ADAPTIVE LOCAL POLYNOMIAL WHITTLE ESTIMATION OF LONG-RANGE DEPENDENCE

BY

DONALD W.K. ANDREWS
AND
YIXIAO SUN

COWLES FOUNDATION PAPER NO. 1080

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281
2004

Econometrica, Vol. 72, No. 2 (March, 2004), 569–614

ADAPTIVE LOCAL POLYNOMIAL WHITTLE ESTIMATION OF LONG-RANGE DEPENDENCE

BY DONALD W. K. ANDREWS AND YIXIAO SUN[1]

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. We specify a data-dependent adaptive procedure that adjusts the degree of the polynomial to the smoothness of $\varphi(\lambda)$ at zero and selects the bandwidth. The resulting "adaptive LPW" estimator is shown to achieve the optimal rate of convergence, which depends on the smoothness of $\varphi(\lambda)$ at zero, up to a logarithmic factor.

KEYWORDS: Adaptive estimator, asymptotic bias, asymptotic normality, bias reduction, local polynomial, long memory, minimax rate, optimal bandwidth, Whittle likelihood.

1. INTRODUCTION

IN THIS PAPER, we consider estimation of the long-memory parameter d_0 for a stationary process $\{x_t\}$. The spectral density, $f(\lambda)$, of $\{x_t\}$ is taken to be of the form

$$f(\lambda) = |\lambda|^{-2d_0}\varphi(\lambda), \tag{1.1}$$

where $d_0 \in [d_1, d_2]$, $-1/2 < d_1 < d_2 < 1/2$, and $0 < \varphi(0) < \infty$.

The parameter d_0 determines the long-memory properties of $\{x_t\}$ and $\varphi(\lambda)$ determines its short-run dynamics. To maintain generality of the short-run dynamics of $\{x_t\}$, we do not impose a specific functional form on $\varphi(\lambda)$. Instead, we take $\varphi(\lambda)$ to belong to a family that is characterized by regularity conditions near frequency zero. This is a narrow-band semiparametric 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 Porter-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

[1]The authors thank Patrik Guggenberger, John Geweke, Doug Hodgson, the editor, and three anonymous referees for comments; and Carol Copeland for proofreading the paper. Andrews thanks the National Science Foundation for research support via Grant Numbers SBR-9730277 and SES-0001706. Sun thanks the Cowles Foundation for support under a Cowles prize.

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.

Motivation for LPW estimators comes from the nonparametric regression literature. In that literature, one of the most popular estimators, especially among the cognoscente, is a local polynomial estimator. For example, see Fan (1992) and references in Härdle and Linton (1994). Also, the LPW estimator can be viewed as a semiparametric local (to frequency zero) version of an approximate maximum likelihood estimator of a parametric FEXP model considered by Diggle (1990) and Beran (1993) that utilizes polynomials, rather than trigonometric polynomials.

Let m denote the number of frequencies near zero used in the LPW likelihood. Results of Andrews and Guggenberger (2003) (AG), which in turn rely on results of Giraitis, Robinson, and Samarov (1997), show that the optimal choices of r and m (in terms of the rate of convergence of the estimator of d_0) depend on the smoothness of $\varphi(\lambda)$ at zero. We provide an adaptive estimator of d_0, denoted the ALPW estimator, based on the method of Lepskii (1990), that uses the data to select r and the exponent on m (although not the constant). This estimator is shown to obtain the optimal rate of convergence, up to a logarithmic factor. If $\varphi(\lambda)$ is infinitely smooth at zero, then this estimator is $n^{1/2-\delta}$-consistent for all $\delta > 0$ and, hence, has rate of convergence that is arbitrarily close to the parametric rate.

In comparison to the adaptive GPH estimator in Giraitis, Robinson, and Samarov (2000) (GRS), our estimator has the following advantages: (i) its rate of convergence is faster for spectral densities that are smooth of order s for $s > 2$, (ii) it does not delete two-thirds of the frequencies or require tapering, which avoids substantial inflation of its variance, (iii) its asymptotic properties are shown to hold for non-Gaussian processes with finite fourth moments, and (iv) it does not impose an upper bound on the amount of smoothness and, in consequence, achieves rates of convergence that are arbitrarily close to the parametric rate in the infinitely smooth case. Points (iii) and (iv) also are in contrast to the procedure of Lepskii (1990). Point (iii) is typical for local Whittle estimators; see Robinson (1995a). Part of the reason that we are able to establish advantages (iii) and (iv) is that we consider zero–one loss rather than squared-error loss. Results for zero–one loss are sufficient to obtain rates of convergence, which is the item of greatest interest here. On the other hand, computationally, the LPW estimator requires minimization of a globally convex criterion function whereas the adaptive GPH estimator has a closed form.

In comparison to the adaptive FEXP estimators of Iouditsky, Moulines, and Soulier (2002) (IMS) and Hurvich, Moulines, and Soulier (2002) (HMS), our

estimator has the following advantages: (i) its asymptotic properties hold without any restrictions on the spectral density of the process outside a neighborhood of the origin, which allows for a much wider class of processes, (ii) it does not require tapering or the deletion of some fraction of the frequencies, which circumvents inflation of its variance, and (iii) its asymptotic properties are shown to hold for non-Gaussian processes.

Our method suffers the same drawback as those of Lepskii (1990), GRS, IMS, and HMS in that there are constants in the adaptive procedure that are arbitrary. In the simulation section, we specify values of these constants that work fairly well in the contexts considered.

If desired, one can adjust the choice of m used by the ALPW estimator to be slightly smaller than the choice that is optimal (in terms of rate of convergence). The adjustment can be done so that the adaptive estimator with data-dependent r and m is asymptotically normal with zero asymptotic mean, but has a slightly slower rate of convergence than the optimal rate. This result holds for all finite values of the smoothness index of $\varphi(\lambda)$ at zero except for values in a set with Lebesgue measure zero.

Although the results of this paper are for stationary processes, they also can be utilized when the underlying process is nonstationary. Suppose the long-memory parameter d_0 lies in the interval $(-.5, 1.5)$, which is plausible for most economic data (and which corresponds to a nonstationary process when $d_0 \geq .5$). Suppose one has a preliminary consistent estimator of d_0 for d_0 in this range, such as the estimator of Velasco (1999), Velasco and Robinson (2000), or Shimotsu and Phillips (2002). Consider the following procedure: One first differences the data if this estimator exceeds .5 and one leaves the data as is otherwise. Then, one applies the adaptive LPW estimator to the data. If the data are differenced the estimator of d_0 equals the adaptive LPW estimator plus one. Otherwise the estimator of d_0 is just the adaptive LPW estimator.

With probability that goes to one as $n \to \infty$, the data are properly differenced or left in levels, provided $d_0 \neq .5$, and the results of this paper are applicable to the differenced or levels data. The advantage of applying the adaptive LPW estimator considered in this paper over other estimators, such as one that is consistent for $d_0 \in (-.5, 1.5)$, is that it achieves the optimal rate of convergence (up to a logarithmic factor). If one believes a priori that d_0 is near .5, then one can .5-difference the data (as in Gil-Alaña and Robinson (1997, p. 249)), rather than first difference it, and the estimator of d_0 is the adaptive LPW estimator plus .5.

The asymptotic properties of the adaptive LPW estimator are obtained by first establishing results for an LPW estimator with a fixed value of r and values of m that depend on the sample size n, but not on the data. Because these results are somewhat unusual, we briefly describe them here.

First, we concentrate out the constant G from the LPW log-likelihood. Then, for fixed nonnegative integer r, we let $(\widehat{d}(r), \widehat{\theta}(r))$ denote the LPW estimator that minimizes the (negative) concentrated LPW log-likelihood with respect

to (d, θ) over the parameter space $[d_1, d_2] \times \Theta$, where $\theta = (\theta_1, \ldots, \theta_r)'$ and Θ is a compact and convex subset of $\mathbb{R}^r$. One can show that the LPW estimator, $\widehat{d}(r)$, is consistent for d_0 by extending the argument of Robinson (1995a) (see Andrews and Sun (2001)). To establish asymptotic normality of $\widehat{d}(r)$, a typical argument would first establish consistency of $(\widehat{d}(r), \widehat{\theta}(r))$. But, showing that $(\widehat{d}(r), \widehat{\theta}(r))$ is consistent is problematic, because the concentrated LPW log-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 establish consistency and asymptotic normality of the LPW estimator simultaneously using the following steps. First, we show: (i) there exists a solution $(\widetilde{d}(r), \widetilde{\theta}(r))$ to the first-order conditions (FOC's) with probability that goes to one as $n \to \infty$ and this solution is consistent and asymptotically normal. The FOC approach is effective because one can use different normalizations of the FOC's for the different parameters $d, \theta_1, \ldots, \theta_r$. By doing so, one can ensure that the gradient and Hessian matrix of the normalized log-likelihood are asymptotically nondegenerate.

Next, we show: (ii) the (negative) concentrated LPW log-likelihood is a strictly convex function of (d, θ). This implies that it has a unique minimum. Furthermore, it implies that if there exists a solution to the FOC's, then it is unique and equals the minimizing value. In consequence, $(\widehat{d}(r), \widehat{\theta}(r))$ equals $(\widetilde{d}(r), \widetilde{\theta}(r))$ with probability that goes to one as $n \to \infty$ and, hence, 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 $s \geq 1$. Strict convexity of the LPW log-likelihood also yields obvious computational advantages for the LPW estimator over typical nonlinear estimators. Local minima and multiple solutions to the FOC's do not exist. (Note that Marinucci and Robinson (2001, p. 237) point out the convexity of the negative local Whittle log-likelihood.)

Our method of proof has some advantages even in the context of establishing consistency and asymptotic normality of the local Whittle estimator analyzed in Robinson (1995a). It is comparable to a proof of asymptotic normality with consistency given and, hence, circumvents the need for a separate proof of consistency, which occupies about six pages in Robinson (1995a).

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

$$m^{1/2}\big(\widehat{d}(r) - d_0\big) \to_d N(Ab_{2+2r}\tau_r, c_r/4) \quad \text{as} \quad n \to \infty, \tag{1.2}$$

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 (1.2), the rate of convergence of $\widehat{d}(r)$ is $n^{-(2+2r)/(5+4r)}$, whereas the rate of convergence for $\widehat{d}(0)$ is $n^{-2/5}$.

When r and m are selected adaptively, using the data, the rate of convergence of the (adaptive) estimator depends on the smoothness of $\varphi(\lambda)$. For example, if $\varphi(\lambda)$ is smooth of order s, then the rate of convergence is shown to be $n^{-s/(2s+1)}\zeta(n)$, where $\zeta(n)$ is less than $\log^2 n$. This is the optimal rate up to the factor $\zeta(n)$.

We note that the results of the paper provide some new results for the local Whittle estimator $\widehat{d}(0)$ 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, 1995b) 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). Also, the results of this paper establish the validity of an adaptive procedure for selecting m for the local Whittle estimator that is analogous to the method of GRS for the GPH estimator.

The results of this paper are similar to those of AG, who consider adding the regressors $\lambda_j^2, \ldots, \lambda_j^{2r}$ to a log-periodogram regression that is used to estimate d_0. But, AG does not consider *adaptive* selection of r and m. In addition, for fixed r, the estimator considered by AG has the same asymptotic bias as the LPW 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 estimator of AG are determined under the assumption of Gaussianity of $\{x_t\}$, whereas the properties of the LPW estimator considered here are determined without requiring $\{x_t\}$ to be Gaussian.

The LPW estimators considered in this paper also are related to estimators introduced by Robinson and Henry (2003). They consider a general class of semiparametric M-estimators of d_0 that utilize higher-order kernels to obtain bias-reduction like that of the LPW estimator. Other papers in the literature that are related to this paper include Henry and Robinson (1996) and Hurvich and Deo (1999). These papers approximate $\log \varphi(\lambda)$ by a more flexible function than a constant in order to obtain a data-driven choice of m.

In the paper, we compare the root mean-squared error performance of the ALPW estimator with the adaptive estimators of GRS and IMS, the FEXP estimator coupled with a local C_L criterion, as in Hurvich (2001), and the Gaussian ARFIMA$(1, d, 0)$ Whittle quasi-maximum likelihood (QML) estimator analyzed by Fox and Taqqu (1986). We consider two or three variants of each estimator—one that is theoretically justified by results in the literature (or is close to it) and one or more that are not. We consider three models: ARFIMA$(1, d, 0)$; DARFIMA$(1, d, 0)$, which is a model whose spectral density is *discontinuous* and equals that of the ARFIMA$(1, d, 0)$ model for frequencies on an interval $[0, \lambda_0]$ and zero elsewhere, where $\lambda_0 = \pi/2$ in the present case; and long-memory component (LMC) models, which are designed to have smoothness of the short-run component of the spectral density to be finite—equal to 1.5 in the present case. We consider three different distributions for the innovations: normal, t_5, and χ^2_2. Sample sizes $n = 512$ and $n = 4{,}096$ are considered.

The Monte Carlo results can be summarized as follows. (i) The simulation results are not sensitive to the value of d_0 (within the stationary region) or the innovation distribution. (ii) The RMSE of the ALPW estimator is lower than those of the theoretically-justified GRS and IMS adaptive estimators and the Hurvich (2001) FEXP estimator, often by a substantial margin, in all but a few of the fifty cases reported in the tables. Hence, of the theoretically-justified adaptive estimators or the Hurvich (2001) FEXP estimator, the ALPW estimator is clearly the best. (iii) Both trimming and tapering of the GRS estimator, as required for the theoretical results in GRS, hurt the performance of the adaptive GRS estimator. Similarly, tapering of the IMS estimator, as required for the theoretical results of IMS, hurts the performance of the adaptive IMS estimator. (iv) The best estimators in an overall sense are the ALPW estimator, the GRS estimator without trimming or tapering, and the IMS estimator without tapering but with some pooling. (v) As expected, the parametric Whittle QML estimator performs very well when the parametric model is correctly specified; moderately well when the degree of misspecification is moderate; and very poorly when the degree of misspecification is large.

The remainder of the paper is organized as follows. Section 2 defines the LPW log-likelihood function. Section 3 states the assumptions used. Section 4 shows that there exists a sequence of solutions to the FOC's that is consistent and asymptotically normal. Section 5 shows that this sequence is the LPW estimator and, hence, the LPW estimator is consistent and asymptotically normal. Section 6 establishes that the LPW estimator attains the optimal rate of convergence for estimation of d_0. Section 7 introduces the adaptive method for choosing the order, r, of the polynomial and the bandwidth, m. Section 8 provides the Monte Carlo simulation results. An Appendix contains proofs.

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

2. DEFINITION OF THE LPW LOG-LIKELIHOOD

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

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

$$I_j = |w_j|^2.$$

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

$$Q_r(d, G, \theta) = m^{-1} \sum_{j=1}^{m} \left\{ \log\left[G\lambda_j^{-2d} \exp(-p_r(\lambda_j, \theta))\right] + \frac{I_j}{G\lambda_j^{-2d} \exp(-p_r(\lambda_j, \theta))} \right\}, \quad \text{where} \tag{2.2}$$

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

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 log-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 (negative) concentrated LPW log-likelihood $R_r(d, \theta)$:

$$R_r(d, \theta) = \log \widehat{G}(d, \theta) - m^{-1} \sum_{j=1}^{m} p_r(\lambda_j, \theta) - 2dm^{-1} \sum_{j=1}^{m} \log \lambda_j + 1, \quad \text{where} \tag{2.3}$$

$$\widehat{G}(d, \theta) = m^{-1} \sum_{j=1}^{m} I_j \exp(p_r(\lambda_j, \theta)) \lambda_j^{2d}.$$

The LPW estimator $(\widehat{d}(r), \widehat{\theta}(r))$ of (d, θ) solves the following minimization problem:

$$\left(\widehat{d}(r), \widehat{\theta}(r)\right) = \underset{d \in [d_1, d_2], \theta \in \Theta}{\arg\min} \; R_r(d, \theta), \tag{2.4}$$

where Θ is a compact and convex set in $\mathbb{R}^r$. Existence and uniqueness of $(\widehat{d}(r), \widehat{\theta}(r))$ is a consequence of strict convexity of $R_r(d, \theta)$ (shown below) and convexity and compactness of the parameter space.

By definition, the estimator of G is

$$\widehat{G}(r) = \widehat{G}\big(\widehat{d}(r), \widehat{\theta}(r)\big). \tag{2.5}$$

3. ASSUMPTIONS

We now introduce the assumptions that are employed 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., $|h^{([s])}(\lambda) - h^{([s])}(0)| \le C|\lambda|^{s-[s]}$ for some constant $C < \infty$ and all λ in a neighborhood of zero.

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* $-1/2 < d_1 < d_2 < 1/2$.

ASSUMPTION 2: $\varphi(\lambda)$ *is smooth of order* s *at* $\lambda = 0$, *where* $s > 2r$ *and* $s \ge 1$.

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:

$$\log \varphi(\lambda) = \log \varphi(0) + \sum_{k=1}^{[s/2]} \frac{b_{2k}}{(2k)!}\lambda^{2k} + O(\lambda^s) \quad \text{as} \quad \lambda \to 0+, \quad \text{where} \tag{3.1}$$

$$b_k = \frac{d^k}{d\lambda^k} \log \varphi(\lambda)\bigg|_{\lambda=0}.$$

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

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

ASSUMPTION 3: (a) *The time series* $\{x_t : t = 1, \ldots, n\}$ *satisfies*

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

where

$$\sum_{j=0}^{\infty} \alpha_j^2 < \infty, \quad E(\varepsilon_t|F_{t-1}) = 0 \quad a.s., \quad E(\varepsilon_t^2|F_{t-1}) = 1 \quad a.s.,$$

$$E(\varepsilon_t^3|F_{t-1}) = \sigma_3 \quad a.s.,$$

$$E(\varepsilon_t^4|F_{t-1}) = \sigma_4 \quad a.s. \quad \textit{for} \quad t = \ldots, -1, 0, 1, \ldots,$$

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. Assumptions 3(a) and (b) are 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 $\beta = 2$. His Assumption A4′ on m is weakest when $\beta = 2$ and in this case it requires that $1/m + m^5(\log^2 m)/n^4 \to 0$. In contrast, if $r = 0$ and our Assumption 2 holds with $s = 2$, then our 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 FOC's 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 log-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 $\gamma_0 \in \Gamma \subset \mathbb{R}^k$, 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 $k \times k$ nonrandom 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* $\eta > 0$, $\lambda_{\min}((B_n^{-1})' \nabla^2 L_n(\gamma_0) B_n^{-1} \geq \eta$ *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*

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

The proofs of Lemma 1 and other results below are given in the Appendix of Proofs.

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 jth diagonal element $[B_n]_{jj}$ defined by

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

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

$$S_n(d,\theta) = \widehat{G}^{-1}(d,\theta)\sum_{j=1}^{m}\left(y_j(d,\theta) - m^{-1}\sum_{k=1}^{m} y_k(d,\theta)\right)X_j \quad \text{and} \tag{4.2}$$

$$H_n(d,\theta) = \widehat{G}^{-2}(d,\theta)\left(\widehat{G}(d,\theta)\sum_{j=1}^{m} y_j(d,\theta)X_jX_j' - m\left(m^{-1}\sum_{j=1}^{m} y_j(d,\theta)X_j\right)\times\left(m^{-1}\sum_{j=1}^{m} y_j(d,\theta)X_j\right)'\right),$$

where

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

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 kth element $\mu_{r,k}$, Γ_r is an $r\times r$ matrix with (i,k)th element $[\Gamma_r]_{i,k}$,

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

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

$$\nu_n(r,s) = m^{\phi+1/2}n^{-\phi}\big(\mathbb{1}(s\geq 2+2r)b_{2+2r}\kappa_r\xi_r^+ + \mathbb{1}(2r<s<2+2r)O(1)\big) \tag{4.6}$$
$$= \mathbb{1}(s\geq 2+2r)m^{5/2+2r}n^{-(2+2r)}b_{2+2r}\kappa_r\xi_r^+ + \mathbb{1}(2r<s<2+2r)O(m^{s+1/2}n^{-s}),$$

and

$$\xi_r^+ = \begin{pmatrix} 2 \\ \xi_r \end{pmatrix},$$

$$\xi_r = (\xi_{r,1}, \ldots, \xi_{r,r})',$$

$$\xi_{r,k} = \frac{2k(3+2r)}{(2r+2k+3)(2k+1)} \quad \text{for} \quad k = 1, \ldots, r, \quad \text{and}$$

$$\kappa_r = -\frac{(2\pi)^{2+2r}(2+2r)}{(3+2r)!(3+2r)}. \tag{4.7}$$

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:

$$D_m(\eta) = \{d \in [d_1, d_2] : (\log^5 m)|d - d_0| < \eta\} \quad \text{for} \quad \eta > 0 \quad \text{and}$$

$$J_n = \sum_{j=1}^{m} \left(X_j - m^{-1} \sum_{k=1}^{m} X_k \right) \left(X_j - m^{-1} \sum_{k=1}^{m} X_k \right)'. \tag{4.8}$$

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

(a) $B_n^{-1} J_n B_n^{-1} \to \Omega_r$,

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

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

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

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

COMMENTS: 1. Part (c) 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 does not hold.

2. The proof of Lemma 2 relies heavily on the proof of Theorem 2 of Robinson (1995a), as well as 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(a) and (b) 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(c) and (d).

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* $(\partial/\partial(d, \theta')')R_r(d, \theta) = 0$ $wp \to 1$ *and*

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

5. CONSISTENCY AND ASYMPTOTIC NORMALITY

We start by showing that the Hessian $H_n(d, \theta)$ is positive definite (pd) for all (d, θ). By (4.2), for any $c \in \mathbb{R}^{1+r}$ with $c \neq 0$,

$$\begin{aligned} (5.1) \qquad & c' H_n(d, \theta) c \cdot \widehat{G}^2(d, \theta)/m \\ & = \widehat{G}(d, \theta) m^{-1} \sum_{j=1}^{m} y_j(d, \theta)(c' X_j)^2 - \left(m^{-1} \sum_{j=1}^{m} y_j(d, \theta) c' X_j \right)^2. \end{aligned}$$

The right-hand side can be written as $a'a \cdot b'b - (a'b)^2 > 0$, where a and b are m-vectors with $a_j = (m^{-1} y_j(d, \theta))^{1/2}$ and $b_j = (m^{-1} y_j(d, \theta))^{1/2} c' X_j$ for $j = 1, \ldots, m$ and the inequality holds by the Cauchy–Schwarz inequality. This establishes the following lemma.

LEMMA 3: *The* (*negative*) *concentrated* LPW *log-likelihood*, $R_r(d, \theta)$, *is strictly convex on* $[d_1, d_2] \times \Theta$.

The LPW log-likelihood $R_r(d, \theta)$ is a continuous function defined on a compact set. Hence, the LPW estimator exists. Strict convexity of $R_r(d, \theta)$ implies that the LPW estimator is unique. Furthermore, strict convexity and twice continuous differentiability of $R_r(d, \theta)$ imply that if a solution $(\widetilde{d}(r), \widetilde{\theta}(r))$ to the FOC's exists, then it minimizes $R_r(d, \theta)$ over the parameter space and, hence, equals $(\widehat{d}(r), \widehat{\theta}(r))$. This can be shown by a two term Taylor expansion. Let $\widetilde{\gamma}(r) = (\widetilde{d}(r), \widetilde{\theta}(r)')'$. Then, for all $\gamma = (d, \theta')' \neq \widetilde{\gamma}(r)$ in the parameter space,

$$\begin{aligned} (5.2) \qquad & m R_r(\gamma) - m R_r(\widetilde{\gamma}(r)) \\ & = S_n(\widetilde{\gamma}(r))'(\gamma - \widetilde{\gamma}(r)) + \frac{1}{2}(\gamma - \widetilde{\gamma}(r))' H_n(\overline{\gamma}(r))(\gamma - \widetilde{\gamma}(r)) \\ & = \frac{1}{2}(\gamma - \widetilde{\gamma}(r))' H_n(\overline{\gamma}(r))(\gamma - \widetilde{\gamma}(r)) > 0, \end{aligned}$$

where $\overline{\gamma}(r)$ lies between γ and $\widetilde{\gamma}(r)$, the second equality holds by the FOC's, and the inequality holds because the Hessian is pd for all $\gamma = (d, \theta')'$ by strict convexity.

In consequence, Theorem 1 and Lemma 3 imply the following consistency and asymptotic normality result for the LPW estimator.

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

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

as $n \to \infty$.

COMMENTS: 1. By the formula for a partitioned inverse,

$$\Omega_r^{-1} = \begin{pmatrix} \frac{c_r}{4} & -\frac{c_r}{2}\mu_r'\Gamma_r^{-1} \\ -\frac{c_r}{2}\Gamma_r^{-1}\mu_r & \Gamma_r^{-1} + c_r\Gamma_r^{-1}\mu_r\mu_r'\Gamma_r^{-1} \end{pmatrix}, \quad \text{where} \tag{5.3}$$

$$c_r = (1 - \mu_r'\Gamma_r^{-1}\mu_r)^{-1} \quad \text{for} \quad r > 0 \text{ and } c_0 = 1.$$

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, $c_1 = 9/4$, $c_2 = 3.52$, $c_3 = 4.79$, and $c_4 = 6.06$.

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

$$\mathbb{1}(s \geq 2 + 2r)\tau_r b_{2+2r} m^{2+2r} n^{-(2+2r)} \\ + \mathbb{1}(2r < s < 2 + 2r)O(m^s/n^s), \quad \text{where} \tag{5.4}$$

$$\tau_r = \frac{\kappa_r c_r}{2}(1 - \mu_r'\Gamma_r^{-1}\xi_r).$$

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

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

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

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. If $s \geq 2 + 2r$, then using Comments 1 and 2, the "asymptotic MSE" of $\widehat{d}(r)$ is

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

Minimization over m of $\text{AMSE}(\widehat{d}(r))$ 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) (where their $E_\beta(H)$ equals our $b_2/2$). The formula for m_{opt} contains only one unknown, b_{2+2r}.

6. 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 (as a function of the sample size n) than is possible with $\widehat{d}(0)$. See Section 7 for an adaptive choice of m and r.

7. Inflation of the asymptotic variance by the factor c_r due to the addition of parameters (see Comment 1) also is found in AG for a bias-reduced log-periodogram regression estimator of d_0. In consequence, the LPW 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 $(c_r/4) \div (\pi^2 c_r/24) \doteq .608$.

8. The asymptotic bias in (5.4) is the same as that found in AG for the bias-reduced log-periodogram estimator of d_0. Hence, the LPW estimator has the same asymptotic bias, but smaller asymptotic variance, than the latter estimator of d_0.

Theorem 2 provides new results for the local Whittle estimator $\widehat{d}(0)$ that is analyzed in Robinson (1995a).

COROLLARY 1: *Under Assumptions* 1–5, $m^{1/2}(\widehat{d}(0) - d_0) - \nu_n(0, s)/4 \to_d N(0, 1/4)$ *as* $n \to \infty$.

COMMENTS: 1. The "asymptotic bias" of $\widehat{d}(0)$ is

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

2. An analogous result to Corollary 1, but for the GPH estimator, is given by Hurvich and Deo (1999, Theorem 2). A comparison of Corollary 1 with Theorem 2 of Hurvich and Deo (1999) 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 $\nu_n(0, s) = o_p(1)$. Henry and Robinson (1996) provide a heuristic expression for the asymptotic bias of the local Whittle estimator in their equation (1.3).[2]

6. OPTIMAL RATE OF CONVERGENCE

In this section, we show that the LPW estimator attains the optimal rate of convergence for estimation of d_0 established in AG for Gaussian processes. The LPW estimator attains this rate whether or not the process is Gaussian. The optimal rate established in AG 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 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 $a = (a_0, a_{00}, a_1, \ldots, a_{[s/2]})'$, $\delta = (\delta_1, \delta_2, \delta_3)'$, 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:

$$\mathcal{F}(s, a, \delta, K) = \left\{ f : f(\lambda) = |\lambda|^{-2d_f} \varphi(\lambda), |d_f| \leq 1/2 - \delta_1, \int_{-\pi}^{\pi} f(\lambda)\, d\lambda \leq K_1, \right. \tag{6.1}$$

$$\text{and } \varphi \text{ is an even function on } [-\pi, \pi] \text{ that satisfies}$$

[2]Henry and Robinson's (1996) expressions in (1.3) for the bias contain two typos. All three of their expressions are missing a minus sign because the preceding expression for $\widehat{H} - H$ is missing a minus sign. Also, their right-hand side expression in (1.3) should have $1/2$ in place of 2 in the numerator. With these corrections, their expressions for the asymptotic bias are equivalent to ours.

$$\text{(i) } a_0 \le \varphi(0) \le a_{00},$$

$$\text{(ii) } \varphi(\lambda) = \varphi(0) + \sum_{k=1}^{[s/2]} \varphi_k \lambda^{2k} + \Delta(\lambda) \text{ for some constants } \varphi_k$$

$$\text{with } |\varphi_k| \le a_k \text{ for } k = 1, \ldots, [s/2] \text{ and some function } \Delta(\lambda)$$

$$\text{with } |\Delta(\lambda)| \le K_2 \lambda^s \text{ for all } 0 \le \lambda \le \delta_2,$$

$$\left.\text{(iii) } |\varphi(\lambda_1) - \varphi(\lambda_2)| \le K_3 |\lambda_1 - \lambda_2| \text{ for all } 0 < \lambda_1 < \lambda_2 \le \delta_3\right\}.$$

If φ is an even function on $[-\pi, \pi]$ that is smooth of order $s \ge 1$ 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 $k = 1, \ldots, [s/2]$ 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 AG.

THEOREM 3: *Let s and the elements of $a = (a_0, a_{00}, a_1, \ldots, a_{[s/2]})'$, $\delta = (\delta_1, \delta_2, \delta_3)'$, and $K = (K_1, K_2, K_3)'$ be any positive real numbers with $s \ge 1$, $a_0 < a_{00}$, $\delta_1 < 1/2$, and $K_1 \ge 2\pi a_{00}$.*

(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*

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

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 the innovations $\{\varepsilon_t : t = \ldots, 0, 1, \ldots\}$ in Assumption* 3 *having distribution that does not depend on $f \in \mathcal{F}(s, a, \delta, K)$ and with the $O(\cdot)$ term in Assumption* 3(c) *holding uniformly over $f \in \mathcal{F}(s, a, \delta, K)$. Let $m = \psi_1 n^{2s/(2s+1)}$ for some constant $\psi_1 \in (0, \infty)$ and let $r \ge 0$ be the largest integer (strictly) less than $s/2$. Then,*

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

COMMENT: 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. AN ADAPTIVE LPW ESTIMATOR

The definition of $\widehat{d}(r)$ depends on r, the degree of the local polynomial. In turn, a suitable choice of bandwidth m depends on r. In this section, we develop a procedure to choose r and m so they adapt to the smoothness of $\varphi(\lambda)$. The basic method comes from Lepskii (1990) and has been used in the context of estimation of the long-memory parameter by GRS, IMS, and HMS. We show that an adaptive LPW estimator achieves, up to a logarithmic factor, the optimal rate of convergence established in the previous section uniformly over values of the smoothness parameter $s \in [s_*, s^*]$, where s_* and s^* are constants that satisfy $1 \le s_* \le s^* < \infty$. Furthermore, the same estimator achieves this result for all $s^* < \infty$, so there is no need to select an upper bound s^*. This contrasts with the procedures considered in Lepskii (1990) and GRS. The adaptive procedure achieves this rate of convergence without assuming Gaussianity of the process, which contrasts with the results of GRS, IMS, and HMS.

Let $s \in [s_*, \infty)$ for $s_* \ge 1$. For a positive constant ψ_1, set

$$m(s) = \psi_1 n^{\frac{2s}{2s+1}} \quad \text{and} \tag{7.1}$$
$$r(s) = w \quad \text{for} \quad s \in (2w, 2w+2] \quad \text{for} \quad w = 0, 1, \ldots.$$

Equivalently, $r(s) = [s/2]$ if $s/2 \notin \mathbb{N}$ and $r(s) = s/2 - 1$ if $s/2 \in \mathbb{N}$.

Denote $\widehat{d}_s = \widehat{d}(r(s))$ when the bandwidth is $m(s)$. Let $h = 1/\log n$ and $\mathcal{S}_h$ be the h-net of the interval $[s_*, \infty)$: $\mathcal{S}_h = \{\tau : \tau = s_* + kh, k = 0, 1, 2, \ldots\}$. Define

$$\widehat{s} = \sup\big\{s \in \mathcal{S}_h : |\widehat{d}_\tau - \widehat{d}_s| \le m^{-1/2}(\tau)\psi_2(c_{r(\tau)}/4)^{1/2}\zeta(n) \tag{7.2}$$
$$\text{for all } \tau \le s, \tau \in \mathcal{S}_h\big\},$$
$$\zeta(n) = (\log n)(\log\log(n))^{1/2},$$

where ψ_2 is a positive constant. Graphically, one can view the bound in the definition of $\widehat{s}$ as a function of τ. Then, $\widehat{s}$ is the largest value of $s \in \mathcal{S}_h$ such that

$|\widehat{d}_\tau - \widehat{d}_s|$ lies below the bound for all $\tau \le s$, $\tau \in \mathcal{S}_h$. Calculation of $\widehat{s}$ is carried out by considering successively larger s values s_*, $s_* + h$, $s_* + 2h, \ldots$ until for some s the deviation $|\widehat{d}_\tau - \widehat{d}_s|$ exceeds the bound for some $\tau \le s$, $\tau \in \mathcal{S}_h$.

The adaptive estimator $\widehat{d}_{\widehat{s}}$ satisfies the following result.

THEOREM 4: *Let the elements of* $a = (a_0, a_{00}, a_1, \ldots, a_{[s/2]})'$, $\delta = (\delta_1, \delta_2, \delta_3)'$, *and* $K = (K_1, K_2, K_3)'$ *be any positive real numbers with* $a_0 < a_{00}$, $\delta_1 < 1/2$, *and* $K_1 \ge 2\pi a_{00}$. *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 the innovations* $\{\varepsilon_t : t = \ldots, 0, 1, \ldots\}$ *in Assumption* 3 *having a distribution that does not depend on* $f \in \mathcal{F}(s, a, \delta, K)$ *and with the* $O(\cdot)$ *term in Assumption* 3(c) *holding uniformly over* $f \in \mathcal{F}(s, a, \delta, K)$. *Let* $s_* \ge 1$. *For all* $s^* \in [s_*, \infty)$,

$$\lim_{C_1 \to \infty} \limsup_{n \to \infty} \sup_{s \in [s_*, s^*]} \sup_{f \in \mathcal{F}(s,a,\delta,K)} P_f\left(n^{\frac{s}{2s+1}} \zeta^{-1}(n) |\widehat{d}_{\widehat{s}} - d_f| \ge C_1\right) = 0.$$

COMMENTS: 1. By Theorem 3, for $f \in \mathcal{F}(s, a, \delta, K)$, the optimal rate of convergence of an estimator of d_f for a given value of s is $n^{-s/(2s+1)}$. Theorem 4 shows that the adaptive LPW estimator, $\widehat{d}_{\widehat{s}}$, achieves this rate up to the logarithmic factor $\zeta(n)$ uniformly over $s \in [s_*, s^*]$ for any $s^* \in [s_*, \infty)$.

2. Theorem 4 does not require $\{x_t\}$ to be a Gaussian process. In fact, it only requires x_t to have finite fourth moments.

3. For a density f that is in $\mathcal{F}(s, a, \delta, K)$ for all $s < \infty$, Theorem 4 shows $\widehat{d}_{\widehat{s}}$ is $n^{1/2-\delta}$-consistent for all $\delta > 0$. That is, $\widehat{d}_{\widehat{s}}$ has a rate of convergence that is arbitrarily close to the parametric rate. For example, this is the rate obtained for ARFIMA(p, d, q) processes.

4. In the simulations reported in Section 8, we take $s_* = 1$, which allows for the smoothness of $\varphi(\lambda)$ to be any $s \ge 1$.

5. We note that the adaptive procedure considered here is not fully data-dependent. The constants ψ_1 and ψ_2 must be specified. In the simulation section below, we take $\psi_1 = .3$ and $\psi_2 = .2$. These choices work well for a variety of different models and parameter values. Note that analogous constants also appear (or are set equal to arbitrary values) in the adaptive procedures of Lepskii (1990), GRS, IMS, and HMS.

6. Suppose $r(s)$ is defined to equal zero, which corresponds to the standard local Whittle estimator, and $\widehat{s}$ is otherwise defined as above. Then, the result of Theorem 4 still holds, but only for the class of spectral densities $\mathcal{F}(s, a, \delta, K)$ for which $a_k = 0$ for $k = 1, 2, \ldots$. This class of spectral densities is analogous to that considered by GRS. As discussed in AG, this condition is quite restrictive. Spectral densities that are smooth of order s at zero only satisfy this condition if all the coefficients of the Taylor expansion of $\varphi(\lambda)$ about $\lambda = 0$ to order $[s]$ are zero.

7. The adaptive estimator of Theorem 4 is not necessarily asymptotically normal. However, at the cost of a slower rate of convergence, an adaptive estimator can be constructed that is asymptotically normal with zero asymptotic

bias by altering the definition of $m(s)$ so that $m(s)$ diverges to infinity at a slower rate than $n^{2s/(2s+1)}$. For example, suppose one defines $\widehat{s}$ as above but with

$$m(s) = \psi_1 n^{4r(s)/(4r(s)+1)}. \tag{7.3}$$

Then, it can be shown that the result of Theorem 4 holds but with $n^{s/(2s+1)}$ replaced by $n^{r(s)/(2r(s)+1)}$. Now, suppose the true spectral density of $\{x_t\}$ is f and Assumptions 3 and 5 hold as in Theorem 4. Let s_f be the supremum of $\{s : f \in \mathcal{F}(s, a, \delta, K)$ for some (a, δ, K), where (a, δ, K) are as in Theorem 4$\}$. Provided $s_f < \infty$ and s_f is not an even integer, Theorems 3 and 4 imply that $r(\widehat{s}) = r(s_f)$ wp$\to$ 1. Thus, $r(\widehat{s})$ and $m(\widehat{s})$ are essentially nonrandom for large n. In consequence, the asymptotic normality result of Theorem 2 applies to the adaptive estimator $\widehat{d}_{\widehat{s}}$ with $r = r(s_f)$ and $\nu_n(r(s_f), s_f) = o(1)$ in the result of Theorem 2, where $o(1)$ holds by (4.6) with $s_f \in (2r(s_f), 2 + 2r(s_f))$, $m = m(s_f)$, and $m(s_f)^{s_f+1/2} n^{-s_f} = o(1)$. Of course, one would expect that a given level of accuracy of approximation by the normal distribution would require a larger sample size when r and m are adaptively selected than otherwise.

8. MONTE CARLO SIMULATIONS

8.1. *Experimental Design*

In this section, we present some simulation results that compare the root mean-square error (RMSE) performance of the adaptive LPW estimator with several adaptive estimators in the literature. Additional simulation results for nonadaptive LPW estimators are given in Andrews and Sun (2001).

We consider three models and several parameter combinations for each model.

The first model we consider for the time series $\{x_t : t \geq 1\}$ is a first-order autoregressive fractionally integrated (ARFIMA$(1, d, 0)$) model with autoregressive (AR) parameter ϕ and long-memory parameter d_0:

$$(1 - \phi L)(1 - L)^{d_0} x_t = u_t, \tag{8.1}$$

where the innovations $\{u_t : t = \ldots, 0, 1, \ldots\}$ are iid random variables and L denotes the lag operator. We consider three distributions for u_t: standard normal, t_5, and χ^2_2. The t_5 and χ^2_2 distributions are considered because they exhibit thick tails and asymmetry, respectively. All of the estimators of d_0 that we consider are invariant with respect to the mean and variance of the time series. In consequence, the choice of location and scale of the innovations is irrelevant. Note that the spectral density of an ARFIMA$(1, d, 0)$ process is continuous and infinitely differentiable on $(0, \pi]$.

The second model we consider is a stationary ARFIMA$(1, d, 0)$-*like* model that has a discontinuity in its spectral density at frequency $\lambda = \lambda_0$. We call this model a DARFIMA$(1, d, 0)$ model. Its spectral density is that of an

ARFIMA$(1, d, 0)$ process for $\lambda \in (0, \lambda_0]$, but is zero for $\lambda \in (\lambda_0, \pi]$. A DARFIMA$(1, d, 0)$ process $\{x_t : t \geq 1\}$ is defined as in (8.1), but with innovations $\{u_t : t = \ldots, 0, 1, \ldots\}$ that are an iid Gaussian process filtered by a low pass filter. Specifically,

$$u_t = \sum_{j=-\infty}^{\infty} c_j \varepsilon_{t-j} \quad \text{for} \quad t = \ldots, 0, 1, \ldots, \quad \text{where} \tag{8.2}$$

$$c_j = \begin{cases} \dfrac{\lambda_0}{\pi} & \text{for} \quad j = 0, \\ \dfrac{\sin(\lambda_0 j)}{j\pi} & \text{for} \quad j \neq 0, \end{cases}$$

and $\{\varepsilon_t : t = \ldots, 0, 1, \ldots\}$ are iid random variables with standard normal, t_5, or χ_2^2 distribution. The spectral density $f_u(\lambda)$ of $\{u_t : t \geq 1\}$ equals $\sigma_u^2/(2\pi)$ for $0 < \lambda \leq \lambda_0$ and equals 0 for $\lambda_0 < \lambda \leq \pi$, where σ_u^2 denotes the variance of u_t; e.g., see Brillinger (1975, equation (3.3.25), p. 58). The spectral density of $\{x_t : t \geq 1\}$ is $f_u(\lambda)$ times the spectral density of the ARFIMA$(1, d, 0)$ process that has the same AR parameter. Thus, the spectral density of a DARFIMA process is a truncated discontinuous version of that of the corresponding ARFIMA process.

The third model we consider is a model that we call a *long-memory components* (LMC) model. It is designed to have a finite degree of smoothness s_0 at frequency zero in the short-run part, $\varphi(\lambda)$, of its spectral density. The process $\{x_t : t \geq 1\}$ is defined by

$$(1 - L)^{d_0} x_t = u_t + k(1 - L)^{s_0/2} v_t, \tag{8.3}$$

where $\{u_t : t \geq 1\}$ and $\{v_t : t \geq 1\}$ are independent iid processes both with normal, t_5, or χ_2^2 distribution. The spectral density function of $\{u_t + k(1-L)^{s_0/2} v_t : t \geq 1\}$ is

$$\varphi_{s_0,k}(\lambda) = \frac{\sigma^2}{2\pi} + \frac{k^2\sigma^2}{2\pi} |1 - e^{i\lambda}|^{s_0}, \tag{8.4}$$

where σ^2 denotes the variance of u_t and v_t. Because $|1 - e^{i\lambda}| \sim \lambda$ as $\lambda \to 0$, the smoothness of $\varphi_{s_0,k}(\lambda)$ at $\lambda = 0$ is s_0 for non-integer s_0.

For the ARFIMA and DARFIMA models, we consider seven values of ϕ, viz., 0, .3, .6, .9, $-.3$, $-.6$, and $-.9$. For the DARFIMA model, we take λ_0 to equal $\pi/2$. For the LMC model, we take $s_0 = 1.5$ and consider five values of k, viz., 1/3, 1/2, 1, 2, and 3. For all three models, we consider three values of d_0, viz., $d_0 = -.4, 0$, and .4. For each model, we consider two sample sizes $n = 512$ and $n = 4{,}096$. In all cases, 1,000 simulation repetitions are used. This produces simulation standard errors that are roughly 3% of the magnitudes of the reported RMSE's.

The estimators that we consider include the adaptive LPW (ALPW) estimator defined in the previous section, the adaptive log-periodogram regression estimator of GRS, the adaptive FEXP estimator of IMS, and the FEXP estimator of Hurvich (2001) (H) with the number of terms in the expansion chosen by his local C_L method. Each of these estimators requires the specification of certain constants in the adaptive or local C_L procedure. In addition, the estimator analyzed by GRS requires trimming of frequencies near zero and tapering of the periodogram, and the estimator analyzed by IMS requires tapering of the periodogram and allows for pooling of the periodogram.

The constants in the adaptive procedures are tuned to the Gaussian ARFIMA model with $\phi = .6$ with $n = 512$. That is, they are determined by simulation to be the values (from a grid) that yield the smallest RMSE for the Gaussian ARFIMA model with $\phi = .6$ and $n = 512$. These values are then used for all of the processes considered in the experiment. For the ALPW procedure, the grid for the constant ψ_1 is $\{.1, .2, \ldots, .5\}$ and the grid for ψ_2 is $\{.05, .10, \ldots, .70\}$. For the GRS procedure, their constant β^* is taken to be two (as suggested on their p. 192). In addition, we introduce two constants ψ_1 and ψ_2 that are analogous to the constants that appear in the adaptive procedure for ALPW.[3] The grid for ψ_1 is $\{.1, .2, \ldots, 1.0\}$ and the grid for ψ_2 is $\{.05, .10, \ldots, .70\}$. The constants ψ_1 and ψ_2 are introduced in order to give the GRS adaptive procedure a degree of flexibility that is comparable to that of the ALPW procedure. For the IMS procedure, their constant κ is analogous to the constant ψ_2 of the ALPW estimator and is chosen from the same grid as ψ_2 and the pooling size (m in IMS's notation and denoted *pool* below) is determined simultaneously with the constant κ from the grid $\{1, 2, \ldots, 6\}$.[4]

We analyze several versions of the adaptive estimators. The first version is a version that is closest to being covered by the theoretical results in the literature. We refer to these as being the "theoretically-justified" estimators and they are denoted ALPW1, GRS1, and IMS1. (The ALPW1 estimator *is* covered by the results of this paper. The GRS1 estimator uses constants ψ_1 and ψ_2 that are not covered by the results of GRS and the IMS1 estimator uses a constant κ and an upper bound on the number p of Fourier terms that are not covered by the results of IMS; see footnote 4.) The GRS1 estimator uses the cosine-bell

[3]The constants ψ_1 and ψ_2 appear as multiplicative constants on the right-hand side of GRS's formulae for the bandwidth $m(\gamma)$ in their (3.6) and for $d(\beta')$ in their (3.8), respectively.

[4]IMS require that their constant $\kappa > 6$. Such a choice provides poor finite sample performance for all of the cases that we considered. In consequence, we did not impose this bound and we selected κ from the same grid $\{.05, .10, \ldots, .70\}$ as for the analogous constant ψ_2 for the ALPW and GRS procedures. The selected κ value was not at the upper end of the grid.

In addition, IMS require that the number of terms, p, in their Fourier series expansion be less than a bound (their $K\varepsilon_n$) that is very restrictive. For example, with $pool = 1$, the bound allows for at most zero terms when $n \leq 18{,}000$ and at most one term when $n \leq 58{,}000$. The bound is more restrictive when $pool > 1$. We do not impose this bound because it would eviscerate the semiparametric nature of the estimator. Instead, we required that $p \leq 20$.

taper with two out of every three frequencies dropped (as in GRS) and three frequencies near the origin are trimmed ($trim = 3$) when $n = 512$ and six are trimmed ($trim = 6$) when $n = 4{,}096$. The IMS1 estimator uses the Hurvich taper (of order one) with one frequency dropped between each pool group of frequencies, as in Section 2.2 of HMS. (Note that for the case where no differencing is carried out to eliminate potential trends, the IMS and HMS adaptive estimators are essentially the same except that HMS uses a scheme that deletes fewer frequencies, which we employ here.)

The second and third versions of the adaptive estimators that we consider do not have known theoretical properties. The ALPW2 estimator is the same as the ALPW1 estimator except that a bound is placed on the degree of the polynomial. Specifically, $\widehat{s}$ is defined as in (7.2), the bandwidth is taken to be $m(\widehat{s})$, and the degree of the polynomial is taken to be $\min\{r(\widehat{s}), 2\}$, when $n = 512$ and $\min\{r(\widehat{s}), 4\}$, when $n = 4{,}096$.

The estimator GRS2 differs from GRS1 in that it does not trim any frequencies near zero. The estimator GRS3 differs from GRS1 in that it does not trim frequencies near zero, use a taper, or drop two out of every three frequencies. The estimator IMS2 differs from IMS1 in that it does not use a taper or drop any frequencies.

The constants determined by simulation are: ALPW1: $(\psi_1, \psi_2) = (.3, .2)$; ALPW2: $(\psi_1, \psi_2) = (.3, .5)$; GRS1: $(\psi_1, \psi_2) = (.6, .5)$; GRS2: $(\psi_1, \psi_2) = (.3, .6)$; GRS3: $(\psi_1, \psi_2) = (.2, .25)$; IMS1: $(pool, \kappa) = (2, .65)$; and IMS2: $(pool, \kappa) = (2, .45)$.

The H procedure requires the specification of a constant α. We consider the two values $\alpha = .5$ and $\alpha = .8$ that are considered in the Hurvich paper. The corresponding estimators are denoted H1 and H2. (The value $\alpha = .8$ turns out to minimize RMSE of the FEXP estimator for the ARFIMA process with $\phi = .6$ and $n = 512$ over α values in $\{.1, .2, \ldots, .8\}$.[5]) The theoretical properties of the H procedure, such as its rate of convergence, are not given in Hurvich (2001). For this reason, we do not put the H1 and H2 estimators in with the first group of "theoretically-justified" estimators.

The final estimator that we consider is the parametric Whittle quasi-maximum likelihood (QML) estimator for a Gaussian ARFIMA$(1, d, 0)$ model. This estimator is misspecified when the model under consideration is the DARFIMA$(1, d, 0)$ model or the LMC model. This estimator is included in the simulations for comparative purposes.

8.2. *Monte Carlo Results*

Tables I–III give the results for the three models. Each table has a separate panel of results for $n = 512$ and $n = 4{,}096$. The numbers reported in the tables are the RMSE's of the estimators. The first three rows of each panel give

[5]Note that $\alpha = .9$ is not included in the grid because it does not yield a *local* C_L criterion given that $m = n^{.9} \geq n/2$.

TABLE I

RMSE FOR ARFIMA$(1, d, 0)$ PROCESSES WITH AR PARAMETER ϕ

	$d=0$ Normal ϕ				$d=0$ t_5 ϕ		$d=.4$ Normal ϕ	
Estimator	0	.3	.6	.9	.6	.9	.6	.9
				(a) $n=512$				
ALPW1	.145	.142	.145	.423	.140	.420	.151	.425
GRS1 (*taper*, *trim* $=3$)	.160	.197	.397	.855	.393	.864	.394	.857
IMS1 (*pool* $=2$, *taper*)	.234	.252	.291	.448	.311	.460	.379	.579
ALPW2 ($r \leq 2$)	.098	.098	.118	.551	.121	.552	.126	.551
GRS2 (*taper*, *no trim*)	.164	.172	.244	.662	.254	.657	.253	.668
GRS3 (*no taper*, *no trim*)	.131	.133	.159	.502	.157	.501	.166	.499
IMS2 (*pool* $=2$, *no taper*)	.203	.209	.217	.315	.215	.320	.222	.301
H1 ($\alpha=.5$)	.288	.309	.321	.438	.308	.420	.310	.436
H2 ($\alpha=.8$)	.182	.206	.233	.459	.235	.457	.241	.480
Parametric Whittle QML	.066	.128	.134	.112	.141	.107	.140	.156
				(b) $n=4096$				
ALPW1	.061	.061	.060	.207	.058	.201	.062	.213
GRS1 (*taper*, *trim* $=6$)	.056	.081	.169	.586	.168	.581	.169	.587
IMS1 (*pool* $=2$, *taper*)	.045	.133	.122	.218	.120	.220	.142	.273
ALPW2 ($r \leq 4$)	.041	.041	.045	.336	.043	.333	.048	.340
GRS2 (*taper*, *no trim*)	.063	.066	.108	.396	.109	.388	.114	.401
GRS3 (*no taper*, *no trim*)	.052	.052	.065	.222	.064	.219	.070	.228
IMS2 (*pool* $=2$, *no taper*)	.046	.076	.073	.155	.073	.154	.071	.145
H1 ($\alpha=.5$)	.067	.076	.083	.125	.082	.123	.084	.132
H2 ($\alpha=.8$)	.052	.062	.073	.147	.075	.144	.077	.156
Parametric Whittle QML	.013	.028	.046	.025	.045	.025	.045	.032

results for the "theoretically-justified" adaptive estimators. The next four rows give results for the adaptive estimators that do not have theoretical justification (currently). The last three rows give results for the H1, H2, and parametric Whittle QML estimators.

For brevity, the tables do not provide results for all combinations of parameters considered. For example, results for $\phi = -.3, -.6, -.9$ are not given because they are quite close to those for $\phi = 0, .3$. We only give selected results for $d = .4$ and we give no results for $d = -.4$, because the value of d has only a small effect on RMSE (except for the IMS1 estimator, which exhibits some sensitivity to d). Similarly, we only give selected results for t_5 innovations

TABLE II

RMSE FOR DARFIMA$(1, d, 0)$ PROCESSES WITH $\lambda_0 = \pi/2$ AND AR PARAMETER ϕ

	$d=0$ Normal ϕ				$d=0$ t_5 ϕ		$d=.4$ Normal ϕ	
Estimator	0	.3	.6	.9	.6	.9	.6	.9
				(a) $n=512$				
ALPW1	.143	.142	.146	.423	.149	.430	.146	.425
GRS1 (*taper*, *trim* = 3)	.160	.200	.390	.857	.395	.868	.394	.859
IMS1 (*pool* = 2, *taper*)	.448	.448	.445	.464	.468	.490	.520	.578
ALPW2 ($r \leq 2$)	.181	.185	.244	.676	.248	.676	.244	.637
GRS2 (*taper*, *no trim*)	.176	.181	.243	.656	.254	.657	.253	.668
GRS3 (*no taper*, *no trim*)	.131	.142	.159	.503	.157	.501	.165	.499
IMS2 (*pool* = 2, *no taper*)	.268	.266	.262	.303	.252	.311	.246	.299
H1 ($\alpha = .5$)	.356	.348	.317	.404	.326	.401	.309	.427
H2 ($\alpha = .8$)	.396	.393	.384	.389	.403	.409	.363	.429
Parametric Whittle QML	.866	.616	.949	1.202	.951	1.209	.852	.855
				(b) $n=4096$				
ALPW1	.060	.060	.059	.207	.055	.204	.061	.213
GRS1 (*taper*, *trim* = 6)	.056	.081	.169	.586	.168	.581	.169	.587
IMS1 (*pool* = 2, *taper*)	.146	.146	.146	.127	.148	.130	.121	.120
ALPW2 ($r \leq 4$)	.096	.096	.091	.334	.089	.327	.091	.303
GRS2 (*taper*, *no trim*)	.063	.066	.108	.396	.109	.388	.114	.401
GRS3 (*no taper*, *no trim*)	.051	.052	.066	.222	.064	.219	.070	.229
IMS2 (*pool* = 2, *no taper*)	.096	.096	.096	.106	.093	.103	.100	.125
H1 ($\alpha = .5$)	.126	.125	.123	.117	.124	.119	.110	.127
H2 ($\alpha = .8$)	.120	.120	.119	.113	.122	.119	.119	.139
Parametric Whittle QML	.346	.579	.930	1.256	.933	1.257	.905	1.068

and we give no results for χ^2_2 innovations, because the innovation distribution turns out to have a small effect on RMSE.

We start by noting two broad features of the results presented in the tables. First, the results for $d = 0$ and t_5 innovations are quite similar to those for $d = 0$ and normal innovations. Second, with one exception, the results for $d = .4$ and normal innovations are quite similar to those for $d = 0$ and normal innovations. The exception is: the IMS1 estimator is noticeably worse with $d = .4$ than with $d = 0$ for all ARFIMA models and for the DARFIMA models with $n = 512$.

Now, we compare the estimators. Among the three theoretically-justified estimators, the ALPW1 estimator performs the best across all three models and both sample sizes. Its performance in the ARFIMA and DARFIMA mod-

TABLE III

RMSE FOR LMC MODEL WITH SMOOTHNESS INDEX $s_0 = 1.5$ AND WEIGHT k

	$d=0$ Normal k					$d=0$ t_5 k		$d=.4$ Normal k	
Estimator	1/3	1/2	1	2	3	1/2	2	1/2	2
				(a) $n=512$					
ALPW1	.139	.139	.139	.168	.216	.145	.171	.138	.164
GRS1 (*taper*, *trim* = 3)	.172	.190	.230	.384	.486	.172	.384	.177	.375
IMS1 (*pool* = 2, *taper*)	.247	.249	.264	.298	.336	.262	.297	.307	.323
ALPW1 ($r \leq 2$)	.096	.097	.106	.167	.245	.098	.167	.098	.164
GRS2 (*taper*, *no trim*)	.165	.173	.175	.262	.337	.182	.269	.181	.265
GRS3 (*no taper*, *no trim*)	.131	.131	.136	.181	.245	.121	.184	.128	.176
IMS2 (*pool* = 2, *no taper*)	.207	.207	.209	.209	.229	.194	.215	.215	.233
H1 ($\alpha = .5$)	.264	.265	.276	.303	.333	.259	.307	.271	.311
H2 ($\alpha = .8$)	.178	.182	.199	.241	.298	.169	.229	.182	.245
Parametric Whittle QML	.064	.069	.129	.273	.368	.064	.269	.069	.274
				(b) $n=4096$					
ALPW1	.061	.060	.060	.065	.100	.060	.068	.061	.058
GRS1 (*taper*, *trim* = 6)	.061	.066	.108	.213	.305	.066	.210	.066	.212
IMS1 (*pool* = 2, *taper*)	.056	.075	.123	.136	.166	.075	.138	.070	.105
ALPW2 ($r \leq 4$)	.043	.043	.045	.067	.113	.041	.070	.041	.067
GRS2 (*taper*, *no trim*)	.068	.068	.079	.137	.201	.067	.137	.066	.135
GRS3 (*no taper*, *no trim*)	.051	.052	.056	.085	.122	.053	.087	.049	.080
IMS2 (*pool* = 2, *no taper*)	.050	.058	.070	.085	.085	.058	.086	.062	.089
H1 ($\alpha = .5$)	.072	.073	.078	.089	.100	.078	.089	.079	.091
H2 ($\alpha = .8$)	.056	.058	.065	.087	.107	.060	.089	.061	.087
Parametric Whittle QML	.023	.033	.104	.236	.320	.032	.235	.031	.232

els is almost the same, as desired. This is due to its narrow-band character and the adaptive nature of the bandwidth selection method. In these models the performance of ALPW1 does not vary much with ϕ except for $\phi = .9$. When $\phi = .9$, its performance deteriorates because of the difficulty of distinguishing an AR root close to unity from the long memory parameter d_0. The same deterioration occurs for most of the other estimators. In the LMC model, the performance of the ALPW1 estimator is best when k is small and worst when k is large. This is to be expected and is true of all of the other estimators as well. Not surprisingly, the performance of the ALPW1 and all other estimators improves substantially as n increases from 512 to 4,096.

The GRS1 estimator performs poorly, in an absolute sense and relative to ALPW1, when $\phi = .6$ or .9 and when $k = 1, 2$, or 3. In contrast, the IMS1 estimator performs poorly, in an absolute sense and relative to ALPW1, for $\phi \leq .6$ in the ARFIMA models; for most values of ϕ in DARFIMA models; and for many values of k in LMC models. The performance of the IMS1 estimator deteriorates for DARFIMA models compared to ARFIMA models. This reflects its broad-band character, which is not designed to be robust against discontinuous spectral densities.

Next, we consider the non-theoretically-justified adaptive estimators ALPW2, GRS2, GRS3, and IMS2. The ALPW2 estimator outperforms the ALPW1 estimator and all other semiparametric estimators for both ARFIMA and LMC processes except when $\phi = .9$ or $k = 3$. However, its performance deteriorates for DARFIMA processes. It is outperformed by ALPW1 for most cases with DARFIMA processes. In consequence, it is not possible to order the overall relative performance of ALPW1 and ALPW2.

The GRS2 and GRS3 estimators perform noticeably better than GRS1, especially when ϕ or k is large. Hence, trimming is found to have a negative impact. The GRS3 estimator outperforms the GRS2 estimator across all cases and all models considered. Hence, tapering also is found to have a negative impact. In an overall sense, the GRS3 estimator performs quite well relative to other semiparametric estimators. Compared to the ALPW1 estimator, it does better for small values of ϕ and k, but worse for large values.

The IMS2 estimator outperforms the IMS1 estimator in all cases but one. In many cases, the difference is substantial. Hence, again we find the effect of tapering to be negative. Compared with the other semiparametric estimators, IMS2 performs very well (in fact, the best) in the cases where the short-run serial correlation is highest, i.e., $\phi = .9$ or $k = 3$. But, in most other cases, it is outperformed by the ALPW1, ALPW2, and GRS3 estimators. It appears that the relative performance of the IMS2 estimator compared to the narrow-band ALPW and GRS estimators improves as the sample size increases from 512 to 4,096.

The H1 and H2 estimators perform well when the sample size is 4,096 and either $\phi = .9$ or $k = 3$. In other cases, they do not perform well relative to the ALPW1, ALPW2, or GRS3 estimators. In particular, they perform poorly for DARFIMA processes except when $\phi = .9$. The relative strengths of the H1 and H2 estimators are similar to those of the IMS1 and IMS2 estimators. This is not surprising, because all of these estimators are broadband FEXP estimators. The H2 estimator outperforms the H1 estimator for ARFIMA and LMC processes, but the opposite is true for DARFIMA processes with $n = 512$. In an overall sense, H2 outperforms H1.

The parametric Whittle QML estimator performs as expected. For ARFIMA processes, it outperforms all semiparametric estimators by a substantial margin especially when $\phi = .9$ or when $n = 4{,}096$. For DARFIMA processes, for which it is misspecified, it performs very poorly. It is substantially outperformed by all

semiparametric estimators. For LMC processes, for which it is misspecified, it outperforms the semiparametric estimators for small values of k, which are close to ARFIMA processes. But, it is outperformed for larger values of k.

To conclude, among the three theoretically-justified estimators, the ALPW1 estimator is clearly the best. Trimming hurts the performance of the GRS1 estimator. Tapering hurts the performance of the GRS1 and IMS1 estimators. Of all the semiparametric estimators, the three best ones in an overall sense are the ALPW1, GRS3, and IMS2 estimators. The narrow-band estimators, ALPW1 and GRS3, perform well over a broad range of parameter values, but are outperformed by the broad-band estimator, IMS2, when the serial correlation in the short-run part of the spectrum is quite large. The broadband estimator IMS2 performs relatively well when the sample size is large and the amount of serial correlation is high. The parametric Whittle QML estimator performs very well when the model is correctly specified, moderately well when the amount of misspecification is modest, and poorly when the amount of misspecification is large.

Cowles Foundation for Research in Economics, Yale University, 30 Hillhouse Ave., Box 208281, New Haven, CT 06520-8281, U.S.A.; donald.andrews@yale.edu

and

Dept. of Economics, 0508, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0508, U.S.A.; yisun@weber.usc.edu.

Manuscript received October, 2001; final revision received June, 2003.

APPENDIX: PROOFS

PROOF OF LEMMA 1: Let $\Gamma_{n0} = \{\gamma \in \Gamma : \|B_n(\gamma - \gamma_0)\| \le K_n,\ \|\gamma - \gamma_0\| < \delta\}$ 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 $\gamma \in \Gamma_{n0}$,

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

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

$$
\begin{aligned}
(\text{A.2}) \qquad |\rho_n(\gamma)| &\le \sup_{\overline{\gamma} \in \Gamma_{n0}} \big|(\gamma - \gamma_0)'\big(\nabla^2 L_n(\overline{\gamma}) - \nabla^2 L_n(\gamma_0)\big)(\gamma - \gamma_0)\big| \\
&\le \|B_n(\gamma - \gamma_0)\|^2 \sup_{\overline{\gamma} \in \Gamma_{n0}} \big\|(B_n^{-1})'\big(\nabla^2 L_n(\overline{\gamma}) - \nabla^2 L_n(\gamma_0)\big)B_n^{-1}\big\| \\
&= \|B_n(\gamma - \gamma_0)\|^2 o_p(1).
\end{aligned}
$$

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$\to 1$. In consequence, by (A.1) and (A.2),

$$\text{(A.3)} \qquad L_n(\gamma_n^*) - L_n(\gamma_0) = -\frac{1}{2} Y_n'(B_n^{-1})' \nabla L_n(\gamma_0) + \rho_n(\gamma_n^*) \quad \text{and}$$

$$\rho_n(\gamma_n^*) = o_p(1).$$

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$\to 1$. We have $\Gamma_n(\varepsilon) \subset \Gamma_{n0}$ wp$\to 1$, and so, $\sup_{\gamma \in \Gamma_n(\varepsilon)} |\rho_n(\gamma)| = o_p(1)$ by (A.2). Let $\partial\Gamma_n(\varepsilon)$ denote the boundary of $\Gamma_n(\varepsilon)$. Combining (A.1)–(A.3), for $\gamma \in \partial\Gamma_n(\varepsilon)$,

$$\text{(A.4)} \qquad L_n(\gamma) - L_n(\gamma_n^*) = \frac{1}{2} \mu_n' (B_n^{-1})' \nabla^2 L_n(\gamma_0) B_n^{-1} \mu_n + o_p(1)$$

for some k-vector μ_n with $\|\mu_n\| = \varepsilon > 0$. The right-hand side is bounded away from zero wp$\to 1$ 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$\to 1$. This point satisfies the first-order conditions $\nabla L_n(\widetilde{\gamma}_n(\varepsilon)) = 0$ wp$\to 1$.

In consequence, for all $J \geq 1$, $P(\nabla L_n(\widetilde{\gamma}_n(1/j)) = 0 \ \forall j = 1, \ldots, J) \to 1$ as $n \to \infty$. Thus, there is a sequence $\{J_n : n \geq 1\}$ such that $J_n \uparrow \infty$ and $P(\nabla L_n(\widetilde{\gamma}_n(1/j)) = 0 \ \forall j = 1, \ldots, J_n) \to 1$ as $n \to \infty$. 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 $n = 2, 3, \ldots$. Define $\widetilde{\gamma}_n = \widetilde{\gamma}_n(1/J_n)$ for $n \geq 1$. Then, $P(\nabla L_n(\widetilde{\gamma}_n) = 0) \geq 1 - 1/J_{n-1} \to 1$ as $n \to \infty$. In addition, $\widetilde{\gamma}_n \in \Gamma_n(1/J_n)$ for all $n \geq 1$. Hence, $B_n(\widetilde{\gamma}_n - \gamma_0) = -Y_n + o_p(1) = O_p(1)$. *Q.E.D.*

PROOF OF LEMMA 2(a): Part (a) holds by approximating sums by integrals. See Lemma 2(a), (h), and (i) in AG for details (noting that $X_j = -2 \log \lambda_j$ in AG). *Q.E.D.*

PROOF OF LEMMA 2(b): The normalized Hessian can be written as

$$\text{(A.5)} \qquad B_n^{-1} H_n(d, \theta) B_n^{-1}$$

$$= \widehat{G}^{-2}(d, \theta) \left(\widehat{G}(d, \theta) m^{-1} \sum_{j=1}^{m} y_j(d, \theta) \widetilde{X}_j \widetilde{X}_j' - \left(m^{-1} \sum_{j=1}^{m} y_j(d, \theta) \widetilde{X}_j \right) \right.$$

$$\left. \times \left(m^{-1} \sum_{j=1}^{m} y_j(d, \theta) \widetilde{X}_j \right)' \right), \quad \text{where}$$

$$\widetilde{X}_j = \left(2 \log j, (j/m)^2, \ldots, (j/m)^{2r}\right)'.$$

Let

$$\text{(A.6)} \qquad \widehat{G}_{a,b}(d, \theta) = m^{-1} \sum_{j=1}^{m} I_j \exp(p_r(\lambda_j, \theta)) \lambda_j^{2d} (2 \log j)^a (j/m)^{2b}$$

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

$$\text{(A.7)} \qquad \widehat{G}_{0,0}^{-2} (\widehat{G}_{0,0} \widehat{G}_{2,0} - \widehat{G}_{1,0}^2),$$

$$\widehat{G}_{0,0}^{-2} (\widehat{G}_{0,0} \widehat{G}_{1,k-1} - \widehat{G}_{1,0} \widehat{G}_{0,k-1}), \quad \text{and}$$

$$\widehat{G}_{0,0}^{-2} (\widehat{G}_{0,0} \widehat{G}_{0,k+\ell-2} - \widehat{G}_{0,k-1} \widehat{G}_{0,\ell-1}),$$

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

Let $J_{a,b}$ be defined 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,

$$J_{a,b} = G_0 m^{-1} \sum_{j=1}^{m} (2\log j)^a (j/m)^{2b} \tag{A.8}$$

for $a = 0, 1, 2$ and $b = 0, \ldots, r$. The elements of $B_n^{-1} J_n B_n^{-1}$ are given by the formulae in (A.7) with $\widehat{G}_{a,b}(d_0, \theta_0)$ replaced by $J_{a,b}$. Note that $J_{a,b} = O(\log^a m)$ and $J_{0,0} = G_0 > 0$. Hence, to prove Lemma 2(b), it suffices to show that

$$\Delta_{a,b} := |\widehat{G}_{a,b}(d_0, \theta_0)/G_0 - J_{a,b}/G_0| = o_p(\log^{-2} m) \tag{A.9}$$

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

Let

$$g_j = \lambda_j^{-2d_0} G_0 \exp(-p_r(\lambda_j, \theta_0)). \tag{A.10}$$

By summation by parts, we have

$$\begin{aligned}
\Delta_{a,b} &= \left| m^{-1} \sum_{j=1}^{m} \left(\frac{I_j}{g_j} - 1 \right) (2\log j)^a \left(\frac{j}{m} \right)^{2b} \right| \\
&\le \left| m^{-1} \sum_{k=1}^{m-1} \left[(2\log k)^a \left(\frac{k}{m} \right)^{2b} - (2\log(k+1))^a \left(\frac{k+1}{m} \right)^{2b} \right] \sum_{j=1}^{k} \left(\frac{I_j}{g_j} - 1 \right) \right| \\
&\quad + \left| (2\log m)^a m^{-1} \sum_{j=1}^{m} \left(\frac{I_j}{g_j} - 1 \right) \right| \\
&:= \vartheta_{1,a,m} + \vartheta_{2,a,m}.
\end{aligned} \tag{A.11}$$

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

$$\begin{aligned}
\vartheta_{1,a,m} &\le m^{-1} \sum_{k=1}^{m-1} \left(\left| (2\log k)^a \left(\frac{k}{m} \right)^{2b} - (2\log k)^a \left(\frac{k+1}{m} \right)^{2b} \right| \right. \\
&\quad \left. + \left| (2\log k)^a \left(\frac{k+1}{m} \right)^{2b} - (2\log(k+1))^a \left(\frac{k+1}{m} \right)^{2b} \right| \right) \left| \sum_{j=1}^{k} \left(\frac{I_j}{g_j} - 1 \right) \right| \\
&\le 2^a m^{-1} \sum_{k=1}^{m-1} \left((\log k)^a 2b \left(\frac{k+1}{m} \right)^{2b-1} m^{-1} + a(\log(k+1))^{a-1} k^{-1} \left(\frac{k+1}{m} \right)^{2b} \right) \\
&\quad \times \left| \sum_{j=1}^{k} \left(\frac{I_j}{g_j} - 1 \right) \right| \\
&\le 2^a (\log m)^a (2b + a) m^{-1} \sum_{k=1}^{m-1} k^{-1} \left| \sum_{j=1}^{k} \left(\frac{I_j}{g_j} - 1 \right) \right|.
\end{aligned} \tag{A.12}$$

By altering the statement and proof of (4.8) of Robinson (1995a), as described in Andrews and Sun (2001), and using (4.9) of Robinson (1995a) without change, or by (A.58) below, we obtain:

$$\text{(A.13)(i)} \qquad \sum_{j=1}^{k}\left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j}\right) = O_p(k^{1/3}\log^{2/3} k + k^{\phi+1}n^{-\phi} + k^{1/2}n^{-1/4}) \quad \text{and}$$

$$\text{(A.13)(ii)} \qquad \sum_{j=1}^{k}(2\pi I_{\varepsilon j} - 1) = O_p(k^{1/2}), \quad \text{where}$$

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

as $n \to \infty$ uniformly over $k = 1, \ldots, m$. Combining (A.11)–(A.13), $\vartheta_{1,a,m}$ and $\vartheta_{2,a,m}$ 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.[6] *Q.E.D.*

PROOF OF LEMMA 2(c): By (A.9) and $J_{a,b} = O(\log^a m)$, we obtain $\widehat{G}_{a,b}(d_0, \theta_0) = O_p(\log^a m)$ for $a = 0, 1, 2$ and $b = 0, \ldots, r$ and $\widehat{G}_{0,0}(d_0, \theta_0) = G_0 + o_p(\log^{-2} m)$, where $G_0 > 0$. These results and (A.7) imply that it suffices to show that

$$\text{(A.14)} \qquad \sup_{\theta\in\Theta}|\widehat{G}_{a,b}(d_0, \theta) - \widehat{G}_{a,b}(d_0, \theta_0)| = o_p(\log^{-2} m)$$

for all $a = 0, 1, 2$ and $b = 0, \ldots, r$. The left-hand side of (A.14) equals

$$\begin{aligned}
\text{(A.15)} \qquad & \sup_{\theta\in\Theta}\left|m^{-1}\sum_{j=1}^{m} I_j[\exp(p_r(\lambda_j, \theta)) - \exp(p_r(\lambda_j, \theta_0))]\lambda_j^{2d_0}(2\log j)^a(j/m)^{2b}\right| \\
& \le \sup_{\theta\in\Theta, k=1,\ldots,m} |\exp\{p_r(\lambda_k, \theta) - p_r(\lambda_k, \theta_0)\} - 1| \\
& \qquad \times m^{-1}\sum_{j=1}^{m} I_j \exp(p_r(\lambda_j, \theta_0))\lambda_j^{2d_0}(2\log j)^a \\
& = O(\lambda_m^2)\widehat{G}_{a,0}(d_0, \theta_0), \\
& = O_p\big((m/n)^2(\log^a m)\big) \\
& = o_p(\log^{-2} m),
\end{aligned}$$

where the first equality holds by a mean-value expansion using the compactness of Θ, the second equality holds by (A.9) and $J_{a,b} = O(\log^a m)$, and the third equality holds by Assumption 4. *Q.E.D.*

PROOF OF LEMMA 2(d): We have (i) $\widehat{G}_{a,b}(d_0, \theta) = J_{a,b} + o_p(\log^{-2} m)$ by (A.9) and (A.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 $O(\log^2 m)$ cancels with an identical term in $J_{1,0}^2$,

[6]Note that the summing of terms of the sort $O(k^a)$ over $k = 1, \ldots, m$, which is done implicitly here and explicitly below, is notationally convenient and is justified provided the $O(\cdot)$ holds uniformly over $k = 1, \ldots, m$. The requisite uniformity holds here, as is stated explicitly. Robinson (1995a, 1995b) also utilizes this notation.

(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 (A.7), to establish Lemma 2(d) it suffices to show that

$$\text{(A.16)} \qquad \sup_{d\in D_m(\eta_n),\,\theta\in\Theta} |\widehat{G}_{a,b}(d,\theta) - \widehat{G}_{a,b}(d_0,\theta)| = o_p(\log^{-2} m).$$

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 (A.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

$$\text{(A.17)} \qquad Z_{a,b}(\eta_n) := \sup_{d\in D_m(\eta_n),\,\theta\in\Theta} |\widehat{E}_{a,b}(d,\theta) - \widehat{E}_{a,b}(d_0,\theta)| = o_p(n^{2d_0}\log^{-2} m)$$

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

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(d). 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 $(\log^5 m)|\widetilde{H} - H_0| = o_p(1)$. The latter holds by the same argument as given by Robinson (1995a, pp. 1642–1643) 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 (A.17) is similar to a proof of Robinson (1995a, p. 1642). We have

$$\text{(A.18)} \qquad Z_{a,b}(\eta_n) = \sup_{d\in D_m(\eta_n),\,\theta\in\Theta} \left| m^{-1}\sum_{j=1}^{m} I_j \exp(p_r(\lambda_j,\theta))(2\log j)^a (j/m)^{2b} j^{2d_0}(j^{2(d-d_0)} - 1) \right|$$

$$\le C \sup_{d\in D_m(\eta_n)} m^{-1}\sum_{j=1}^{m} I_j(\log j)^a j^{2d_0}\left|j^{2(d-d_0)} - 1\right|$$

$$\le 2Ce^{2\eta_n \log^{-4} m} \sup_{d\in D_m(\eta_n)} m^{-1}\sum_{j=1}^{m} I_j(\log j)^{a+1} j^{2d_0}|d - d_0|$$

$$\le \eta_n(\log^{-2} m)2Ce^{2\eta_n \log^{-4} m} m^{-1}\sum_{j=1}^{m} I_j \lambda_j^{2d_0}(2\pi/n)^{-2d_0}$$

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 (A.9) and (A.14). In consequence, the left-hand side of (A.18) is $o_p(n^{2d_0}\log^{-2} m)$, as desired. *Q.E.D.*

PROOF OF LEMMA 2(e): Using (4.2) and (A.10), the normalized score is

$$B_n^{-1}S_n(d_0,\theta_0) = \widehat{G}^{-1}(d_0,\theta_0)m^{-1/2}\sum_{j=1}^{m}\left(y_j(d_0,\theta_0) - m^{-1}\sum_{k=1}^{m}y_k(d_0,\theta_0)\right)\widetilde{X}_j \tag{A.19}$$

$$= (1+o_p(1))m^{-1/2}\sum_{j=1}^{m}\left(\frac{I_j}{g_j}-1\right)\left(\widetilde{X}_j - m^{-1}\sum_{k=1}^{m}\widetilde{X}_k\right),$$

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 (A.9).[7] The right-hand side, with $(1+o_p(1))$ deleted, can be written as

$$T_{1,n} + T_{2,n} + T_{3,n} + T_{4,n}, \quad \text{where}$$

$$T_{1,n} = m^{-1/2}\sum_{j=1}^{m}\left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j} - E\left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j}\right)\right)\left(\widetilde{X}_j - m^{-1}\sum_{k=1}^{m}\widetilde{X}_k\right), \tag{A.20}$$

$$T_{2,n} = m^{-1/2}\sum_{j=1}^{m}\left(\frac{EI_j}{f_j}-1\right)\frac{f_j}{g_j}\left(\widetilde{X}_j - m^{-1}\sum_{k=1}^{m}\widetilde{X}_k\right),$$

$$T_{3,n} = m^{-1/2}\sum_{j=1}^{m}(2\pi I_{\varepsilon j}-1)\left(\widetilde{X}_j - m^{-1}\sum_{k=1}^{m}\widetilde{X}_k\right),$$

$$T_{4,n} = m^{-1/2}\sum_{j=1}^{m}\left(\frac{f_j}{g_j}-1\right)\left(\widetilde{X}_j - m^{-1}\sum_{k=1}^{m}\widetilde{X}_k\right),$$

and $f_j = f(\lambda_j)$, using the fact that $E2\pi I_{\varepsilon j} = 1$. We show that $T_{1,n} = o_p(1)$, $T_{2,n} = o(1)$, $T_{3,n} \to_d N(0,\Omega_r)$, 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:

$$\sum_{j=1}^{k}\left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j} - E\left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j}\right)\right) \tag{A.21}$$

$$= O_p(k^{1/3}\log^{2/3}k + k^{\phi+1/2}n^{-\phi} + k^{1/2}n^{-1/4})$$

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

$$T_{1,n} = m^{-1/2}\sum_{k=1}^{m-1}\sum_{j=1}^{k}\left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j} - E\left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j}\right)\right)\left(\widetilde{X}_k - \widetilde{X}_{k+1}\right) \tag{A.22}$$

$$+ m^{-1/2}\sum_{j=1}^{m}\left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j} - E\left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j}\right)\right)\left(\widetilde{X}_m - m^{-1}\sum_{k=1}^{m}\widetilde{X}_k\right)$$

[7] It might seem, and indeed it was suggested to us by a discussant, that the proof given below can be simplified and the assumptions on m weakened by establishing the asymptotic normality of the right-hand side of (A.19) with g_j replaced by f_j. Ostensibly, this replacement would be justified by noting that f_j/g_j is uniformly bounded away from infinity and zero. Such a replacement, however, is not correct. If it were, the normalized score would be asymptotically normal with mean zero, which is not the case.

$$
\begin{aligned}
&= m^{-1/2}\sum_{k=1}^{m-1} O_p(k^{1/3}\log^{2/3} k + k^{\phi+1/2}n^{-\phi} + k^{1/2}n^{-1/4})O(k^{-1}) \\
&\quad + m^{-1/2}O_p(m^{1/3}\log^{2/3} m + m^{\phi+1/2}n^{-\phi} + m^{1/2}n^{-1/4})O(1) \\
&= O_p\big(m^{-1/6}\log^{2/3} m + (m/n)^{\phi} + n^{-1/4}\big) \\
&= o_p(1),
\end{aligned}
$$

where the second equality uses $\widetilde{X}_k - \widetilde{X}_{k+1} = O(k^{-1})$ uniformly over $k = 1, \ldots, m$ (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 $i = 1, \ldots, r$) 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 and $(m/m)^{2i} - m^{-1}\sum_{j=1}^m (j/m)^{2i} = O(1)$ for $i = 1, \ldots, r$.

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

$$
EI_j/f_j = 1 + O(j^{-1}\log j) \tag{A.23}
$$

uniformly over $j = 1, \ldots, m$. 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 (A.23) is different from that in Theorem 2 of Robinson (1995b) because the proof of (A.23) only requires (4.1), and not (4.2), of Robinson (1995b) to hold, given the normalization by f_j.

By (A.23),

$$
\begin{aligned}
T_{2,n} &= m^{-1/2}\sum_{j=1}^m O(j^{-1}\log j)O(1)\left(\widetilde{X}_j - m^{-1}\sum_{k=1}^m \widetilde{X}_k\right) \\
&= O\left(m^{-1/2}\log m \sum_{j=1}^m j^{-1}\log j\right) \\
&= O(m^{-1/2}\log^3 m) = o(1).
\end{aligned} \tag{A.24}
$$

We show that $\beta' T_{3,n} \to_d N(0, \beta'\Omega_r\beta)$ for all $\beta \neq 0$ using the same proof as Robinson's (1995a, pp. 1644–1647) proof that $m^{-1/2}\sum_{j=1}^m (2\pi I_{ej} - 1)2\nu_j \to_d N(0,4)$, 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) $m^{-1}\sum_{j=1}^m \zeta_j^2 \to \beta'\Omega_r\beta$ as $n \to \infty$ by Lemma 2(d) and (ii) $|\zeta_j - \zeta_{j+1}| \le \|\beta\| \cdot \|\widetilde{X}_j - \widetilde{X}_{j+1}\| \le Cj^{-1}$ for some constant $C < \infty$ independent of j, which is needed in (4.21) of Robinson's proof.

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

$$
\begin{aligned}
\log(f_j/g_j) &= \log\varphi(\lambda_j) - \log G_0 + p_r(\lambda_j, \theta_0) \\
&= \mathbb{1}(s \ge 2 + 2r)\frac{b_{2+2r}}{(2+2r)!}\lambda_j^{2+2r} + O(\lambda_j^q) \quad \text{and} \\
f_j/g_j &= 1 + \mathbb{1}(s \ge 2+2r)\frac{b_{2+2r}}{(2+2r)!}\lambda_j^{2+2r} + O(\lambda_j^q), \quad \text{where} \\
q &= \min\{s, 4+2r\},
\end{aligned} \tag{A.25}
$$

uniformly over $j = 1, \ldots, m$, 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 $s \geq 2 + 2r$,

$$
\begin{aligned}
T_{4,n} &= m^{-1/2} \sum_{j=1}^{m} \left(\frac{b_{2+2r}}{(2+2r)!} \lambda_j^{2+2r} + O(\lambda_j^q) \right) \left(\widetilde{X}_j - m^{-1} \sum_{k=1}^{m} \widetilde{X}_k \right) \\
&= m^{5/2+2r} n^{-(2+2r)} m^{-1} \sum_{j=1}^{m} \frac{(2\pi)^{2+2r} b_{2+2r}}{(2+2r)!} \left(\frac{j}{m} \right)^{2+2r} \left(\widetilde{X}_j - m^{-1} \sum_{k=1}^{m} \widetilde{X}_k \right) \\
&\quad + m^{-1/2} \sum_{j=1}^{m-1} (\widetilde{X}_j - \widetilde{X}_{j+1}) \sum_{i=1}^{j} O(\lambda_i^q) + m^{-1/2} \sum_{j=1}^{m} O(\lambda_j^q) \left(\widetilde{X}_m - m^{-1} \sum_{k=1}^{m} \widetilde{X}_k \right),
\end{aligned} \tag{A.26}
$$

where the second equality uses summation by parts. The second and third summands on the right-hand side of (A.26) are $O(m^{q+1/2} n^{-q})$ because $\widetilde{X}_j - \widetilde{X}_{j+1} = O(j^{-1})$ uniformly over $j = 1, \ldots, m$ and $\widetilde{X}_m - m^{-1} \sum_{k=1}^{m} \widetilde{X}_k = 1 + o(1)$ by the calculations following (A.22).

The following results are proved by approximating sums by integrals; see the proof of Lemma 1 in AG for details. Suppose $m \to \infty$; then for $k = 1, \ldots, r$,

$$
\begin{aligned}
&\frac{1}{m} \sum_{j=1}^{m} \left(\frac{j}{m} \right)^{2+2r} \left(\left(\frac{j}{m} \right)^{2k} - \frac{1}{m} \sum_{i=1}^{m} \left(\frac{i}{m} \right)^{2k} \right) \\
&\quad = \frac{1}{2r+2k+3} - \frac{1}{(3+2r)(2k+1)} + o(1) \\
&\quad = \frac{(2+2r)}{(3+2r)^2} \xi_{r,k} + o(1) \quad \text{and} \\
&\frac{1}{m} \sum_{j=1}^{m} \left(\frac{j}{m} \right)^{2+2r} \left(2 \log j - \frac{1}{m} \sum_{i=1}^{m} 2 \log i \right) = \frac{2(2r+2)}{(3+2r)^2} + o(1).
\end{aligned} \tag{A.27}
$$

For the case where $s > 2 + 2r$, the combination of (A.26) and (A.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 (A.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 (A.26) with the term that contains b_{2+2r} deleted and with $q = s$. Hence, by the remarks following (A.26), $T_{4,n} = O(m^{s+1/2} n^{-s}) = -\nu_n(r, s)$.

Now we prove (A.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

$$
\sum_{j=1}^{\ell} \left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j} - E\left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j} \right) \right) = O_p(k^{1/3} \log^{2/3} k) \quad \text{as} \quad n \to \infty \tag{A.28}
$$

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

$$
\begin{aligned}
&\sum_{j=\ell+1}^{k} \left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j} - E\left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j} \right) \right) \\
&\quad = \sum_{j=\ell+1}^{k} \left(\left(\frac{I_j}{f_j} - 2\pi I_{\varepsilon j} \right) \frac{f_j}{g_j} - E\left(\frac{I_j}{f_j} - 2\pi I_{\varepsilon j} \right) \frac{f_j}{g_j} \right) + 2\pi \sum_{j=\ell+1}^{k} (I_{\varepsilon j} - E I_{\varepsilon j}) \left(\frac{f_j}{g_j} - 1 \right) \\
&\quad := A_1 + A_2.
\end{aligned} \tag{A.29}
$$

We have

$$EA_1^2 \le E\left(\sum_{j=\ell+1}^{k}\left(\frac{I_j}{f_j}-2\pi I_{\varepsilon j}\right)\frac{f_j}{g_j}\right)^2 = O(k^{2/3}\log^{4/3}k + kn^{-1/2}), \tag{A.30}$$

where the equality holds by the same proof as in Robinson (1995a, pp. 1648–1651) 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 $j = 1, \ldots, m$.

Next, we have

$$\begin{aligned} EA_2^2 &= 4\pi^2\sum_{j=\ell+1}^{k}\operatorname{var}(I_{\varepsilon j})\left(\frac{f_j}{g_j}-1\right)^2 + 8\pi^2\sum_{i=\ell+1}^{k}\sum_{j=\ell+1}^{i-1}\operatorname{cov}(I_{\varepsilon i}, I_{\varepsilon j})\left(\frac{f_i}{g_i}-1\right)\left(\frac{f_j}{g_j}-1\right) \\ &= O(1)\sum_{j=\ell+1}^{k}\lambda_j^{2\phi} + O(n^{-1})\sum_{i=\ell+1}^{k}\sum_{j=\ell+1}^{i-1}\lambda_i^{\phi}\lambda_j^{\phi} \\ &= O\left(k\left(\frac{k}{n}\right)^{2\phi}\right) + O\left(\frac{k^2}{n}\left(\frac{k}{n}\right)^{2\phi}\right) \\ &= O\left(k\left(\frac{k}{n}\right)^{2\phi}\right), \end{aligned} \tag{A.31}$$

where the second equality uses (A.25) and Theorem 5.2.4 of Brillinger (1975, p. 125), which states that $\operatorname{var}(I_{\varepsilon j}) = O(1)$ and $\operatorname{cov}(I_{\varepsilon i}, I_{\varepsilon j}) = O(n^{-1})$ uniformly over $i, j = 1, \ldots, n$ with $i \ne j$. 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 $\{\varepsilon_t : t = 1, 2, \ldots\}$ is the same as that of an iid process with finite fourth moment, which is sufficient.

Combining (A.28)–(A.31) gives (A.21). *Q.E.D.*

PROOF OF THEOREM 3(b): The choice of r as the largest integer less than $s/2$ implies that $s > 2r$ and $s \le 2 + 2r$. Hence, $\nu_n(r, s) = O(m^{s+1/2}n^{-s}) = O(1)$. By the definition of m, $m^{1/2} = \psi_1^{1/2}n^{s/(2s+1)}$. In consequence, the result of Theorem 3(b) follows from

$$\sup_{f\in\mathcal{F}(s,a,\delta,K)}\left|P_f\left(\begin{pmatrix} m^{1/2}(\widehat{d}(r) - d_0) \\ m^{1/2}\operatorname{Diag}((2\pi m/n)^2, \ldots, (2\pi m/n)^{2r})(\widehat{\theta}(r) - \theta_0)\end{pmatrix} - \Omega_r^{-1}\nu_n(r, s) \le x\right) - \Phi(\Omega_r^{1/2}x)\right| \to 0 \tag{A.32}$$

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

To prove (A.32), 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(b) 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 $\{K_n : n \ge 1\}$ in condition (iv) does not depend on f. In consequence, the conclusion of the lemma is that a solution to the FOC's 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 (b)–(d) of Lemma 2 hold with $o_p(1)$ holding unif-f. Inspection of their proofs shows that they do. Part (a) 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 AG; 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

$$\sup_{f\in\mathcal{F}(s,a,\delta,K)} \left|P_f\left(B_n^{-1}S_n(d_0,\theta_0) + \nu_n(r,s) \le x\right) - \Phi(\Omega_r^{-1/2}x)\right| \to 0 \quad \text{as} \quad n\to\infty \tag{A.33}$$

for all $x \in \mathbb{R}^{r+1}$.

Given the above unif-f extensions of the results of Lemmas 1 and 2, Theorems 1 and 2 have analogous extensions. The result of the extended Theorem 1 is the same as that of (A.32) with $\widehat{d}(r)$ replaced by $\widetilde{d}(r)$. The result of the extended Theorem 2 is exactly (A.32). *Q.E.D.*

Before proving Theorem 4, we introduce some notation and state two lemmas. The proofs of the lemmas are given after that of Theorem 4. Let $R_{r(\tau),\tau}(d,\theta)$, $B_{n,\tau}$, $S_{n,\tau}(d,\theta)$, $H_{n,\tau}(d,\theta)$, $X_{j,\tau}$, $D_{m(\tau)}(\eta)$, and J_n^τ be defined as the corresponding quantities without the τ subscripts (or superscript) are defined in the text, but with m and r replaced by $m(\tau)$ and $r(\tau)$, respectively. (We use the symbol J_n^τ rather than $J_{n,\tau}$ because $J_{a,b}$ is used for a different quantity in Proof of Lemma 2(b) above and also below.) Let θ_0 denote θ_0 defined as in (3.2) with $r = r(\tau)$. Let $\widehat{d}_\tau$ and $\widehat{\theta}_\tau$ denote the LPW estimators with $m = m(\tau)$ and $r = r(\tau)$.

Let $1 \le s_* \le s^* < \infty$. For any sequences of sets $\{E_{n,\tau} : n \ge 1\}$ for $\tau \in [s_*, s^*]$, we say that the sets $\{E_{n,\tau} : n \ge 1\}$ are "*uniformly* $\zeta^{-2}(n)$ *small*" if for some positive finite constant C

$$\sup_{s\in[s_*,s^*]} \sup_{f\in\mathcal{F}(s,a,\delta,K)} \sup_{\tau\in[s_*,s]} P_f(E_{n,\tau}) \le C\zeta^{-2}(n) \quad \text{for all} \quad n \ge 1. \tag{A.34}$$

LEMMA 4: *Suppose the assumptions of Theorem* 4 *hold. Let* $s_* \ge 1$. *Then, for each* $s^* \in [s_*, \infty)$, *the* LPW *estimators* $\{(\widehat{d}_\tau, \widehat{\theta}_\tau) : n \ge 1\}$ *satisfy*

$$B_{n,\tau}\begin{pmatrix}\widehat{d}_\tau - d_f\\ \widehat{\theta}_\tau - \theta_0\end{pmatrix} = -\Omega_{r(\tau)}^{-1}B_{n,\tau}^{-1}S_{n,\tau}(d_f,\theta_0) + \varepsilon_{n,\tau},$$

where the sets $\{\{\|\varepsilon_{n,\tau}\| > C_*\zeta(n)\} : n \ge 1\}$ *are uniformly* $\zeta^{-2}(n)$ *small for all* $C_* > 0$.

The dimensions of θ, $S_{n,\tau}(d,\theta)$, $H_{n,\tau}(d,\theta)$, $X_{j,\tau}$, and J_n^τ depend on τ through $r(\tau)$, the degree of the polynomial. But, by definition, $r(\tau)$ is constant for all $\tau \in T_w$, where

$$T_w = (2w, 2w+2] \quad \text{for} \quad w = 0, 1, \ldots. \tag{A.35}$$

Given this, in the following lemma, we consider $\tau \in T_w \cap [s_*, s]$ separately for a finite number of values w rather than considering $\tau \in [s_*, s]$ all at once.

LEMMA 5: *Suppose the assumptions of Theorem* 4 *hold. Let* $s_* \ge 1$. *For each* $s^* \in [s_*, \infty)$, *each integer* w *in the set* $\{0, 1, \ldots, [(s^*-2)/2]+1\}$, *and each constant* $C_* > 0$, *there exists a positive finite constant* C *such that*

$$\sup_{s\in[s_*,s^*]} \sup_{f\in\mathcal{F}(s,a,\delta,K)} \sup_{\tau\in T_w\cap[s_*,s]} P_f\left(\|B_{n,\tau}^{-1}S_{n,\tau}(d_f,\theta_0)\| > C_*\zeta(n)\right) \le C\zeta^{-2}(n).$$

PROOF OF THEOREM 4: We write

$$P_f\left(n^{\frac{s}{2s+1}}\zeta^{-1}(n)|\widehat{d}_{\hat{s}} - d_f| \ge C_1\right) := \Pi_n^+ + \Pi_n^-, \quad \text{where} \tag{A.36}$$

$$\Pi_n^+ = P_f\left(n^{\frac{s}{2s+1}}\zeta^{-1}(n)|\widehat{d}_{\hat{s}} - d_f| \ge C_1,\ \hat{s} \ge s\right) \quad \text{and}$$

$$\Pi_n^- = P_f\left(n^{\frac{s}{2s+1}}\zeta^{-1}(n)|\widehat{d}_{\hat{s}} - d_f| \ge C_1,\ \hat{s} < s\right).$$

We want to show that $\lim_{C_1\to\infty}\limsup_{n\to\infty}\sup_{s\in[s_*,s^*]}\sup_{f\in\mathcal{F}(s,a,\delta,K)}\Pi_n^+=0$ and likewise for Π_n^-.

We consider Π_n^+ first. By the triangle inequality and the definition of $\hat{s}$, we have

$$\begin{aligned}\Pi_n^+ &\le P_f\big(n^{\frac{s}{2s+1}}\zeta^{-1}(n)|\widehat{d}_{\hat{s}}-\widehat{d}_s|\ge C_1/2,\hat{s}\ge s\big)\\ &\quad+P_f\big(n^{\frac{s}{2s+1}}\zeta^{-1}(n)|\widehat{d}_s-d_f|\ge C_1/2\big)\\ &\le P_f\big(n^{\frac{s}{2s+1}}\zeta^{-1}(n)m^{-1/2}(s)\psi_2(c_{r(s)}/4)^{1/2}\zeta(n)\ge C_1/2\big)\\ &\quad+P_f\big(n^{\frac{s}{2s+1}}\zeta^{-1}(n)|\widehat{d}_s-d_f|\ge C_1/2\big)\\ &:=\Pi_{n,1}^+ +\Pi_{n,2}^+.\end{aligned}\tag{A.37}$$

We have

$$\begin{aligned}&\lim_{C_1\to\infty}\limsup_{n\to\infty}\sup_{s\in[s_*,s^*]}\sup_{f\in\mathcal{F}(s,a,\delta,K)}\Pi_{n,1}^+\\ &\quad=\lim_{C_1\to\infty}\mathbb{1}\big(\psi_1^{-1/2}\psi_2(c_{r(s^*)}/4)^{1/2}\ge C_1/2\big)=0\end{aligned}\tag{A.38}$$

using the fact that c_r is nondecreasing in r and $r(s)$ is nondecreasing in s.

Note that the $(1,1)$ element of the diagonal matrix $B_{n,\tau}$ is $m^{1/2}(\tau)$. Thus, Lemmas 4 and 5 and the nonsingularity of $\Omega_{r(\tau)}$ combine to give: for each $C_*>0$ there is a constant $C<\infty$ such that

$$\sup_{s\in[s_*,s^*]}\sup_{f\in\mathcal{F}(s,a,\delta,K)}\sup_{\tau\in[s_*,s]}P_f\big(m^{1/2}(\tau)|\widehat{d}_\tau-d_f|>C_*\zeta(n)\big)\le C\zeta^{-2}(n).\tag{A.39}$$

This establishes the desired result for $\Pi_{2,n}^+$.

Next, we consider Π_n^-. We have

$$\begin{aligned}\Pi_n^- &=\sum_{\tau\in\mathcal{S}_h:\tau+h<s}P_f\big(n^{\frac{s}{2s+1}}\zeta^{-1}(n)|\widehat{d}_\tau-d_f|\ge C_1,\hat{s}=\tau\big)\\ &\quad+P_f\big(n^{\frac{s}{2s+1}}\zeta^{-1}(n)|\widehat{d}_{\tau_s}-d_f|\ge C_1,\hat{s}=\tau_s\big)\\ &\le\sum_{\tau\in\mathcal{S}_h:\tau+h<s}P_f(\hat{s}=\tau)+P_f\big(n^{\frac{s}{2s+1}}\zeta^{-1}(n)|\widehat{d}_{\tau_s}-d_f|\ge C_1\big)\\ &:=\Pi_{n,1}^- +\Pi_{n,2}^-,\end{aligned}\tag{A.40}$$

where $\tau_s\in\mathcal{S}_h$ and $s-h\le\tau_s<s$.

Now, we bound $P_f(\hat{s}=\tau)$. By the definition of $\hat{s}$, if $\hat{s}=\tau$, there exists $\tau'\le\tau$, $\tau'\in\mathcal{S}_h$ such that $|\widehat{d}_{\tau+h}-\widehat{d}_{\tau'}|>m^{-1/2}(\tau')\psi_2(c_{r(\tau')}/4)^{1/2}\zeta(n)$. In consequence, for all $\tau\in\mathcal{S}_h$ with $\tau+h<s$,

$$\begin{aligned}P_f(\hat{s}=\tau) &\le\sum_{\tau'\in\mathcal{S}_h:\tau'\le\tau}P_f\big(|\widehat{d}_{\tau+h}-\widehat{d}_{\tau'}|>m^{-1/2}(\tau')\psi_2(c_{r(\tau')}/4)^{1/2}\zeta(n)\big)\\ &\le\sum_{\tau'\in\mathcal{S}_h:\tau'\le\tau}P_f\big(m^{1/2}(\tau+h)|\widehat{d}_{\tau+h}-d_f|>\kappa\zeta(n)\big)\\ &\quad+\sum_{\tau'\in\mathcal{S}_h:\tau'\le\tau}P_f\big(m^{1/2}(\tau')|\widehat{d}_{\tau'}-d_f|>\kappa\zeta(n)\big),\\ &\le 2(s^*-s_*)(\log n)\sup_{\tau''<s}P_f\big(m^{1/2}(\tau'')|\widehat{d}_{\tau''}-d_f|>\kappa\zeta(n)\big),\end{aligned}\tag{A.41}$$

where $\kappa=\psi_2(c_{r(s_*)}/4)^{1/2}/2$. The third inequality holds because there are at most $(s^*-s_*)(\log n)$ elements $\tau'\in\mathcal{S}_h$ for which $\tau'\le\tau$. Note that the third inequality only applies for τ such that $\tau+h<s$. It is for this reason that we decompose Π_n^- into $\Pi_{n,1}^-+\Pi_{n,2}^-$ in (A.40).

Equations (A.39)–(A.41) give: for some $C < \infty$,

$$\sup_{s\in[s_*,s^*]}\sup_{f\in\mathcal{F}(s,a,\delta,K)} \Pi_{n,1}^- \le 2(s^*-s_*)^2(\log n)^2 C\zeta^{-2}(n) \tag{A.42}$$
$$= O((\log\log n)^{-1}) = o(1) \quad \text{as} \quad n\to\infty.$$

Next, we have

$$n^{s/(2s+1)}n^{-\tau_s/(2\tau_s+1)} \le n^{s/(2s+1)}n^{-(s-h)/(2s-2h+1)} = n^{\kappa_{s,h}h} \le n^h = n^{\log^{-1}n} = e, \tag{A.43}$$

where $\kappa_{s,h} = (2s+1)^{-1}(2s-2h+1)^{-1} \le 1$. This, $\tau_s < s$, and (A.39) give: for some $C < \infty$,

$$n^{s/(2s+1)} \le m^{1/2}(\tau_s)e\psi_1^{-1/2} \quad \text{and} \tag{A.44}$$

$$\Pi_{n,2}^- \le P_f\big(m^{1/2}(\tau_s)|\widehat{d}_{\tau_s} - d_f| \ge C_1 e^{-1}\psi_1^{1/2}\zeta(n)\big)$$
$$\le C\zeta^{-2}(n) = o(1) \quad \text{as} \quad n\to\infty.$$

This completes the proof of the theorem. *Q.E.D.*

Next, we prove Lemma 4. For convenience, what we show is that for some $C < \infty$,

$$\sup_{s\in[s_*,s^*]}\sup_{f\in\mathcal{F}(s,a,\delta,K)}\sup_{\tau\in[s_*,s]} P_f(\|\varepsilon_{n,\tau}\| > C\zeta(n)) \le C\zeta^{-2}(n) \quad \text{for all} \quad n\ge 1. \tag{A.45}$$

This is sufficient because the proof of (A.45) goes through unchanged with $\zeta(n)$ replaced by $\widetilde{C}\zeta(n)$ for all constants $\widetilde{C} > 0$.

To show (A.45), we establish three lemmas that reflect the same steps as are used in the text to prove Theorem 2. We start by establishing an analogue to Lemma 1. Let $\{L_{n,\tau}(\gamma) : n \ge 1\}$ be a sequence of minimands for estimation of the parameter $\gamma_0 \in \Gamma \subset \mathbb{R}^k$, where Γ is the parameter space and $\tau \in T$ indexes different minimands. Suppose the distribution of $L_{n,\tau}(\gamma)$ depends on $f \in \mathcal{F}$, where $\mathcal{F}$ denotes some index set. Let $\mathcal{G} \subset T \times \mathcal{F}$. Denote the gradient and Hessian of $L_{n,\tau}(\gamma)$ by $\nabla L_{n,\tau}(\gamma)$ and $\nabla^2 L_{n,\tau}(\gamma)$ respectively. Let $\{\zeta(n) : n \ge 1\}$ be a sequence of positive constants for which $\zeta(n) \to \infty$ as $n \to \infty$. In our application of the lemma, we take $\zeta(n) = (\log n)(\log\log(n))^{1/2}$.

LEMMA 6: *Suppose γ_0 is in the interior of Γ, $L_{n,\tau}(\gamma)$ is twice continuously differentiable in γ on a neighborhood of γ_0 for all $\tau \in T$, and there exist a positive finite constant C and sequences of $k\times k$ nonrandom nonsingular matrices $B_{n,\tau}$ such that*

(i) $\sup_{\tau\in T}\|B_{n,\tau}^{-1}\|\zeta(n) \to 0$ *as* $n\to\infty$,

(ii) $\sup_{(\tau,f)\in\mathcal{G}} P_f(\|(B_{n,\tau}^{-1})'\nabla L_{n,\tau}(\gamma_0)\| > C\zeta(n)) \le C\zeta^{-2}(n)$ *for all* $n \ge 1$,

(iii) $\sup_{(\tau,f)\in\mathcal{G}} P_f(\lambda_{\min}((B_{n,\tau}^{-1})'\nabla^2 L_{n,\tau}(\gamma_0)B_{n,\tau}^{-1}) < C) \le C\zeta^{-2}(n)$ *for all* $n\ge 1$, *and*

(iv)

$$\sup_{(\tau,f)\in\mathcal{G}} P_f\Bigg(\sup_{\gamma\in\Gamma:\|B_{n,\tau}(\gamma-\gamma_0)\|\le K_{n,\tau}} \big\|(B_{n,\tau}^{-1})'\big(\nabla^2 L_{n,\tau}(\gamma) - \nabla^2 L_{n,\tau}(\gamma_0)\big)B_{n,\tau}^{-1}\big\| > C\zeta^{-1}(n)\Bigg)$$
$$\le C\zeta^{-2}(n) \quad \textit{for all} \quad n\ge 1$$

for some sequences of scalar constants $\{K_{n,\tau} : n \ge 1\}$ for which $K_{n,\tau}\zeta^{-1}(n) \to \infty$ as $n\to\infty$ for each $\tau \in T$.

Then, there exist a positive finite constant C_1 and sequences of estimators $\{\widetilde{\gamma}_{n,\tau} : n \ge 1\}$ for each $\tau \in T$ such that the probability that the first-order conditions do not hold at $\widetilde{\gamma}_{n,\tau}$ goes to zero at the following rate:

$$\sup_{(\tau,f)\in\mathcal{G}} P_f\big(\nabla L_{n,\tau}(\widetilde{\gamma}_{n,\tau}) \ne 0\big) \le C_1\zeta^{-2}(n) \quad \textit{for all} \quad n\ge 1$$

and

$$B_{n,\tau}(\widetilde{\gamma}_{n,\tau}-\gamma_0)=-Y_{n,\tau}+\varepsilon_{n,\tau},\quad \textit{where}$$

$$Y_{n,\tau}=\big((B_{n,\tau}^{-1})'\nabla^2 L_{n,\tau}(\gamma_0)B_{n,\tau}^{-1}\big)^{-1}(B_{n,\tau}^{-1})'\nabla L_{n,\tau}(\gamma_0)\quad \textit{and}$$

$$\sup_{(\tau,f)\in\mathcal{G}} P_f\big(\|\varepsilon_{n,\tau}\|>C_1\zeta(n)\big)\le C_1\zeta^{-2}(n).$$

PROOF OF LEMMA 6: Throughout let C' denote a positive finite constant that may differ across equations. Let

$$\text{(A.46)}\qquad \Gamma_{n0}=\big\{\gamma\in\Gamma:\|B_{n,\tau}(\gamma-\gamma_0)\|\le K_{n,\tau},\ \|\gamma-\gamma_0\|<\delta\big\}$$

for some $\delta>0$ such that $L_{n,\tau}(\gamma)$ is twice differentiable on $\{\gamma\in\mathbb{R}^k:\|\gamma-\gamma_0\|<\delta\}$ and $\{\gamma\in\mathbb{R}^k:\|\gamma-\gamma_0\|<\delta\}\subset\Gamma$.

By a Taylor expansion about γ_0 and some algebra, we obtain: for $\gamma\in\Gamma_{n0}$,

$$\begin{aligned}\text{(A.47)}\qquad & L_{n,\tau}(\gamma)-L_{n,\tau}(\gamma_0)\\ &=\nabla L_{n,\tau}(\gamma_0)'(\gamma-\gamma_0)+\frac{1}{2}(\gamma-\gamma_0)'\nabla^2 L_{n,\tau}(\gamma_0)(\gamma-\gamma_0)+\rho_{n,\tau}(\gamma)\\ &=\frac{1}{2}(B_{n,\tau}(\gamma-\gamma_0)+Y_{n,\tau})'\big[(B_{n,\tau}^{-1})'\nabla^2 L_{n,\tau}(\gamma_0)B_{n,\tau}^{-1}\big](B_{n,\tau}(\gamma-\gamma_0)+Y_{n,\tau})\\ &\quad-\frac{1}{2}Y_{n,\tau}'(B_{n,\tau}^{-1})'\nabla L_{n,\tau}(\gamma_0)+\rho_{n,\tau}(\gamma),\end{aligned}$$

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

$$\text{(A.48)}\qquad |\rho_{n,\tau}(\gamma)|\le\|B_{n,\tau}(\gamma-\gamma_0)\|^2\sup_{\overline{\gamma}\in\Gamma_{n0}}\big\|(B_{n,\tau}^{-1})'\big(\nabla^2 L_{n,\tau}(\overline{\gamma})-\nabla^2 L_{n,\tau}(\gamma_0)\big)B_{n,\tau}^{-1}\big\|.$$

Let $\gamma_{n,\tau}^*=\gamma_0-B_{n,\tau}^{-1}Y_{n,\tau}$. Conditions (ii) and (iii) imply that for some C',

$$\text{(A.49)}\qquad \sup_{(\tau,f)\in\mathcal{G}} P_f(\|Y_{n,\tau}\|>C'\zeta(n))\le C'\zeta^{-2}(n)\quad\text{for all}\quad n\ge 1.$$

This, $K_{n,\tau}\zeta^{-1}(n)\to\infty$ as $n\to\infty$, and condition (i) imply that $\sup_{(\tau,f)\in\mathcal{G}}P_f(\gamma_{n,\tau}^*\notin\Gamma_{n0})\le C'\zeta^{-2}(n)$ for all $n\ge 1$. In consequence, by (A.47), (A.48), and condition (iv), for some C',

$$\begin{aligned}\text{(A.50)}\qquad & L_{n,\tau}(\gamma_{n,\tau}^*)-L_{n,\tau}(\gamma_0)=-\frac{1}{2}Y_{n,\tau}'(B_{n,\tau}^{-1})'\nabla L_{n,\tau}(\gamma_0)+\rho_{n,\tau}(\gamma_{n,\tau}^*),\\ &|\rho_{n,\tau}(\gamma_{n,\tau}^*)|\le\|Y_{n,\tau}\|^2\sup_{\overline{\gamma}\in\Gamma_{n0}}\big\|(B_{n,\tau}^{-1})'\big(\nabla^2 L_{n,\tau}(\overline{\gamma})-\nabla^2 L_{n,\tau}(\gamma_0)\big)B_{n,\tau}^{-1}\big\|,\quad\text{and}\\ &\sup_{(\tau,f)\in\mathcal{G}} P_f\big(|\rho_{n,\tau}(\gamma_{n,\tau}^*)|>C'\zeta(n)\big)\le C'\zeta^{-2}(n)\quad\text{for all}\quad n\ge 1.\end{aligned}$$

Let C'' be any positive finite constant. Define

$$\begin{aligned}\text{(A.51)}\qquad & \Gamma_{n,\tau}(x)=\{\gamma\in\Gamma:\|B_{n,\tau}(\gamma-\gamma_0)+Y_{n,\tau}\|\le x\}\quad\text{and}\\ &C_n=C''\zeta(n).\end{aligned}$$

Note that $\gamma_{n,\tau}^*$ is in the interior of $\Gamma_{n,\tau}(C_n)$ unless $\gamma_{n,\tau}^*$ is not in Γ. The latter occurs with probability less than or equal to $C'\zeta^{-2}(n)$ for n large by (A.49) and condition (i).

We have $\Gamma_{n,\tau}(C_n)\subset\Gamma_{n0}$ except on a sequence of sets with small probabilities:

$$\text{(A.52)}\qquad \sup_{(\tau,f)\in\mathcal{G}}\big(1-P_f(\Gamma_{n,\tau}(C_n)\subset\Gamma_{n0})\big)\le C'\zeta^{-2}(n)$$

for n large, by (A.49) and the assumption that $K_{n,\tau}\zeta^{-1}(n) \to \infty$ as $n \to \infty$. In consequence, for some C',

$$\sup_{(\tau,f)\in\mathcal{G}} P_f\left(\sup_{\gamma\in\Gamma_{n,\tau}(C_n)} |\rho_{n,\tau}(\gamma)| > C'\zeta(n) \right) \le C'\zeta^{-2}(n) \tag{A.53}$$

using (A.48) and condition (iv).

Let $\partial\Gamma_{n,\tau}(C_n)$ denote the boundary of $\Gamma_{n,\tau}(C_n)$. Combining (A.47), (A.50), and (A.53), for all $\gamma \in \partial\Gamma_{n,\tau}(C_n)$ and some C',

$$\big(L_{n,\tau}(\gamma) - L_{n,\tau}(\gamma^*_{n,\tau})\big)C_n^{-2} = \frac{1}{2}\mu_n'(B_{n,\tau}^{-1})'\nabla^2 L_{n,\tau}(\gamma_0)B_{n,\tau}^{-1}\mu_n + \lambda_{n,\tau}(\gamma), \tag{A.54}$$

for some k-vector μ_n with $\|\mu_n\| = 1$, where

$$\lambda_{n,\tau}(\gamma) = \big(\rho_{n,\tau}(\gamma) - \rho_{n,\tau}(\gamma^*_{n,\tau})\big)(C'')^{-2}\zeta^{-2}(n) \quad \text{and} \tag{A.55}$$

$$\sup_{(\tau,f)\in\mathcal{G}} P_f\left(\sup_{\gamma\in\Gamma_{n,\tau}(C_n)} |\lambda_{n,\tau}(\gamma)| > C'\zeta^{-1}(n) \right) \le C'\zeta^{-2}(n).$$

Let $A_{n,\tau}$ be the set on which the first summand on the right-hand side of (A.54) is greater than $C > 0$ uniformly over all k-vectors μ_n with $\|\mu_n\| = 1$ and $\sup_{\gamma\in\Gamma_{n,\tau}(C_n)} |\lambda_{n,\tau}(\gamma)| < C$. Then, the complement of $A_{n,\tau}$, i.e., $A^c_{n,\tau}$, satisfies for some C'

$$\sup_{(\tau,f)\in\mathcal{G}} P_f(A^c_{n,\tau}) \le C'\zeta^{-2}(n) \tag{A.56}$$

for n large by condition (iii) and (A.55). Hence, the minimum of $L_{n,\tau}(\gamma)$ over $\gamma \in \partial\Gamma_n(C_n)$ is greater than its value at the interior point $\gamma^*_{n,\tau}$ except on $A^c_{n,\tau}$. In consequence, for each n large and all $\tau \in T$, the minimum of $L_{n,\tau}(\gamma)$ over $\gamma \in \Gamma_n(C_n)$ is attained at a point, say $\widetilde{\gamma}_{n,\tau}$ (not necessarily unique), in the interior of $\Gamma_n(C_n)$ except on $A^c_{n,\tau}$. These points satisfy the first-order conditions $\nabla L_n(\widetilde{\gamma}_{n,\tau}) = 0$ except on $A^c_{n,\tau}$. In addition, $\widetilde{\gamma}_{n,\tau} \subset \Gamma_n(C_n)$ and, hence, $\|B_{n,\tau}(\widetilde{\gamma}_{n,\tau} - \gamma_0) + Y_n\| \le C''\zeta(n)$ except on $A^c_{n,\tau}$. Given (A.56), this completes the proof of the lemma. *Q.E.D.*

Next, we apply Lemma 6 with $L_{n,\tau}(\gamma) = m(\tau)R_{r(\tau),\tau}(d,\theta)$, $\gamma = (d,\theta')'$, and $\mathcal{F} = \bigcup_{s\in[s_*,s^*]}\mathcal{F}(s,a,\delta,K)$. We apply Lemma 6 a finite number of times—each time with $T = T_w \cap [s_*, s^*]$ and $\mathcal{G} = \mathcal{G}_w = \{(\tau,f) : \tau \in T_w \cap [s_*, s],\ f \in \mathcal{F}(s,a,\delta,K)$ for some $s \in [s_*, s^*]\}$ for some $w \in \mathcal{W} = \{0,1,\ldots,[(s^*-2)/2]+1\}$—to get results that hold for all $(\tau,f) \in \bigcup_{w\in\mathcal{W}}\mathcal{G}_w$. Note that the supremum of a function over $(\tau,f) \in \bigcup_{w\in\mathcal{W}}\mathcal{G}_w$ equals the supremum over $s \in [s_*, s^*]$, $f \in \mathcal{F}(s,a,\delta,K)$, and $\tau \in [s_*, s]$.

The definition of $m(\tau)$ guarantees that $\|B_{n,\tau}^{-1}\|\zeta(n) \to 0$, as required by condition (i) of Lemma 6. Condition (ii) holds by Lemma 5. To verify conditions (iii) and (iv) of Lemma 6, we need to establish some properties of the Hessian of $m(\tau)R_{r(\tau),\tau}(d,\theta)$, viz., $H_{n,\tau}(d,\theta)$. This is done in the following lemma.

LEMMA 7: *Suppose the assumptions of Theorem* 4 *hold. Let* $s_* \ge 1$. *For each* $s^* \in [s_*, \infty)$ *and each integer* w *in the set* $\mathcal{W}$, *there exists a positive finite constant* C *such that*

(a) $$\sup_{\tau\in T_w\cap[s_*,s^*]} \|B_{n,\tau}^{-1}J_n^\tau B_{n,\tau}^{-1} - \Omega_{r(\tau)}\| \to 0 \quad \text{as} \quad n \to \infty,$$

(b) $$\sup_{s\in[s_*,s^*]}\sup_{f\in\mathcal{F}(s,a,\delta,K)}\sup_{\tau\in T_w\cap[s_*,s]} P_f\big(\big\|B_{n,\tau}^{-1}(H_{n,\tau}(d_f,\theta_0) - J_n^\tau)B_{n,\tau}^{-1}\big\| > C\zeta^{-1}(n)\big)$$
$$\le C\zeta^{-2}(n),$$

(c) $$\sup_{s\in[s_*,s^*]}\sup_{f\in\mathcal{F}(s,a,\delta,K)}\sup_{\tau\in T_w\cap[s_*,s]} P_f\Bigg(\sup_{\theta\in\Theta}\big\|B_{n,\tau}^{-1}\big(H_{n,\tau}(d_f,\theta)-H_{n,\tau}(d_f,\theta_0)\big)B_{n,\tau}^{-1}\big\| > C\zeta^{-1}(n)\Bigg)\le C\zeta^{-2}(n), \quad \textit{and}$$

(d) $$\sup_{s\in[s_*,s^*]}\sup_{f\in\mathcal{F}(s,a,\delta,K)}\sup_{\tau\in T_w\cap[s_*,s]} P_f\Bigg(\sup_{d\in D_{m(\tau)}(\eta_n),\theta\in\Theta}\big\|B_{n,\tau}^{-1}\big(H_{n,\tau}(d,\theta)-H_{n,\tau}(d_f,\theta)\big)B_{n,\tau}^{-1}\big\| > C\zeta^{-1}(n)\Bigg)\le C\zeta^{-2}(n), \quad \textit{where} \quad \eta_n=\zeta^{-2}(n).$$

PROOF OF LEMMA 7(a): For $\tau\in T_w\cap[s_*,s^*]$, $B_{n,\tau}^{-1}J_n^\tau B_{n,\tau}^{-1}-\Omega_{r(\tau)}$ does not depend on τ and the proof of Lemma 2(a) applies. *Q.E.D.*

PROOF OF LEMMA 7(b): We employ the same definitions as in Proof of Lemma 2(b), but with $m=m(\tau)$ and $r=r(\tau)$. By (A.7), (A.8), $J_{a,b}=O(\log^{2a}(m(\tau))$, and some fairly standard manipulations, it suffices to show that the sets $\{|\widehat{G}_{a,b}-J_{a,b}|>C\zeta^{-1}(n)\log^{-2}m(\tau):n\ge1\}$ are uniformly $\zeta^{-2}(n)$ small for $a=0,1,2$, $b=0,1,\ldots,r(\tau)$. By Markov's inequality, the latter is implied by

(A.57) $$E_f|\widehat{G}_{a,b}(d_f,\theta_0)-J_{a,b}|^2\le 2G_0(E_f\vartheta_{1,a,m}^2+E_f\vartheta_{2,a,m}^2)=O\big(\zeta^{-4}(n)\log^{-4}m(\tau)\big)$$

uniformly over $s\in[s_*,s^*]$, $f\in\mathcal{F}(s,a,\delta,K)$, $\tau\in T_w\cap[s_*,s]$, for $a=0,1,2$. The inequality in (A.57) holds by (A.11).

To show the equality in (A.57) holds, we use the following results:

(A.58)(i) $$E_f\left(\sum_{j=1}^k\left(\frac{I_j}{g_j}-2\pi I_{\varepsilon j}\right)\right)^2=O(k^{2/3}\log^{4/3}k+k^{2+2\tau}n^{-2\tau}+kn^{-1/2}) \quad \text{and}$$

(A.58)(ii) $$E_f\left(\sum_{j=1}^k(2\pi I_{\varepsilon j}-1)\right)^2=O(k) \quad \text{as} \quad n\to\infty$$

uniformly over $k\in\{1,2,\ldots,m(\tau)\}$, s, f, and τ. Using (A.12), the Cauchy–Schwarz inequality, and (A.58), we obtain: for some $C<\infty$ independent of s, f, and τ,

(A.59) $$\begin{aligned}E_f\vartheta_{1,a,m}^2&\le C(\log^{2a}m)m^{-2}\left(\sum_{k=1}^{m-1}k^{-1}\right)\sum_{k=1}^{m-1}k^{-1}E\left(\sum_{j=1}^k\left(\frac{I_j}{g_j}-1\right)\right)^2\\&\le C(\log^{2a+1}m)m^{-2}\sum_{k=1}^{m-1}k^{-1}O(k+k^{2/3}\log^{4/3}k+k^{2+2\tau}n^{-2\tau}+kn^{-1/2})\\&\le C(\log^{2a+1}m)(m^{-1}+m^{2\tau}n^{-2\tau}),\end{aligned}$$

where $m=m(\tau)$. Using (A.12) and (A.58), we have

(A.60) $$\begin{aligned}E_f\vartheta_{2,a,m}^2&\le C(\log^{2a}m)m^{-2}O(m+m^{2/3}\log^{4/3}m+m^{2+2\tau}n^{-2\tau}+mn^{-1/2})\\&\le C(\log^{2a}m)(m^{-1}+m^{2\tau}n^{-2\tau})\end{aligned}$$

uniformly over s, f, and τ. The right-hand side in (A.59) and (A.60) is $o(\zeta^{-4}(n)\log^{-4}m)$, so the equality in (A.57) holds.

Finally, we prove (A.58). Part (ii) follows from Robinson's (1995a, p. 1647) proof of his (4.9). To prove part (i), we choose an integer $\ell \le k$ and write $E_f(\sum_{j=1}^k (I_j/g_j - 2\pi I_{\varepsilon j}))^2/3$ as

$$
\begin{aligned}
(A.61)\qquad & E_f\left(\sum_{j=1}^{\ell}\left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j}\right) + \sum_{j=\ell+1}^{k}\left(\frac{I_j}{g_j} - \frac{I_j}{f_j}\right) + \sum_{j=\ell+1}^{k}\left(\frac{I_j}{f_j} - 2\pi I_{\varepsilon j}\right)\right)^2 \Big/ 3 \\
& \le E_f\left(\sum_{j=1}^{\ell}\left(\frac{I_j}{g_j} - 2\pi I_{\varepsilon j}\right)\right)^2 + E_f\left(\sum_{j=\ell+1}^{k}\frac{I_j}{g_j}\left(1 - \frac{g_j}{f_j}\right)\right)^2 \\
& \quad + E_f\left(\sum_{j=\ell+1}^{k}\left(\frac{I_j}{f_j} - 2\pi I_{\varepsilon j}\right)\right)^2.
\end{aligned}
$$

The first expectation on the right-hand side of (A.61) divided by two is uniformly bounded by

$$
(A.62)\qquad E_f\left(\sum_{j=1}^{\ell}\frac{I_j}{g_j}\right)^2 + E_f\left(\sum_{j=1}^{\ell}2\pi I_{\varepsilon j}\right)^2 \le \ell\sum_{j=1}^{\ell}E_f(I_j/g_j)^2 + \ell\sum_{j=1}^{\ell}E_f(2\pi I_{\varepsilon j})^2 = O(\ell^2),
$$

where the last equality follows from the fact that $E_f(I_j/g_j)^2 = O(1)$ and $E_f(2\pi I_{\varepsilon j})^2 = O(1)$ uniformly over $j \in \{1, 2, \ldots, m(\tau)\}$, s, f, and τ. The second expectation on the right-hand side of (A.61) is uniformly bounded by

$$
(A.63)\qquad k\sum_{j=\ell+1}^{k}E_f\frac{I_j^2}{g_j^2}\left(1 - \frac{g_j}{f_j}\right)^2 = O\left(k\sum_{j=\ell+1}^{k}\lambda_j^{2\tau}\right) = O\big(k^2(k/n)^{2\tau}\big),
$$

where the first equality uses $E_f(I_j/g_j)^2 = O(1)$ and a Taylor expansion of $\log\varphi(\lambda)$ to order $2r(\tau)$. The third expectation on the right-hand side of (A.61) is uniformly bounded by

$$
(A.64)\qquad O\big(\log^2 k + (k\log^2 k)/\ell + k^{1/2}\log k + kn^{-1/2}\big),
$$

which follows from Robinson's (1995a) proof on pp. 1648–1651. Setting $\ell = k^{1/3}\log^{2/3} k$, (A.61)–(A.64) combine to give part (i) of (A.58). *Q.E.D.*

PROOF OF LEMMA 7(c): As in the proof of Lemma 7(b), it suffices to show that

$$
(A.65)\qquad E_f\sup_{\theta\in\Theta}\big(\widehat{G}_{a,b}(d_f, \theta_0) - \widehat{G}_{a,b}(d_f, \theta)\big)^2 = O(\zeta^{-4}(n)\log^{-4} m(\tau))
$$

uniformly over s, f, and τ, for all $a = 0, 1, 2$ and $b = 0, \ldots, r(\tau)$. Using (A.15) and the definition of $\widehat{G}_{a,0}(d_f, \theta_0)$ in (A.6), the left-hand side of (A.65) is bounded by

$$
(A.66)\qquad O(\lambda_{m(\tau)}^2)E_f\widehat{G}_{a,0}^2(d_f, \theta_0) = O\big((m(\tau)/n)^2(\log^{2a} m(\tau))\big),
$$

where the equality holds by (A.57) and $J_{a,0} = O(\log^{2a} m(\tau))$. The right-hand side of (A.66) is $O(\zeta^{-4}(n)\log^{-4} m(\tau))$ by the definitions of $m(\tau)$ and $\zeta(n)$. *Q.E.D.*

PROOF OF LEMMA 7(d): As in the proof of part (c) above, it suffices to show that the expectation of the square of the left-hand side of (A.16) is $O(\zeta^{-4}(n)\log^{-4} m(\tau))$ uniformly over s, f, and τ. As in (A.17), this is implied by

$$
(A.67)\qquad E_f Z_{a,b}^2(\eta_n) = O\big(n^{4d_f}\zeta^{-4}(n)\log^{-4} m(\tau)\big)
$$

uniformly over s, f, and τ for all $a = 0, 1, 2$, and $b = 0, \ldots, r(\tau)$, where $Z_{a,b}(\eta_n)$ is defined in (A.17). Equation (A.67) holds because (A.18) and $\widehat{G}_{0,0}(d_f, 0) = O(1)$ (by (A.57) and (A.65)) imply that $E_f Z_{a,b}^2(\eta_n) = O(\eta_n^2(\log^{-4} m(\tau))n^{4d_f})$. *Q.E.D.*

We now use the results of Lemma 7 to verify conditions (iii) and (iv) of Lemma 6 with $\mathcal{G} = \mathcal{G}_w$ for each $w \in \mathcal{W}$. Condition (iii) holds by Lemma 7(a) and (b) and the positive definiteness of $\Omega_{r(\tau)}$. Condition (iv) holds with $K_{n,\tau} = m(\tau)^{1/2}\eta_n \log^{-5} m(\tau)$, where $\eta_n = \zeta^{-2}(n)$, by Lemma 7(c) and (d). In consequence, the following lemma holds.

LEMMA 8: *Suppose the assumptions of Theorem* 4 *hold. Let* $s_* \geq 1$. *Then, for each* $s^* \in [s_*, \infty)$, *there exist solutions* $(\widetilde{d}_\tau, \widetilde{\theta}_\tau)$ *to the* FOC's $(\partial/\partial(d, \theta')')\, R_{r(\tau),\tau}(d, \theta) = 0$ *except on sets that are uniformly* $\zeta^{-2}(n)$ *small and*

$$B_{n,\tau}\begin{pmatrix} \widetilde{d}_\tau - d_f \\ \widetilde{\theta}_\tau - \theta_0 \end{pmatrix} = -\Omega_{r(\tau)}^{-1} B_{n,\tau}^{-1} S_{n,\tau}(d_f, \theta_0) + \varepsilon_{n,\tau},$$

where the sets $\{\{\|\varepsilon_{n,\tau}\| > C_*\zeta(n)\} : n \geq 1\}$ *are uniformly* $\zeta^{-2}(n)$ *small for all* $C_* < \infty$.

Using strict convexity of $R_r(d, \theta)$, the solutions $(\widetilde{d}_\tau, \widetilde{\theta}_\tau)$ to the FOC's equal the LPW estimators $(\widehat{d}_\tau, \widehat{\theta}_\tau)$ except on sets that are uniformly $\zeta^{-2}(n)$ small (by the argument given at the beginning of Section 5). Hence, Lemma 8 implies Lemma 4.

It remains to prove Lemma 5.

PROOF OF LEMMA 5: For convenience, we show that: for some $C < \infty$, the result of Lemma 5 holds with C_* replaced by C. This is sufficient because the proof of the latter holds without change with $\zeta(n)$ replaced by $\widetilde{C}\zeta(n)$ for all constants $\widetilde{C} > 0$.

By (A.19) and (A.20),

$$B_{n,\tau}^{-1} S_{n,\tau}(d_f, \theta_0) = \widehat{G}^{-1}(d_f, \theta_0) \sum_{k=1}^{4} T_{k,n,\tau}, \tag{A.68}$$

where $T_{k,n,\tau}$ denotes $T_{k,n}$, defined in (A.20), with $m = m(\tau)$ and $r = r(\tau)$ for $k = 1, 2, 3, 4$ and $\widehat{G}(d_f, \theta_0)$ is defined in (2.3) with $m = m(\tau)$. By definition, $\widehat{G}(d_f, \theta_0) = \widehat{G}_{0,0}(d_f, \theta_0)$. By (A.57) and Markov's inequality, for some $C < \infty$,

$$P_f\big(|\widehat{G}_{0,0}(d_f, \theta_0) - J_{0,0}| > C\zeta(n)\big) \leq C\zeta^{-2}(n), \tag{A.69}$$

where the inequality holds for all $\tau \leq s$, $f \in \mathcal{F}(s, a, \delta, K)$, and $s \in [s_*, s^*]$. Note that $J_{0,0} = G_0 > 0$. In consequence, it suffices to show that for some $C < \infty$,

$$P_f(\|T_{k,n,\tau}\| > C\zeta(n)) \leq C\zeta^{-2}(n) \quad \text{for} \quad k = 1, 2, 3, 4, \tag{A.70}$$

where the inequality holds for all $\tau \leq s$, $f \in \mathcal{F}(s, a, \delta, K)$, and $s \in [s_*, s^*]$.

For $k = 2$ and $k = 4$, (A.70) holds because

$$T_{2,n,\tau} = o(1) \quad \text{and} \quad T_{4,n,\tau} = O(1) \tag{A.71}$$

uniformly over $\tau \leq s$, $f \in \mathcal{F}(s, a, \delta, K)$, and $s \in [s_*, s^*]$, where the former result holds by (A.23) and (A.24) and the latter holds by (A.25)–(A.27).

To establish (A.70) for $k = 3$, let $h_j(\tau) = \widetilde{X}_j - m^{-1}(\tau)\sum_{k=1}^{m(\tau)} \widetilde{X}_k$. Then, by Markov's inequality, we have

$$\begin{aligned} P_f(\|T_{3,n,\tau}\| \geq C\zeta(n)) &= P_f\left(\left\|m^{-1/2}(\tau)\sum_{j=1}^{m(\tau)}(2\pi I_{\varepsilon j} - 1)h_j(\tau)\right\| \geq C\zeta(n)\right) \\ &\leq C^{-2}\zeta^{-2}(n)m^{-1}(\tau)E\left\|\sum_{j=1}^{m(\tau)}(2\pi I_{\varepsilon j} - 1)h_j(\tau)\right\|^2. \end{aligned} \tag{A.72}$$

By the second last paragraph of the proof of Lemma 2(e), $\mathrm{var}(I_{\varepsilon j}) = O(1)$ and $\mathrm{cov}(I_{\varepsilon j}, I_{\varepsilon k}) = O(n^{-1})$ uniformly over $j, k = 1, \ldots, n$ with $j \neq k$. Also, by Lemma 2(a) of AG, $m^{-1}(\tau) \sum_{j=1}^{m(\tau)} h_j(\tau)' h_j(\tau) = O(1)$ uniformly over $\tau \leq s$. In consequence,

$$
\begin{aligned}
(\text{A.73}) \qquad & E \left\| \sum_{j=1}^{m(\tau)} (2\pi I_{\varepsilon j} - 1) h_j(\tau) \right\|^2 \\
& = (2\pi)^2 \sum_{j=1}^{m(\tau)} \mathrm{var}(I_{\varepsilon j}) h_j(\tau)' h_j(\tau) + 8\pi^2 \sum_{j,k=1, j \neq k}^{m(\tau)} \mathrm{cov}(I_{\varepsilon j}, I_{\varepsilon k}) h_j(\tau)' h_k(\tau) \\
& = O(m(\tau))
\end{aligned}
$$

uniformly over $\tau \leq s$, $f \in \mathcal{F}(s, a, \delta, K)$, and $s \in [s_*, s^*]$. Combining (A.72) and (A.73) establishes (A.70) for $k = 3$.

For $k = 1$, (A.70) holds using (A.21)–(A.22) and the proof of (A.21) given in (A.28)–(A.31). In particular, the moment bounds in (A.30) and (A.31) are used to bound the tail probabilities of interest using Markov's inequality. *Q.E.D.*

REFERENCES

AGIAKLOGLOU, C., P. NEWBOLD, AND M. WOHAR (1993): "Bias in an Estimator of the Fractional Difference Parameter," *Journal of Time Series Analysis*, 14, 235–246.

ANDREWS, D. W. K., AND P. GUGGENBERGER (2003): "A Bias-Reduced Log-Periodogram Regression Estimator for the Long-Memory Parameter," *Econometrica*, 71, 675–712.

ANDREWS, D. W. K., AND Y. SUN (2001): "Local Polynomial Whittle Estimation of Long-Range Dependence," Cowles Foundation Discussion Paper No. 1293, Yale University. Available at http://cowles.econ.yale.edu.

BERAN, J. (1993): "Fitting Long-Memory Models by Generalized Linear Regression," *Biometrika*, 80, 817–822.

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

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: Oxford University Press.

FAN, J. (1992): "Design-Adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998–1004.

FOX, R., AND M. S. TAQQU (1986): "Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series," *Annals of Statistics*, 14, 517–532.

GEWEKE, J., AND S. PORTER-HUDAK (1983): "The Estimation and Application of Long-Memory Time Series Models," *Journal of Time Series Analysis,* 4, 221–237.

GIL-ALAÑA, L. A., AND P. M. ROBINSON (1997): "Testing of Unit Root and Other Nonstationary Hypotheses in Macroeconomic Time Series," *Journal of Econometrics*, 80, 241–268.

GIRAITIS, L., P. M. ROBINSON, AND A. SAMAROV (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.

——— (2000): "Adaptive Semiparametric Estimation of the Memory Parameter," *Journal of Multivariate Analysis*, 72, 183–207.

HÄRDLE, W., AND O. LINTON (1994): "Applied Nonparametric Methods," in *Handbook of Econometrics*, Vol. 4, ed. by R. F. Engle and D. McFadden. New York: Elsevier, Ch. 38.

HEIJMANS, R. D. H., AND J. R. MAGNUS (1986): "On the First-Order Efficiency and Asymptotic Normality of Maximum Likelihood Estimators Obtained from Dependent Observations," *Statistica Nederlandica*, 40, 169–188.

HENRY, M., AND P. M. ROBINSON (1996): "Bandwidth Choice in Gaussian Semiparametric Estimation of Long Range Dependence," in *Athens Conference on Applied Probability and Time Series*, Vol. II: *Time Series in Memory of E. J. Hannan*, ed. by P. M. Robinson and M. Rosenblatt. New York: Springer-Verlag.

HURVICH, C. M. (2001): "Model Selection for Broadband Semiparametric Estimation of Long Memory in Time Series," *Journal of Time Series Analysis*, 22, 679–709.

HURVICH, C. M., AND R. S. DEO (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.

HURVICH, C. M., E. MOULINES, AND P. SOULIER (2002): "The FEXP Estimator for Potentially Non-stationary Linear Time Series," *Stochastic Processes and Their Applications*, 97, 307–340.

IOUDITSKY, A., E. MOULINES, AND P. SOULIER (2002): "Adaptive Estimation of the Fractional Differencing Coefficient," *Bernoulli*, 7, 699–731.

KÜNSCH, H. R. (1987): "Statistical Aspects of Self-Similar Processes," in *Proceedings of the First World Congress of the Bernoulli Society*, 1, ed. by Yu. Prohorov and V. V. Sazanov. Utrecht: VNU Science Press, pp. 67–74.

LEPSKII, O. V. (1990): "On a Problem of Adaptive Estimation in Gaussian White Noise," *Theory of Probability and Its Applications*, 35, 454–466.

MARINUCCI, D., AND P. M. ROBINSON (2001): "Semiparametric Fractional Cointegration Analysis," *Journal of Econometrics*, 105, 225–247.

ROBINSON, P. M. (1995a): "Gaussian Semiparametric Estimation of Long Range Dependence," *Annals of Statistics,* 23, 1630–1661.

——— (1995b): "Log-Periodogram Regression of Time Series with Long Range Dependence," *Annals of Statistics,* 23, 1048–1072.

ROBINSON, P. M., AND M. HENRY (1999): "Long and Short Memory Conditional Heteroskedasticity in Estimating the Memory Parameter of Levels," *Econometric Theory*, 15, 299–336.

——— (2003): "Higher-Order Kernel Semiparametric M-estimation of Long Memory," *Journal of Econometrics*, 114, 1–27.

SHIMOTSU, K., AND P. C. B. PHILLIPS (2002): "Exact Local Whittle Estimation of Fractional Integration," Cowles Foundation Discussion Paper No. 1367, Yale University. Available at http://cowles.econ.yale.edu.

VELASCO, C. (1999): "Non-Stationary Log-Periodogram Regression," *Journal of Econometrics*, 91, 325–371.

VELASCO, C., AND P. M. ROBINSON (2000): "Whittle Pseudo-Maximum Likelihood Estimation for Nonstationary Time Series," *Journal of the American Statistical Association*, 95, 1229–1243.

WEISS, L. (1971): "Asymptotic Properties of Maximum Likelihood Estimators in Some Nonstandard Cases I," *Journal of the American Statistical Association*, 66, 345–350.

——— (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, ed. by R. F. Engle and D. L. McFadden. Amsterdam: North-Holland.