ that is $O_p(\log^2 m)$ cancels with an identical term in $J_{1,0}^2$, (iv) $J_{0,0}J_{1,k-1}-J_{1,0}J_{0,k-1}=O(1)$ by the same sort of argument as for (iii), and (v) $J_{0,0}=G_0>0$. Given (i)-(v) and (8.7), to establish Lemma 2(a) it suffices to show that

$$\sup_{d\in D_m(\eta_n),\theta\in\Theta}|\widehat{G}_{a,b}(d,\theta)-\widehat{G}_{a,b}(d_0,\theta)|=o_p(\log^{-2}m). \tag{8.16}$$

Define $\widehat{E}_{a,b}(d,\theta)$ as $\widehat{G}_{a,b}(d,\theta)$ is defined, but with λ_j^{2d} replaced by j^{2d}. The formulae in (8.7) for the elements of $B_n^{-1}H_n(d,\theta)B_n^{-1}$ also hold with $\widehat{G}_{a,b}(d,\theta)$ replaced by $\widehat{E}_{a,b}(d,\theta)$. Hence, it suffices to show that

$$Z_{a,b}(\eta_n) := \sup_{d\in D_m(\eta_n),\theta\in\Theta} |\widehat{E}_{a,b}(d,\theta) - \widehat{E}_{a,b}(d_0,\theta)| = o_p(n^{2d_0}\log^{-2} m) \tag{8.17}$$

for all $a = 0, 1, 2$, and $b = 0, ..., r$.

We note that in Robinson's (1995a) proof of the asymptotic normality of the local Whittle estimator $\widetilde{H}$ (using his notation) he shows that the Hessian is well behaved for $H \in M = \{H : (\log^3 m)|H - H_0| \le \varepsilon\}$ on p. 1642 and he shows that $(\log^3 m)(\widetilde{H} - H_0) = o_p(1)$. There is a slight error in his proof (which can be fixed without difficulty) that leads us to define $D_m(\eta_n)$ with $\log^5 m$ rather than $\log^3 m$ in the statement of Lemma 2(a) and to show that $\widetilde{d}(r) - d_0 = o_p(\log^{-5} m)$ rather than $o_p(\log^{-3} m)$ in Lemma 3. In particular, the second equality in his equation following (4.9) on p. 1643 is not correct. The left-hand side of this equality is unchanged if E is replaced by F and $o_p(n^{2H_0-1})$ is replaced by $o_p(1)$ throughout. The problem in his proof is that $\widehat{F}_2(H_0) = O_p(\log^2 m)$, not $O_p(1)$, so that $\widehat{F}_2(H_0)o_p(1) = o_p(\log^2 m)$, not $o_p(1)$, as is necessary for the stated equality to hold. To obtain the desired result, one needs to show that $\widehat{E}_k(\widetilde{H}) - \widehat{E}_k(H_0) = o_p(n^{2H_0-1}\log^{-k} m)$ for $k = 0, 1, 2$, rather than $o_p(n^{2H_0-1})$, in (4.4) on p. 1642. This can be achieved by (i) redefining M on p. 1642 to be $M = \{H : (\log^5 m)|H - H_0| \le \varepsilon\}$ and (ii) showing that $(\log^5 m)|H - H_0| = o_p(1)$. The latter holds by the same argument as given by Robinson (1995a, pp. 1642-3) except that the left-hand side of (4.6) needs to be $o_p(\log^{-10} m)$, which holds by the argument given on p. 1643.

The proof of (8.17) is similar to a proof of Robinson (1995a, p. 1642). We have

$$\begin{aligned} Z_{a,b}(\eta_n) &= \sup_{d\in D_m(\eta_n),\theta\in\Theta} |m^{-1}\sum_{j=1}^{m} I_j \exp(p_r(\lambda_j,\theta))(2\log j)^a (j/m)^{2b} j^{2d_0}(j^{2(d-d_0)} - 1)| \\ &\le C \sup_{d\in D_m(\eta_n)} m^{-1}\sum_{j=1}^{m} I_j(\log j)^a j^{2d_0}|j^{2(d-d_0)} - 1| \\ &\le 2Ce^{2\eta_n \log^{-4} m} \sup_{d\in D_m(\eta_n)} m^{-1}\sum_{j=1}^{m} I_j(\log j)^{a+1} j^{2d_0}|d - d_0| \\ &\le \eta_n(\log^{-2} m)2Ce^{2\eta_n \log^{-4} m} m^{-1}\sum_{j=1}^{m} I_j\lambda_j^{2d_0}(2\pi/n)^{-2d_0} \end{aligned} \tag{8.18}$$

for some constant $C < \infty$, where the first inequality uses the fact that $\sup_{0\le\lambda\le 2\pi,\theta\in\Theta} \exp(p_r(\lambda,\theta)) < \infty$ since Θ is compact, the second inequality uses $|j^{2(d-d_0)} - 1| / |d - d_0| \le 2m^{2|d-d_0|}\log j \le 2m^{2\eta_n \log^{-5} m}\log j = 2e^{2\eta_n\log^{-4} m}\log j$ for $d \in D_m(\eta_n)$ by a mean-value expansion and using $m^{\log^{-1} m} = e$, and the third inequality uses $d \in D_m(\eta_n)$. We have $m^{-1}\sum_{j=1}^{m} I_j\lambda_j^{2d_0} = \widehat{G}_{0,0}(d_0, 0) = G_0 + o_p(\log^{-2} m)$ by (8.9) and (8.14). In consequence, the left-hand side of (8.18) is $o_p(n^{2d_0}\log^{-2} m)$, as desired. □

Proof of Lemma 2(e). Using (4.2) and (8.10), the normalized score is

$$\begin{aligned} B_n^{-1}S_n(d_0,\theta_0) &= \widehat{G}^{-1}(d_0,\theta_0)m^{-1/2}\sum_{j=1}^{m}\left(y_j(d_0,\theta_0)-m^{-1}\sum_{k=1}^{m}y_k(d_0,\theta_0)\right)\widetilde{X}_j \\ &= (1+o_p(1))m^{-1/2}\sum_{j=1}^{m}(\frac{I_j}{g_j}-1)\left(\widetilde{X}_j-m^{-1}\sum_{k=1}^{m}\widetilde{X}_k\right), \qquad (8.19)\end{aligned}$$

where the second equality uses $\widehat{G}(d_0,\theta_0)=\widehat{G}_{0,0}(d_0,\theta_0)=G_0+o_p(1)$ by (8.9). The right-hand side, with $(1+o_p(1))$ deleted, can be written as

$$\begin{aligned} & T_{1,n}+T_{2,n}+T_{3,n}+T_{4,n}, \text{ where} \\ T_{1,n} &= m^{-1/2}\sum_{j=1}^{m}\left(\frac{I_j}{g_j}-2\pi I_{\varepsilon j}-E(\frac{I_j}{g_j}-2\pi I_{\varepsilon j})\right)\left(\widetilde{X}_j-m^{-1}\sum_{k=1}^{m}\widetilde{X}_k\right), \\ T_{2,n} &= m^{-1/2}\sum_{j=1}^{m}\left(\frac{EI_j}{f_j}-1\right)\frac{f_j}{g_j}\left(\widetilde{X}_j-m^{-1}\sum_{k=1}^{m}\widetilde{X}_k\right), \\ T_{3,n} &= m^{-1/2}\sum_{j=1}^{m}(2\pi I_{\varepsilon j}-1)\left(\widetilde{X}_j-m^{-1}\sum_{k=1}^{m}\widetilde{X}_k\right), \\ T_{4,n} &= m^{-1/2}\sum_{j=1}^{m}\left(\frac{f_j}{g_j}-1\right)\left(\widetilde{X}_j-m^{-1}\sum_{k=1}^{m}\widetilde{X}_k\right), \qquad (8.20)\end{aligned}$$

and $f_j=f(\lambda_j)$, using the fact that $E2\pi I_{\varepsilon j}=1$. We show that $T_{1,n}=o_p(1)$, $T_{2,n}=o(1)$, $T_{3,n}\rightarrow_d N(0,\Omega_r)$, and $T_{4,n}=-\nu_n(r,s)+o(1)$.

To show $T_{1,n}=o_p(1)$, we use the following result, which is proved below:

$$\sum_{j=1}^{k}\left(\frac{I_j}{g_j}-2\pi I_{\varepsilon j}-E(\frac{I_j}{g_j}-2\pi I_{\varepsilon j})\right)=O_p(k^{1/3}\log^{2/3}k+k^{\phi+1/2}n^{-\phi}+k^{1/2}n^{-1/4}) \qquad (8.21)$$

as $n\rightarrow\infty$ uniformly over $k=1,...,m$. By summation by parts,

$$\begin{aligned} T_{1,n} &= m^{-1/2}\sum_{k=1}^{m-1}\sum_{j=1}^{k}\left(\frac{I_j}{g_j}-2\pi I_{\varepsilon j}-E(\frac{I_j}{g_j}-2\pi I_{\varepsilon j})\right)\left(\widetilde{X}_k-\widetilde{X}_{k+1}\right) \\ &\quad +m^{-1/2}\sum_{j=1}^{m}\left(\frac{I_j}{g_j}-2\pi I_{\varepsilon j}-E(\frac{I_j}{g_j}-2\pi I_{\varepsilon j})\right)\left(\widetilde{X}_m-m^{-1}\sum_{k=1}^{m}\widetilde{X}_k\right) \\ &= m^{-1/2}\sum_{k=1}^{m-1}O_p(k^{1/3}\log^{2/3}k+k^{\phi+1/2}n^{-\phi}+k^{1/2}n^{-1/4})O(k^{-1}) \\ &\quad +m^{-1/2}O_p(m^{1/3}\log^{2/3}m+m^{\phi+1/2}n^{-\phi}+m^{1/2}n^{-1/4})O(1) \\ &= O_p(m^{-1/6}\log^{2/3}m+(m/n)^{\phi}+n^{-1/4}) \\ &= o_p(1), \qquad (8.22)\end{aligned}$$

where the second equality uses $\widetilde{X}_k - \widetilde{X}_{k+1} = O(k^{-1})$ uniformly over $k = 1, ..., m$ (because $\log k - \log(k+1) = O(k^{-1})$ by a mean-value expansion and $|(k/m)^{2i} - ((k+1)/m)^{2i}| = (k/m)^{2i}|1-(1+k^{-1})^{2i}| = O(k^{-1})$ for $i = 1, ..., r$) and $\widetilde{X}_m - m^{-1}\sum_{k=1}^m \widetilde{X}_k = O(1)$ because $\log m - m^{-1}\sum_{k=1}^m \log k = \log m - m^{-1}(m \log m - m + O(\log m)) = 1 + O((\log m)/m)$ by approximating sums by integrals (e.g., see (6.11) of Andrews and Guggenberger (1999)) and $(m/m)^{2i} - m^{-1}\sum_{j=1}^m (j/m)^{2i} = O(1)$ for $i = 1, ..., r$.

To show $T_{2,n} = o(1)$, we use the result that

$$EI_j/f_j = 1 + O(j^{-1}\log j) \tag{8.23}$$

uniformly over $j = 1, ..., m$. Because Assumptions 1 and 2 imply Assumptions 1–3 of Robinson (1995b), this holds by Theorem 2 of Robinson (1995b) using the normalization of I_j by f_j rather than $G_0\lambda_j^{-2d_0}$. The remainder term in (8.23) is different from that in Theorem 2 of Robinson (1995b) because the proof of (8.23) only requires (4.1), and not (4.2), of Robinson (1995b) to hold, given the normalization by f_j.

By (8.23),

$$\begin{aligned} T_{2,n} &= m^{-1/2}\sum_{j=1}^m O(j^{-1}\log j)O(1)(\widetilde{X}_j - m^{-1}\sum_{k=1}^m \widetilde{X}_k) \\ &= O(m^{-1/2}\log m \sum_{j=1}^m j^{-1}\log j) \\ &= O(m^{-1/2}\log^3 m) = o(1). \end{aligned} \tag{8.24}$$

We show that $\beta' T_{3,n} \rightarrow_d N(0, \beta'\Omega_r\beta)$ for all $\beta \neq 0$ using the same proof as Robinson's (1995a, pp. 1644-47) proof that $m^{-1/2}\sum_{j=1}^m (2\pi I_{\varepsilon j} - 1)2\nu_j \rightarrow_d N(0,4)$, except with Robinson's $2\nu_j = 2\log j - m^{-1}\sum_{k=1}^m 2\log k$ replaced by $\zeta_j = \beta'(\widetilde{X}_j - m^{-1}\sum_{k=1}^m \widetilde{X}_k)$. Robinson's proof goes through with the asymptotic variance 4 replace by $\beta'\Omega_r\beta$ because (i) $m^{-1}\sum_{j=1}^m \zeta_j^2 \rightarrow \beta'\Omega_r\beta$ as $n \rightarrow \infty$ by Lemma 2(d) and (ii) $|\zeta_j - \zeta_{j+1}| \leq ||\beta|| \cdot ||\widetilde{X}_j - \widetilde{X}_{j+1}|| \leq Cj^{-1}$ for some constant $C < \infty$ independent of j, which is needed in (4.21) of Robinson's proof.

Next, we show that $T_{4,n} = -\nu_n(r,s) + o_p(1)$. By (3.1),

$$\begin{aligned} \log(f_j/g_j) &= \log\varphi(\lambda_j) - \log G_0 + p_r(\lambda_j, \theta_0) \\ &= 1(s \geq 2+2r)\frac{b_{2+2r}}{(2+2r)!}\lambda_j^{2+2r} + O(\lambda_j^q) \text{ and} \\ f_j/g_j &= 1 + 1(s \geq 2+2r)\frac{b_{2+2r}}{(2+2r)!}\lambda_j^{2+2r} + O(\lambda_j^q), \text{ where} \\ q &= \min\{s, 4+2r\}, \end{aligned} \tag{8.25}$$

uniformly over $j = 1, ..., m$, using $e^x = 1 + x + x^2 e^{x_*}/2$ for x_* between 0 and x. (If $s = 2+2r$, the remainder $O(\lambda_j^q)$ is actually $o(\lambda_j^q) = o(\lambda_j^{2+2r})$.) Hence, if $s \geq 2+2r$,

$$T_{4,n} = m^{-1/2}\sum_{j=1}^m \left(\frac{b_{2+2r}}{(2+2r)!}\lambda_j^{2+2r} + O(\lambda_j^q)\right)\left(\widetilde{X}_j - m^{-1}\sum_{k=1}^m \widetilde{X}_k\right)$$

$$= m^{5/2+2r} n^{-(2+2r)} m^{-1} \sum_{j=1}^{m} \frac{(2\pi)^{2+2r} b_{2+2r}}{(2+2r)!} (\frac{j}{m})^{2+2r} \left(\widetilde{X}_j - m^{-1} \sum_{k=1}^{m} \widetilde{X}_k \right) \quad (8.26)$$

$$+ m^{-1/2} \sum_{j=1}^{m-1} (\widetilde{X}_j - \widetilde{X}_{j+1}) \sum_{i=1}^{j} O(\lambda_i^q) + m^{-1/2} \sum_{j=1}^{m} O(\lambda_j^q) \left(\widetilde{X}_m - m^{-1} \sum_{k=1}^{m} \widetilde{X}_k \right).$$

where the second equality uses summation by parts. The second and third summands on the right-hand side of (8.26) are $O(m^{q+1/2} n^{-q})$ because $\widetilde{X}_j - \widetilde{X}_{j+1} = O(j^{-1})$ uniformly over $j = 1, ..., m$ and $\widetilde{X}_m - m^{-1} \sum_{k=1}^{m} \widetilde{X}_k = 1 + o(1)$ by the calculations following (8.22).

The following results are proved by approximating sums by integrals, see Andrews and Guggenberger (1999, Pf. of Lemma 1) for details. Suppose $m \to \infty$, then for $k = 1, ..., r$,

$$\begin{aligned} \frac{1}{m} \sum_{j=1}^{m} (\frac{j}{m})^{2+2r} \left((\frac{j}{m})^{2k} - \frac{1}{m} \sum_{i=1}^{m} (\frac{i}{m})^{2k} \right) &= \frac{1}{2r+2k+3} - \frac{1}{(3+2r)(2k+1)} + o(1) \\ &= \frac{(2+2r)}{(3+2r)^2} \xi_{r,k} + o(1) \text{ and} \\ \frac{1}{m} \sum_{j=1}^{m} (\frac{j}{m})^{2+2r} \left(2 \log j - \frac{1}{m} \sum_{i=1}^{m} 2 \log i \right) &= \frac{2(2r+2)}{(3+2r)^2} + o(1). \end{aligned} \quad (8.27)$$

For the case where $s > 2 + 2r$, the combination of (8.26) and (8.27) gives $T_{4,n} = -\nu_n(r, s) + o(1)$, using Assumption 4. When $s = 2 + 2r$, the term $O(m^{q+1/2} n^{-q})$ is really $o(m^{q+1/2} n^{-q})$ in (8.26) and the latter is $o(1)$ using Assumption 4. Hence, in this case too, $T_{4,n} = -\nu_n(r, s) + o(1)$.

When $2r < s < 2 + 2r$, $T_{4,n}$ is given by the right-hand side of (8.26) with the term that contains b_{2+2r} deleted and with $q = s$. Hence, by the remarks following (8.26), $T_{4,n} = O(m^{s+1/2} n^{-s}) = -\nu_n(r, s)$.

Now we prove (8.21). Parts of the proof are similar to parts of Robinson's (1995a) proof of his equation (4.8). Let $\ell = k^{1/3} \log^{2/3} k$. We have

$$\sum_{j=1}^{\ell} \left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j} - E(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j}) \right) = O_p(k^{1/3} \log^{2/3} k) \text{ as } n \to \infty \quad (8.28)$$

by the same argument as in Robinson (1995a, p. 1648). We write

$$\begin{aligned} &\sum_{j=\ell+1}^{k} \left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j} - E(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j}) \right) \\ &= \sum_{j=\ell+1}^{k} \left((\frac{I_j}{f_j} - 2\pi I_{\varepsilon j}) \frac{f_j}{g_j} - E(\frac{I_j}{f_j} - 2\pi I_{\varepsilon j}) \frac{f_j}{g_j} \right) + 2\pi \sum_{j=\ell+1}^{k} (I_{\varepsilon j} - E I_{\varepsilon j}) \left(\frac{f_j}{g_j} - 1 \right) \\ &:= A_1 + A_2. \end{aligned} \quad (8.29)$$

We have

$$EA_1^2 \le E\left(\sum_{j=\ell+1}^{k}(\frac{I_j}{f_j}-2\pi I_{\varepsilon j})\frac{f_j}{g_j}\right)^2 = O(k^{2/3}\log^{4/3}k + kn^{-1/2}), \qquad (8.30)$$

where the equality holds by the same proof as in Robinson (1995a, pp. 1648–51) for the quantity given in the third equation on his p. 1648. The only difference is that the factor f_j/g_j does not appear in Robinson (1995a). It can be shown that this factor has no impact on the proof because $f_j/g_j = 1+o(1)$ uniformly over $j = 1, ..., m$.

Next, we have

$$\begin{aligned} EA_2^2 &= 4\pi^2 \sum_{j=\ell+1}^{k} \mathrm{Var}(I_{\varepsilon j})\left(\frac{f_j}{g_j}-1\right)^2 + 8\pi^2 \sum_{i=\ell+1}^{k}\sum_{j=\ell+1}^{i-1} \mathrm{Cov}(I_{\varepsilon i}, I_{\varepsilon j})\left(\frac{f_i}{g_i}-1\right)\left(\frac{f_j}{g_j}-1\right) \\ &= O(1)\sum_{j=\ell+1}^{k}\lambda_j^{2\phi} + O(n^{-1})\sum_{i=\ell+1}^{k}\sum_{j=\ell+1}^{i-1}\lambda_i^{\phi}\lambda_j^{\phi} \\ &= O(k(\frac{k}{n})^{2\phi}) + O(\frac{k^2}{n}(\frac{k}{n})^{2\phi}) \\ &= O(k(\frac{k}{n})^{2\phi}), \end{aligned} \qquad (8.31)$$

where the second equality uses (8.25) and Theorem 5.2.4 of Brillinger (1975, p. 125), which states that $\mathrm{Var}(I_{\varepsilon j}) = O(1)$ and $\mathrm{Cov}(I_{\varepsilon i}, I_{\varepsilon j}) = O(n^{-1})$ uniformly over $i, j = 1, ..., n$. Brillinger's Assumption 2.6.2(1) imposes strict stationarity, which does not hold in the present case. However, his proof only requires fourth-order stationarity. The fourth order cumulant spectrum of $\{\varepsilon_t : t = 1, 2, ...\}$ is the same as that of an iid process with finite fourth moment, which is sufficient.

Combining (8.28)–(8.31) gives (8.21). □

Proof of Lemma 3. To prove part (a), we first establish that $d^*(r)$ is consistent. Due to the non-uniform behavior of $R_r(d,\theta)$ around $d = d_0 - 1/2$, it is necessary to divide the interval $[d_1, d_2]$ into two parts: D_1 and D_2, where

$$\begin{aligned} D_1 &= [\max(d_1, d_0 - 1/2 + \omega), d_2] \text{ and} \\ D_2 &= [d_1, \max(d_1, d_0 - 1/2 + \omega)) \end{aligned} \qquad (8.32)$$

for some small $\omega > 0$.

Let $N_\delta = \{d : |d - d_0| < \delta\}$ for $0 < \delta < 1/2$. We show that $P(d^*(r) \in \overline{N}_\delta) = o(1)$, where $\overline{N}_\delta$ denotes the complement of N_δ. We have

$$\begin{aligned} P(d^*(r) \in \overline{N}_\delta) &\le P\left(\inf_{d\in\overline{N}_\delta\cap D_1} R_r(d,\theta^*(r)) - R_r(d_0,\theta^*(r)) \le 0\right) \\ &\quad + P\left(\inf_{d\in D_2} R_r(d,\theta^*(r)) - R_r(d_0,\theta^*(r)) \le 0\right). \end{aligned} \qquad (8.33)$$

To prove that the first term on the right-hand side of (8.33) is $o_p(1)$, we use Robinson's (1995a, Theorem 1) proof that

$$P\left(\inf_{d\in\overline{N}_\delta\cap D_1} R_r(d,0)-R_r(d_0,0)\le 0\right)=o(1). \tag{8.34}$$

His proof is based on writing $R_r(d,0)-R_r(d_0,0)$ as $U(d)-T(d)$ and showing that $\inf_{d\in\overline{N}_\delta\cap D_1} U(d)>0$ and $\sup_{d\in D_1}|T(d)|=o_p(1)$. We write

$$\begin{aligned}&R_r(d,\theta^*(r))-R_r(d_0,\theta^*(r))\\&=U(d)-T(d)+\left[(R_r(d,\theta^*(r))-R_r(d,0))-(R_r(d_0,\theta^*(r))-R_r(d_0,0))\right].\end{aligned} \tag{8.35}$$

If the last term on the right-hand side of (8.35) is $o_p(1)$ uniformly over $d\in D_1$, then the rest of Robinson's proof of (8.34) goes through without change. The desired property follows from

$$\begin{aligned}\sup_{d\in[d_1,d_2],\theta\in\Theta}|R_r(d,\theta)-R_r(d,0)| &= \sup_{d\in[d_1,d_2],\theta\in\Theta}|\log(\widehat{G}(d,\theta)/\widehat{G}(d,0))-\frac{1}{m}\sum_{j=1}^m p_r(\lambda_j,\theta)|\\&=\log(1+O_p(\lambda_m^2))+O(\lambda_m^2)\\&=O_p(\lambda_m^2)=o_p(1).\end{aligned} \tag{8.36}$$

The second equality holds because $\sup_{\theta\in\Theta}|m^{-1}\sum_{j=1}^m p_r(\lambda_j,\theta)|=O(\lambda_m^2)$ and a mean value expansion of $\exp(\cdot)$ about zero gives

$$\begin{aligned}\exp(p_r(\lambda_j,\theta)) &= 1+(1+c_j(\theta))p_r(\lambda_j,\theta),\ \text{where}\\ \sup_{\theta\in\Theta,j=1,\ldots,m}|c_j(\theta)| &= o(1),\end{aligned} \tag{8.37}$$

which implies that $\sup_{d\in[d_1,d_2],\theta\in\Theta}|\widehat{G}(d,\theta)/\widehat{G}(d,0)-1|=O_p(\lambda_m^2)=o_p(1)$.

It remains to show that the second term on the right-hand side of (8.33) is $o(1)$. We assume that $\max(d_1,d_0-1/2+\omega)=d_0-1/2+\omega$; otherwise, D_2 is an empty set and the conclusion is trivially true. As in Robinson's (1995a) proof on pp. 1638–40, we put $p=\exp(m^{-1}\sum_{k=1}^m\log k)$ and write

$$R_r(d,\theta^*(r))-R_r(d_0,\theta^*(r))=\log\hat{D}(d)/\hat{D}(d_0), \tag{8.38}$$

where $\hat{D}(d)$ is defined in our case by

$$\begin{aligned}\hat{D}(d) &= m^{-1}\sum_{j=1}^m I_j\exp\left(\tilde{p}(\lambda_j,\theta^*(r))\right)\left(\frac{j}{p}\right)^{2d-2d_0} j^{2d_0}\ \text{and}\\ \tilde{p}(\lambda_j,\theta^*(r)) &= p(\lambda_j,\theta^*(r))-m^{-1}\sum_{k=1}^m p(\lambda_k,\theta^*(r)).\end{aligned} \tag{8.39}$$

It is easy to see that

$$\inf_{d\in D_2}\hat{D}(d)\geq m^{-1}\sum_{j=1}^{m}I_j\exp\left(\tilde{p}(\lambda_j,\theta^*(r))\right)a_j j^{2d_0},\ \text{where}$$

$$a_j=\begin{cases}\left(\frac{j}{p}\right)^{2\omega-1} & \text{for } 1\leq j\leq p\\ \left(\frac{j}{p}\right)^{2d_1-2d_0} & \text{for } p<j\leq m.\end{cases}\tag{8.40}$$

Thus, $P(\inf_{d\in D_2}R_r(d,\theta^*(r))-R_r(d_0,\theta^*(r))\leq 0)$ is bounded by

$$P\left(m^{-1}\sum_{j=1}^{m}I_j\exp\left(\tilde{p}(\lambda_j,\theta^*(r))\right)(a_j-1)j^{2d_0}\leq 0\right).\tag{8.41}$$

To show that (8.41) is $o(1)$, we use Robinson's proof that

$$P\left(m^{-1}\sum_{j=1}^{m}I_j(a_j-1)j^{2d_0}\leq 0\right)=o(1).\tag{8.42}$$

His proof remains valid if we replace I_j by $I_j\exp\left(\tilde{p}(\lambda_j,\theta^*(r))\right)$ and $I_{\varepsilon j}$ by $I_{\varepsilon j}\exp\left(\tilde{p}(\lambda_j,\theta^*(r))\right)$. The reason is that $\exp\left(\tilde{p}(\lambda_j,\theta^*(r))\right)=1+o(1)$ uniformly over $j\leq m$. The only step that is not obvious is to show that $m^{-1}\sum_{j=1}^{m}(a_j-1)\left\{2\pi I_{\varepsilon j}\exp(\tilde{p}(\lambda_j,\theta^*(r)))-1\right\}=o_p(1)$. In fact,

$$\begin{aligned}&m^{-1}\sum_{j=1}^{m}(a_j-1)\left\{2\pi I_{\varepsilon j}\exp(\tilde{p}(\lambda_j,\theta^*(r)))-1\right\}\\&=m^{-1}\sum_{j=1}^{m}(a_j-1)2\pi I_{\varepsilon j}\left\{\exp(\tilde{p}(\lambda_j,\theta^*(r)))-1\right\}+m^{-1}\sum_{j=1}^{m}(a_j-1)(2\pi I_{\varepsilon j}-1)\\&=o_p\left(m^{-1}\sum_{j=1}^{m}(a_j+1)2\pi I_{\varepsilon j}\right)+o_p(1),\end{aligned}\tag{8.43}$$

where the last equality follows from Robinson's proof and the fact $\exp(\tilde{p}(\lambda_j,\theta^*(r)))=1+o(1)$. It suffices to show that $m^{-1}\sum_{j=1}^{m}(a_j+1)2\pi I_{\varepsilon j}=O(1)$. This follows from $\sum_{j=1}^{m}a_j=O(m)$ and $I_{\varepsilon_j}=O(1)$.

Next, we use Robinson's (1995a) proof of the $\log^3 m$-consistency of $d^*(0)$ (given on his pp. 1642–43) to obtain $\log^5 m$-consistency of $d^*(r)$. His proof parallels his consistency proof except that he needs to show that $\sup_{d\in D_1\cap N_\delta}|T(d)|=o_p(\log^{-6}m)$ for some $\delta>0$. His proof can be used to show that $d^*(r)$ is $\log^5 m$–consistent by replacing $T(d)$ by $T(d)+R_r(d,\theta^*(r))-R_r(d,0)-(R_r(d_0,\theta^*(r))-R_r(d_0,0))$ as above and by replacing $o_p(\log^{-6}m)$ by $o_p(\log^{-10}m)$. Using (8.36) and Assumption 4, this term is $o_p(\log^{-10}m)$, as desired.

The only other change to Robinson's proof that is needed is to take account of the difference between his Assumption A4$'$, which requires $m^{1+2\beta}(\log^2 m)/n^{2\beta} \to 0$, and our weaker Assumption 4, which requires $m^{\phi+1/2}/n^{\phi} = O(1)$. For Robinson's proof to go through (to show $\log^5 m$-consistency of $d^*(r)$), his (4.7) needs to be $o_p(\log^{-10} m)$. Given his equations (4.8) and (4.9), this holds provided $(\log^{10} m)(m/n)^{\beta} = o(1)$. Our Assumption 2 implies that Robinson's Assumption A1$'$ and (4.8) hold with $\beta = \min\{s, 2\}$. Our Assumption 4 implies that $(\log^{10} m)(m/n)^{\beta} = o(1)$. Hence, Robinson's proof goes through under our weaker assumption on m.

Part (b) of the Lemma is proved in the text. $\square$

Proof of Theorem 3(b). The choice of r as the largest integer less than $s/2$ implies that $s > 2r$ and $s \leq 2+2r$. Hence, $\nu_n(r,s) = O(m^{s+1/2}n^{-s}) = O(1)$. By the definition of m, $m^{1/2} = D_0^{1/2} n^{s/(2s+1)}$. In consequence, the result of Theorem 3(b) follows from

$$\sup_{f \in \mathcal{F}(s,a,\delta,K)} \left| P_f \left(\left(\begin{array}{c} m^{1/2}(\widehat{d}(r) - d_0) \\ m^{1/2} Diag\left((2\pi m/n)^2, ..., (2\pi m/n)^{2r}\right)\left(\widehat{\theta}(r) - \theta_0\right) \end{array} \right) - \Omega_r^{-1}\nu_n(r,s) \leq x \right) - \Phi(\Omega_r^{1/2} x) \right| \to 0 \tag{8.44}$$

as $n \to \infty$ for all $x \in R^{r+1}$.

To prove (8.44), we use the results of Sections 4 and 5. We show that these results hold uniformly over $f \in \mathcal{F}(s,a,\delta,K)$. To this end, we note that although $\varphi(\lambda)$ is not necessarily smooth of order s for $f \in \mathcal{F}(s,a,\delta,K)$, conditions (ii) and (iii) of $\mathcal{F}(s,a,\delta,K)$ provide the Taylor expansion of $\log\varphi(\lambda)$ which is all that is needed in the proofs of Lemma 2(c) and (e), where smoothness of order s is used.

Let unif–f abbreviate uniformly over $f \in \mathcal{F}(s,a,\delta,K)$.

The proof of Lemma 1 goes through unif–f provided conditions (ii)–(iv) hold unif–f and $\{K_n : n \geq 1\}$ in condition (iv) does not depend on f. In consequence, the conclusion of the Lemma is that a solution to the FOCs holds wp$\to$ 1 unif–f and $B_n(\widetilde{\gamma}_n - \gamma_0) = -Y_n + o_p(1)$ unif–f. To verify that conditions (ii)–(iv) of Lemma 1 hold unif–f when $L_n(\gamma) = mR_r(d,\theta)$, we need to show that parts (a)–(c) of Lemma 2 hold with $o_p(1)$ holding unif–f. Inspection of their proofs shows that they do. Part (d) of Lemma 2 does not depend on f, so uniformity over f is not an issue.

Inspection of the proof of part (e) of Lemma 2 shows that $T_{1,n} = o_p(1)$ unif–f; $T_{2,n} = o(1)$ unif–f because Theorem 2 of Robinson (1995b) holds unif–f by Lemma 3(b) of Andrews and Guggenberger (1999); and $T_{4,n} = -\nu_n(r,s) + o(1)$ unif–f using the definition of $\mathcal{F}(s,a,\delta,K)$. The term $T_{3,n}$, which is asymptotically normal, does not depend on f. Hence, its distribution function differs from that of a normal distribution function unif–f trivially. Combining these results yields

$$\sup_{f \in \mathcal{F}(s,a,\delta,K)} |P_f(B_n^{-1}S_n(d_0,\theta_0) + \nu_n(r,s) \leq x) - \Phi(\Omega_r^{-1/2}x)| \to 0 \text{ as } n \to \infty \tag{8.45}$$

for all $x \in R^{r+1}$. Inspection of the proof of Lemma 3 shows that it holds unif–f.

Given the above unif–f extensions of the results of Lemmas 1–3, Theorems 1 and 2 and Corollaries 1 and 2 have analogous extensions. The results of the extended Theorems 1 and 2 are the same as that of (8.44) with $\widehat{d}(r)$ replaced by $\widetilde{d}(r)$ and $\overline{d}(r)$ respectively. The result of the extended Corollary 2 is exactly (8.44). $\square$

Footnotes

[1] The authors thank Patrik Guggenberger for pointing out some errors in an earlier version of this paper and Carol Copeland for proofreading the paper. Andrews thanks the National Science Foundation for research support via grant number SBR–9730277. Sun thanks the Cowles Foundation for support under a Cowles prize. The authors' email addresses are donald.andrews@yale.edu and yixiao.sun@yale.edu.

[2] The confidence intervals are constructed using $[(B_n^{-1} J_n B_n^{-1})^{-1}]_{11}$ to estimate the standard error of $m^{1/2}(\widehat{d}(r) - d_0)$ rather than $[\Omega_r^{-1}]_{11} = c_r/4$, which appears in the asymptotic normality result of Corollary 2 (and is the limit as $n \to \infty$ of $[(B_n^{-1} J_n B_n^{-1})^{-1}]_{11}$, see Lemma 2(d)), because $[(B_n^{-1} J_n B_n^{-1})^{-1}]_{11}$ is closer to the finite sample Hessian matrix than is $[\Omega_r^{-1}]_{11}$ and it performs better in finite samples.

References

Andrews, D. W. K. and P. Guggenberger (1999). A bias-reduced log-periodogram regression estimator for the long-memory parameter. Cowles Foundation Discussion Paper No. 1263, Yale University.

Andrews, D. W. K. and Y. Sun (2001). Finite sample comparisons of semiparametric estimators of the long-memory parameter. Manuscript in preparation, Cowles Foundation, Yale University.

Agiakloglou, C., Newbold, P., and Wohar, M. (1993). Bias in an estimator of the fractional difference parameter. *Journal of Time Series Analysis* **14** 235–246.

Beran, J. (1993). Fitting long-memory models by generalized linear regression. *Biometrika* **80** 817–822.

Bhansali, R. J. and Kokoszka, P. S. (1997). Estimation of the long memory parameter by fitting fractional autoregressive models. Working paper, University of Liverpool.

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

Crowder, M. J. (1976). Maximum likelihood estimation with dependent observations. *Journal of the Royal Statistical Society, Series B* **38** 45–53.

Diggle, P. (1990). *Time Series: A Biostatistical Introduction.* Oxford University Press, Oxford.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long-memory time series models. *Journal of Time Series Analysis* **4** 221–237.

Giraitis, L., Robinson, P. M., and Samarov, A. (1997). Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long-range dependence. *Journal of Time Series Analysis* **18** 49–60.

Heijmans, R. D. H. and Magnus, J. R. (1986). On the first-order efficiency and asymptotic normality of maximum likelihood estimators obtained from dependent observations. *Statistica Nederlandica* **40.**

Henry, M. and Robinson, P. M. (1996). Bandwidth choice in Gaussian semiparametric estimation of long range dependence. In *Athens Conference on Applied Probability and Time Series, Volume II: Time Series in Memory of E. J. Hannan* (P. M. Robinson and M. Rosenblatt, eds.). Springer-Verlag, New York.

Hurvich, C. M. (2000). Model selection for broadband semiparametric estimation of long memory in time series. Statistics and Operations Research Working Paper No. 2000–4, L. N. Stern School of Business, New York University.

Hurvich, C. M. and Brodsky, J. (2000). Broadband semiparametric estimation of the memory parameter of a long-memory time series using fractional exponential models. Statistics and Operations Research Working Paper No. 2000–2, L. N. Stern School of Business, New York University.

Hurvich, C. M. and Deo, R. S. (1999). Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long memory time series. *Journal of Time Series Analysis* **20** 331–341.

Künsch, H. R. (1987). Statistical aspects of self-similar processes. In *Proceedings of the First World Congress of the Bernoulli Society* (Yu. Prohorov and V. V. Sazanov, eds.) **1** 67–74. VNU Science Press, Utrecht.

Moulines, E. and Soulier, P. (1999). Broadband log-periodogram regression of time series with long-range dependence. *Annals of Statistics* **27** 1415–1439.

Moulines, E. and Soulier, P. (2000). Data-driven order selection for long range dependent time series. *Journal of Time Series Analysis* **21** 193–218.

Robinson, P. (1994). Semiparametric analysis of long-memory time series. *Annals of Statistics* **22** 515–539.

Robinson, P. (1995a). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* **23** 1630–1661.

Robinson, P. (1995b). Log periodogram regression of time series with long range dependence. *Annals of Statistics* **23** 1048–1072.

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

Shimotsu, K. and Phillips, P. C. B. (1999). Modified local Whittle estimation of the memory parameter in the nonstationary Case. Working paper, Cowles Foundation, Yale University.

Weiss, L. (1971). Asymptotic properties of maximum likelihood estimators in some nonstandard cases I. *Journal of the American Statistical Association,* **66** 345–350.

Weiss, L. (1973). Asymptotic properties of maximum likelihood estimators in some nonstandard cases II. *Journal of the American Statistical Association,* **68** 428–430.

Wooldridge, J. M. (1994). Estimation and inference for dependent processes. In *Handbook of Econometrics,* Vol. IV (R. F. Engle and D. L. McFadden, eds.). North Holland, Amsterdam.

TABLE 1

Minimum RMSE for Several Semiparametric Estimators of d_0 *for ARFIMA(1, 1, 0) Processes with AR Parameter* ϕ and $d_0 = 0^a$

	ϕ						
Estimator	0	.3	.6	.9	−.3	−.6	−.9
LPW-0	.033*	.081	.126	.297	.067	.072	.073
LPW-1	.052	.075*	.116	.276*	.058*	.055*	.063*
LPW-2	.067	.075*	.115*	.280	.067	.067	.074
GPH-0	.042	.098	.150	.335	.077	.083	.084
GPH-1	.066	.093	.142	.327	.070	.069	.079
GPH-2	.084	.092	.140	.325	.084	.084	.090
FEXP	.047	.090	.138	.325	.086	.117	.170

[a] An asterisk denotes the smallest value in each column.

FIG. 1. *Performances of LPW estimators with $r = 0, 1, 2$ as functions of the bandwidth m for an ARFIMA(1,1,0) process with $d_0 = 0$, AR parameter $\phi = 0.6$, and sample size $n = 512$.*