A BIAS-REDUCED LOG-PERIODOGRAM REGRESSION ESTIMATOR FOR THE LONG-MEMORY PARAMETER

BY

DONALD W. K. ANDREWS and PATRIK GUGGENBERGER

COWLES FOUNDATION PAPER NO. 1051

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

2003

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

Econometrica, Vol. 71, No. 2 (March, 2003), 675–712

A BIAS-REDUCED LOG-PERIODOGRAM REGRESSION ESTIMATOR FOR THE LONG-MEMORY PARAMETER

BY DONALD W. K. ANDREWS AND PATRIK GUGGENBERGER[1]

In this paper, we propose a simple bias-reduced log-periodogram regression estimator, $\hat{d}_r$, of the long-memory parameter, d, that eliminates the first- and higher-order biases of the Geweke and Porter-Hudak (1983) (GPH) estimator. The bias-reduced estimator is the same as the GPH estimator except that one includes frequencies to the power $2k$ for $k = 1, \ldots, r$, for some positive integer r, as additional regressors in the pseudo-regression model that yields the GPH estimator. The reduction in bias is obtained using assumptions on the spectrum only in a neighborhood of the zero frequency.

Following the work of Robinson (1995b) and Hurvich, Deo, and Brodsky (1998), we establish the asymptotic bias, variance, and mean-squared error (MSE) of $\hat{d}_r$, determine the asymptotic MSE optimal choice of the number of frequencies, m, to include in the regression, and establish the asymptotic normality of $\hat{d}_r$. These results show that the bias of $\hat{d}_r$ goes to zero at a faster rate than that of the GPH estimator when the normalized spectrum at zero is sufficiently smooth, but that its variance only is increased by a multiplicative constant.

We show that the bias-reduced estimator $\hat{d}_r$ attains the optimal rate of convergence for a class of spectral densities that includes those that are smooth of order $s \geq 1$ at zero when $r \geq (s-2)/2$ and m is chosen appropriately. For $s > 2$, the GPH estimator does not attain this rate. The proof uses results of Giraitis, Robinson, and Samarov (1997).

We specify a data-dependent plug-in method for selecting the number of frequencies m to minimize asymptotic MSE for a given value of r.

Some Monte Carlo simulation results for stationary Gaussian ARFIMA$(1, d, 1)$ and $(2, d, 0)$ models show that the bias-reduced estimators perform well relative to the standard log-periodogram regression estimator.

KEYWORDS: Asymptotic bias, asymptotic normality, bias reduction, frequency domain, long-range dependence, optimal rate, plug-in estimator, rate of convergence, strongly dependent time series.

1. INTRODUCTION

WE CONSIDER A SEMIPARAMETRIC model for a stationary Gaussian long-memory time series $\{Y_t : t = 1, \ldots, n\}$. The spectral density of the time series is given by

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

where $d \in (-0.5, 0.5)$ is the long-memory parameter, $g(\cdot)$ is an even function on $[-\pi, \pi]$ that is continuous at zero with $0 < g(0) < \infty$, and $f(\lambda)$ is integrable over

[1] The first author gratefully acknowledges the research support of the National Science Foundation via Grant Number SBR-9730277. The second author thanks the Cowles Foundation for research support via a Cowles Prize. The authors thank Marc Henry, Peter Robinson, Alain Monfort, and three anonymous referees for helpful comments.

$(-\pi, \pi)$. The parameter d determines the low frequency properties of the series. When $d > 0$, the series exhibits long memory. The function $g(\cdot)$ determines the high frequency properties of the series, i.e., its short-term correlation structure.

The widely used log-periodogram regression estimator of the long-memory parameter d proposed by Geweke and Porter-Hudak (1983) (GPH) has been criticized because of its finite-sample bias (see Agiakloglou, Newbold, and Wohar (1993)). In this paper, we investigate the asymptotic and finite sample properties of a new bias-reduced log-periodogram estimator $\hat{d}_r$ of the long memory parameter d. Let $\lambda_j = 2\pi j/n$ for $j = 1, \ldots, [n/2]$ denote the fundamental frequencies for a sample of size n. The estimator $\hat{d}_r$ is defined to be the least squares (LS) estimator of the coefficient on $-2\log\lambda_j$ in a regression of the log of the periodogram evaluated at λ_j on a constant, $-2\log\lambda_j$, and $\lambda_j^2, \lambda_j^4, \ldots, \lambda_j^{2r}$ for $j = 1, \ldots, m$, where r is a (fixed) nonnegative integer. We take m such that $m \to \infty$ and $m/n \to 0$ as $n \to \infty$. When $r = 0$, $\hat{d}_r$ is asymptotically equivalent to the well-known GPH estimator $\hat{d}_{GPH}$.

The motivation for the estimator $\hat{d}_r$ is the local polynomial estimator for nonparametric regression functions (see Fan (1992) and additional references in Härdle and Linton (1994)). The latter is a popular nonparametric estimation method that is found to perform well for low order polynomials, such as linear or quadratic polynomials. Analogously, we expect the bias-reduced log-periodogram estimator to perform well for small values of r, such as $r = 1$ or $r = 2$. That is, although the asymptotic results established below hold for arbitrary large values of r, we do not recommend using large values in practice because the asymptotic properties will not be reflected in finite samples.

We determine the asymptotic bias, variance, and MSE of $\hat{d}_r$, calculate the MSE optimal choice of m for $\hat{d}_r$, and establish the asymptotic normality of $\hat{d}_r$. The proofs of these results rely heavily on results of Robinson (1995b) and Hurvich, Deo, and Brodsky (1998) (HDB). We find that the asymptotic bias of $\hat{d}_r$ is of order m^{2+2r}/n^{2+2r} provided g is sufficiently smooth, whereas that of $\hat{d}_{GPH}$ is of the larger order m^2/n^2. The asymptotic variances of $\hat{d}_r$ and $\hat{d}_{GPH}$ are both of order m^{-1}. In consequence, the optimal rate of convergence to zero of the MSE of $\hat{d}_r$ is of order $n^{-(4+4r)/(5+4r)}$ whereas that of $\hat{d}_{GPH}$ is of the larger order $n^{-4/5}$. For example, for $r = 2$, this is $n^{-12/13}$. The rate of convergence of $\hat{d}_r$ also exceeds that of the local Whittle estimator (see Robinson (1995a)) and the average periodogram estimator (see Robinson (1994)), provided g is sufficiently smooth.

We find that $m^{1/2}(\hat{d}_r - d)$ is asymptotically normal with mean zero provided $m = o(n^{(4+4r)/(5+4r)})$. In contrast, $\hat{d}_{GPH}$ is asymptotically normal only under the more stringent condition $m = o(n^{4/5})$.

We determine the optimal rate of convergence of a minimax risk criterion for estimators of d when the true normalized spectral density lies in a class that includes densities that are smooth to order s at zero for some $s \geq 1$. The optimal rate is $n^{-s/(2s+1)}$. The estimator $\hat{d}_r$ is shown to achieve this rate of convergence provided $r \geq (s-2)/2$ and m is chosen suitably. In contrast, when $s > 2$, the GPH estimator does not achieve this rate. The proof of the optimal rate results utilizes results of Giraitis, Robinson, and Samarov (1997) (GRS).

We provide a consistent estimator of the only unknown constant in the formula for the MSE optimal choice of m for a given value of r. This yields a data-dependent plug-in method for choosing m. This procedure does not, however, achieve the optimal MSE rate of convergence. This occurs because the smoothness conditions imposed to yield consistency of the estimator of the unknown constant imply that taking a larger value of r is needed to obtain the optimal MSE rate of convergence. We do not find this troubling because we feel that using a small value of r is preferable in practice in terms of finite sample performance.

Some Monte Carlo simulations show that the bias-reduced estimators $\hat{d}_1$ and $\hat{d}_2$ have lower biases, higher standard deviations, and slightly lower root mean-squared errors (RMSE's) compared to the standard log-periodogram estimator $\hat{d}_0$ (which is a slight variant of $\hat{d}_{GPH}$) for a variety of stationary Gaussian ARFIMA$(1, d, 1)$ and $(2, d, 0)$ processes, as the asymptotic results suggest. (Details of the ARFIMA$(2, d, 0)$ results are not reported in the paper, but are available from the authors.) The lower biases lead to good confidence interval (CI) coverage probabilities for CI's based on $\hat{d}_1$ and $\hat{d}_2$ over a wider range of m values than for $\hat{d}_0$. On the other hand, the lower standard deviation of $\hat{d}_0$ leads to shorter CI intervals than CI's based on $\hat{d}_1$ and $\hat{d}_2$. The RMSE graphs for $\hat{d}_1$ and $\hat{d}_2$ are flatter as a function of m than those for $\hat{d}_0$, which implies that $\hat{d}_1$ and $\hat{d}_2$ are less sensitive to the choice of m than is $\hat{d}_0$. This corroborates asymptotic results which show that the slope of the RMSE function, as a function of m, declines to zero faster when $r > 0$ than when $r = 0$.

For all three estimators, $\hat{d}_0$, $\hat{d}_1$, and $\hat{d}_2$, performance depends primarily on the value of the autoregressive coefficient in ARFIMA$(1, d, 1)$ processes and the sum of the autoregressive coefficients in ARFIMA$(2, d, 0)$ processes. When these are close to one, biases and RMSE's are high and CI coverage probabilities are low.

The simulation results are virtually the same for the three values of d considered: $-.4$, 0, and .4. The basic pattern of results (in terms of the shapes of the bias, standard deviation, RMSE, and coverage probability graphs as functions of m) are the same for sample sizes $n = 128, 512$, and 2048. However, the ratio of the minimum RMSE over $m \in [1, n/2]$ for $\hat{d}_1$ or $\hat{d}_2$ to that of $\hat{d}_0$ decreases as n increases. This is in accord with the asymptotic theory.

Simulation results utilizing the plug-in choice of m show that the estimators $\hat{d}_1$ and $\hat{d}_2$ exhibit reduced bias, increased standard deviations, and for some parameter combinations reduced RMSE compared to $\hat{d}_0$. The estimators $\hat{d}_1$ and $\hat{d}_2$ deliver CI's with much better coverage probabilities than the estimator $\hat{d}_0$. The estimation error in the plug-in method causes a substantial increase in the RMSE of all three estimators $\hat{d}_0$, $\hat{d}_1$, and $\hat{d}_2$ compared to the (infeasible) case in which the unknown in the formula is known. This is to be expected, because the unknown in the plug-in formula is a higher-order derivative that must be estimated nonparametrically.

In sum, the simulations indicate that for stationary Gaussian ARFIMA$(1, d, 1)$ and $(2, d, 0)$ processes the estimators $\hat{d}_1$ and $\hat{d}_2$ usually deliver bias reductions, small RMSE reductions, and improved CI coverage probabilities in finite samples that reflect the asymptotic results.

A closely related paper to this one is Robinson and Henry (2003) (RH). RH introduces a class of narrow-band semiparametric estimators that utilize higher-order kernels to obtain bias-reduction. In contrast to the results given here, the asymptotic results in RH are determined using formal expansions and are not given rigorous proofs.

Other related papers include Hurvich and Brodsky (2001), Bhansali and Kokoszka (1997), Moulines and Soulier (1999), and Hurvich (2001). Each of these papers considers semiparametric estimation of d by specifying a parametric model and letting the number of parameters in the model increase with the sample size. These estimators of d, like $\hat{d}_r$, attain rates of convergence that exceed the rate $n^{2/5}$ of the GPH estimator. These estimators differ from the bias-reduced estimator considered here in that they are broad-band estimators that use all of the frequencies in the range $[0, \pi]$. Correspondingly, they rely on assumptions on the spectrum over the whole interval $[0, \pi]$. In contrast, the bias-reduced estimator considered here is a narrow-band estimator. It only relies on assumptions on the spectrum at the origin.

Still other related papers include Delgado and Robinson (1996), Henry and Robinson (1996), Hurvich and Deo (1999), and Henry (2001). Each of these papers considers a regression of the periodogram or the log-periodogram on several regressors including the squared frequency. The results of these regressions are used to obtain data-dependent bandwidth choices for the GPH, local Whittle, and average periodogram estimators. These papers do not consider bias-reduced estimation of d based on these regressions.

The bias-reduction method utilized here can be extended to a number of other procedures. Andrews and Sun (2001) consider a bias-reduced local polynomial Whittle estimator. The bias-reduction method also could be applied to the pooled and/or multivariate log-periodogram regression estimators of Robinson (1995b), the pooled and/or tapered log-periodogram regression estimators for stationary non-Gaussian series analyzed by Velasco (2000), the modified log-periodogram estimator of Kim and Phillips (1999b) for nonstationary time series, the tapered log-periodogram estimator of Velasco (1999) for nonstationary time series, and the adaptive log-periodogram regression estimator of Giraitis, Robinson, and Samarov (2000). In addition, one could analyze the properties of the bias-reduced log-periodogram estimator with nonstationary time series, along the lines of Kim and Phillips (1999a).

The remainder of this paper is organized as follows. Section 2 motivates the bias-reduced estimator by reviewing results for the GPH estimator. Section 3 establishes the asymptotic bias, variance, and MSE of the bias-reduced log-periodogram estimator, and shows that it is asymptotically normal. Section 4 gives the optimal rate of convergence results. Section 5 provides results for a plug-in method for choosing the bandwidth. Section 6 describes the simulation results. An Appendix provides proofs.

2. THE GPH ESTIMATOR

An alternative parameterization of the model in (1.1) that is often used in the literature (e.g., see HDB) is

$$(2.1) \qquad f(\lambda) = |1-\exp(-i\lambda)|^{-2d} f^*(\lambda),$$

where $f^*(\cdot)$ satisfies the same conditions as $g(\cdot)$. The models in (1.1) and (2.1) are equivalent because $|1-\exp(-i\lambda)|^{-2d} = |\lambda|^{-2d}(1+o(1))$ as $\lambda \to 0$.

Using the parameterization in (2.1), GPH proposed an estimator of d based on the first m periodogram ordinates

$$(2.2) \qquad I_j = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} Y_t \exp(i\lambda_j t)\right|^2 \quad \text{for} \quad j = 1, \ldots, m,$$

where $\lambda_j = 2\pi j/n$ and m is a positive integer smaller than n. The GPH estimator is given by $-1/2$ times the LS estimator of the slope parameter in a regression of $\{\log I_j : j = 1, \ldots, m\}$ on a constant and the regressor variable $\widetilde{X}_j = \log|1-\exp(-i\lambda_j)| = (1/2)\log(2-2\cos\lambda_j)$. By definition, the GPH estimator is

$$(2.3) \qquad \hat{d}_{GPH} = \frac{-0.5\sum_{j=1}^{m}(\widetilde{X}_j - \overline{\widetilde{X}})\log I_j}{\sum_{j=1}^{m}(\widetilde{X}_j - \overline{\widetilde{X}})^2},$$

where $\overline{\widetilde{X}} = (1/m)\sum_{j=1}^{m}\widetilde{X}_j$.

This estimator can be motivated heuristically using model (2.1) by writing

$$(2.4) \qquad \log I_j = (\log f_0^* - C) - 2d\widetilde{X}_j + \log(f_j^*/f_0^*) + \varepsilon_j,$$

where $\varepsilon_j = \log(I_j/f_j) + C, f_j = f(\lambda_j), f_j^* = f^*(\lambda_j), f_0^* = f^*(0)$, and $C = 0.577216\ldots$ is the Euler constant. Equation (2.4) is a pseudo-regression model. If the pseudo-errors $\{\log(f_j^*/f_0^*) + \varepsilon_j : j = 1, \ldots, m\}$ behave like iid random variables, then the regression estimator $\hat{d}_{GPH}$ is a reasonable estimation procedure.

In fact, Robinson (1995b) shows that a variant of $\hat{d}_{GPH}$, which trims out small values of j from the regression, is consistent and asymptotically normal provided $m \to \infty$ as $n \to \infty$ at a rate that is not too quick. Robinson's (1995b) estimator also differs from the GPH estimator in that he uses the model parameterization in (1.1) and, hence, replaces the regressor $\widetilde{X}_j$ by

$$(2.5) \qquad X_j = -2\log\lambda_j$$

(and correspondingly drops the -0.5 term from the definition of $\hat{d}_{GPH}$). The use of X_j rather than $\widetilde{X}_j$ has no effect on the asymptotic bias, variance, MSE, or normality of the estimator. The form of the regressor $\widetilde{X}_j = \log|1-\exp(-i\lambda_j)|$ used by GPH comes from the spectrum of a fractionally differenced time series.

Since $|1-\exp(-i\lambda)|^{-2d} = |\lambda|^{-2d}(1+o(1))$ as $\lambda \to 0$ and the GPH estimator is a consistent estimator of d for a more general class of time series models than fractionally differenced time series, the simpler form for the regressor given in (2.5) is appropriate.

HDB provide further justification for log-periodogram regression estimators. They consider the GPH estimator $\hat{d}_{GPH}$ exactly as defined in (2.3). They establish the asymptotic bias, variance, and MSE of $\hat{d}_{GPH}$, calculate the MSE optimal choice of m, and establish the asymptotic normality (with mean zero) of $\hat{d}_{GPH}$ when $m \to \infty$ at a rate slower than the MSE optimal rate. In addition, it is straightforward to see that their results continue to hold with the regerssor $\widetilde{X}_j$ replaced by X_j.

Using the parameterization of (2.1), HDB suppose that m and f^* satisfy the following assumptions:

ASSUMPTION HDB1: $m = m(n) \to \infty$ *and* $(m(\log m)/n) \to 0$ *as* $n \to \infty$.

ASSUMPTION HDB2: f^* *is three times continuously differentiable in a neighborhood of zero and* $f^{*\prime}(0) = 0$.[2]

Under these assumptions, HDB establish that

$$
\begin{aligned}
(2.6) \qquad & E\hat{d}_{GPH} - d = \frac{-2\pi^2}{9}\frac{f^{*\prime\prime}(0)}{f^*(0)}\frac{m^2}{n^2} + o\left(\frac{m^2}{n^2}\right) + O\left(\frac{\log^3 m}{m}\right), \\
& \mathrm{var}(\hat{d}_{GPH}) = \frac{\pi^2}{24m} + o\left(\frac{1}{m}\right), \quad \text{and} \\
& \mathrm{MSE}(\hat{d}_{GPH}) = E(\hat{d}_{GPH} - d)^2 = \frac{4\pi^4}{81}\left(\frac{f^{*\prime\prime}(0)}{f^*(0)}\right)^2 \frac{m^4}{n^4} + \frac{\pi^2}{24m} + o\left(\frac{m^4}{n^4}\right) \\
& \qquad + O\left(\frac{m\log^3 m}{n^2}\right) + o\left(\frac{1}{m}\right).
\end{aligned}
$$

HDB point out that the choice of m that minimizes $\mathrm{MSE}(\hat{d}_{GPH})$ is given by

$$
(2.7) \qquad m_{GPH,opt} = \left(\frac{27}{128\pi^2}\right)^{\frac{1}{5}} \left(\frac{f^*(0)}{f^{*\prime\prime}(0)}\right)^{\frac{2}{5}} n^{\frac{4}{5}},
$$

provided $f^{*\prime\prime}(0) \neq 0$. With this choice of m, the MSE of $\hat{d}_{GPH}$ is of order $O(n^{-4/5})$.

[2] HDB do not actually assume that $f^{*(3)}$ is continuous, but they use this assumption when taking a three term Taylor expansion of $\log f_j^*$ in the proof of their Lemma 1. HDB also assume that f^* is continuous and bounded away from zero and infinity on $[-\pi, \pi]$, but these assumptions are not used in their proofs.

The dominant bias term of $\hat{d}_{GPH}$ in (2.6) comes from the term $\log(f_j^*/f_0^*)$, rather than the $E\varepsilon_j$ term, in the pseudo-regression model (2.4). Under Assumption HDB2, a Taylor series expansion gives

$$\log(f_j^*/f_0^*) = \frac{\lambda_j^2}{2}\frac{f^{*\prime\prime}(0)}{f^*(0)} + O(\lambda_j^3). \tag{2.8}$$

It is the first term on the right-hand side of (2.8) that is responsible for the dominant bias term of $\hat{d}_{GPH}$. This suggests that the elimination of this term will yield an estimator with reduced bias. This term can be eliminated by adding the regressor λ_j^2 to the pseudo-regression model (2.4). Furthermore, additional bias terms can be eliminated by adding the regressors $\lambda_j^4, \ldots, \lambda_j^{2r}$ for some $r \geq 2$. This is established rigorously in the next section.

3. BIAS-REDUCED LOG-PERIODOGRAM REGRESSION

3.1. *Asymptotic Bias and Variance*

In this section, we define the bias-reduced estimator $\hat{d}_r$, calculate its asymptotic bias and variance, and provide conditions under which it is asymptotically normal. We assume throughout that the model is given by (1.1). Thus, we utilize the regressor X_j, as in Robinson (1995b), rather than $\widetilde{X}_j$.

The bias-reduced estimator $\hat{d}_r$ is the LS estimator of the coefficient on X_j from the regression of $\log I_j$ on $1, X_j, \lambda_j^2, \lambda_j^4, \ldots, \lambda_j^{2r}$ for $j = 1, \ldots, m$ for some nonnegative integer r. It is defined explicitly in (3.8) below. Note that only even powers of λ_j are employed in the regression. Odd powers of λ_j do not help in reducing the asymptotic bias because they have coefficients equal to zero in the Taylor expansion of $\log(g(\lambda_j)/g(0))$, which determines the asymptotic bias of $\hat{d}_r$, as in (2.8). (These coefficients are zero due to the oddness of the odd order derivatives of $\log g(\lambda)$ and their continuity at zero.)

We assume that g is smooth of order s at zero for some $s \geq 1$, which is defined as follows. Let $[s]$ denote the integer part of s. We say that a real function h defined on a neighborhood of zero is smooth of order $s > 0$ at zero if h is $[s]$ times continuously differentiable in some neighborhood of zero and its derivative of order $[s]$, denoted $h^{([s])}$, satisfies a Hölder condition of order $s - [s]$ at zero, i.e., $|h^{([s])}(\lambda) - h^{([s])}(0)| \leq C|\lambda|^{s-[s]}$ for some constant $C < \infty$ and all λ in a neighborhood of zero.

We use the following assumptions:

ASSUMPTION 1: *$m = m(n) \to \infty$ and $m/n \to 0$ as $n \to \infty$.*

ASSUMPTION 2: *g is an even function on $[-\pi, \pi]$ that is smooth of order s at zero for some $s \geq 1$, $0 < g(0) < \infty$, $-1/2 < d < 1/2$, and $\int_{-\pi}^{\pi} |\lambda|^{-2d} g(\lambda)\, d\lambda < \infty$.*

For example, Assumption 2 holds for ARFIMA(p, d, q) processes and fractional Gaussian noise for all s finite.

Assumption 2 allows one to develop an $[s]$ term Taylor expansion of $\log g(\lambda_j)$ about $\lambda = 0$:[3]

$$(3.1) \qquad \log(g_j/g_0) = \sum_{k=1}^{[s]} \frac{b_k}{k!}\lambda_j^k + Rem_j, \quad \text{where}$$

$$\max_{1\le j\le m} |Rem_j/\lambda_j^s| = O(1) \quad \text{as} \quad n\to\infty,$$

$$g_j = g(\lambda_j), \quad g_0 = g(0),$$

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

The function $\log g(\lambda)$ is an even function and its first derivative is a continuous odd function. All continuous odd functions equal zero at zero. Thus, $b_1 = 0$. By analogous reasoning, $b_k = 0$ for all odd integers $k \le [s]$. In consequence,

$$(3.2) \qquad \log(g_j/g_0) = \sum_{k=1}^{[s/2]} \frac{b_{2k}}{(2k)!}\lambda_j^{2k} + Rem_j.$$

For example, $b_2 = g^{(2)}(0)/g(0)$ and $b_4 = g^{(4)}(0)/g(0) - 3g^{(2)}(0)/g(0)$.

We break up the Taylor expansion into the part that is eliminated by the regressors λ_j^{2k} for $k = 1, \ldots, r$ and the remainder:

$$(3.3) \qquad \log(g_j/g_0) = \sum_{k=1}^{\min\{[s/2],r\}} \frac{b_{2k}}{(2k)!}\lambda_j^{2k} + R_j, \quad \text{where}$$

$$R_j = \sum_{k=\min\{[s/2],r\}+1}^{[s/2]} \frac{b_{2k}}{(2k)!}\lambda_j^{2k} + Rem_j$$

$$= 1(s \ge 2+2r)\frac{b_{2+2r}}{(2+2r)!}\lambda_j^{2+2r} + Rem_j^*,$$

$$\max_{1\le j\le m} |Rem_j^*/\lambda_j^q| = O(1), \quad \text{and}$$

$$q = \min\{s, 4+2r\}.$$

If s is an integer, then $\max_{1\le j\le m}|Rem_j/\lambda_j^s| = o(1)$ and $\max_{1\le j\le m}|Rem_j^*/\lambda_j^q| = o(1)$ by the continuity of the sth order derivative of g.

Let

$$(3.4) \qquad Q_{k,j} = \lambda_j^k \quad \text{for} \quad j = 1, \ldots, m \quad \text{and} \quad k = 1, 2, \ldots.$$

Let $\log I, X, Q_k, R$, and ε denote column m-vectors whose jth elements are $\log I_j, X_j, Q_{k,j}, R_j$, and ε_j, respectively where $\varepsilon_j = \log(I_j/f_j) + C$ and $f_j = f(\lambda_j)$.

[3] The proof that Rem_j satisfies the condition in (3.1) requires some care. But, for brevity, we do not give the proof.

Let Q denote the $m \times r$ matrix whose kth column is Q_{2k} for $k = 1, \ldots, r$. Let 1_m denote a column m-vector of ones. Let $b(r)$ denote the column r-vector whose kth element is $b_{2k}/(2k)!$ for $k = 1, \ldots, \min\{[s/2], r\}$ and 0 for $k = \min\{[s/2], r\} + 1, \ldots, r$. Combining (3.3) and (2.4) (with f^* replaced by g and $-2X_j$ replaced by X_j), we can write in m-vector notation:

$$\log I = (\log g_0 - C)1_m + Xd + Qb(r) + R + \varepsilon. \tag{3.5}$$

We define the deviation from column mean regressor vector X^* and matrix Q^* as

$$\begin{aligned} X^* &= X - 1_m \overline{X} \quad \text{and} \\ Q^* &= Q - 1_m \overline{Q}', \quad \text{where} \\ \overline{X} &= \frac{1}{m} X' 1_m \quad \text{and} \quad \overline{Q} = \frac{1}{m} Q' 1_m. \end{aligned} \tag{3.6}$$

The pseudo-regression model in deviation from mean form is

$$\begin{aligned} \log I &= K 1_m + X^* d + Q^* b(r) + R + \varepsilon, \quad \text{where} \\ K &= \log g_0 - C + \overline{X} d + \overline{Q}' b(r). \end{aligned} \tag{3.7}$$

The bias-reduced estimator $\hat{d}_r$ equals the LS estimator of the coefficient on X^* in the regression of $\log I$ on 1_m, X^*, and Q^*. By the partitioned regression formula,

$$\begin{aligned} \hat{d}_r &= (X^{*\prime} M_{Q^*} X^*)^{-1} X^{*\prime} M_{Q^*} \log I, \quad \text{where} \\ M_{Q^*} &= I_m - Q^* (Q^{*\prime} Q^*)^{-1} Q^{*\prime}. \end{aligned} \tag{3.8}$$

(For $r = 0$, we define $M_{Q^*} = I_m$.)

Taking the expectation of $\hat{d}_r$ in (3.8) and using (3.7), we obtain

$$E\hat{d}_r = d + (X^{*\prime} M_{Q^*} X^*)^{-1} X^{*\prime} M_{Q^*} (R + E\varepsilon), \tag{3.9}$$

because $X^{*\prime} 1_m = 0$ and $Q^{*\prime} 1_m = 0$. The term $Q^* b(r)$ in (3.7), which includes the λ_j^{2k} terms for $k = 1, \ldots, \min\{[s/2], r\}$ in the Taylor expansion of $\log(g_j/g_0)$, does not appear in (3.9) because it is eliminated by the inclusion of the Q^* regressors. In consequence, the bias of $\hat{d}_r$ is of smaller order than that of $\hat{d}_{GPH}$.

We now introduce several quantities that arise in the expressions for the asymptotic bias and variance of $\hat{d}_r$. Let μ_r be a column r-vector with kth element $\mu_{r,k}$ and Γ_r be an $r \times r$ matrix with (i, k) element given by $[\Gamma_r]_{i,k}$, where

$$\begin{aligned} \mu_{r,k} &= \frac{2k}{(2k+1)^2} \quad \text{for} \quad k = 1, \ldots, r, \\ [\Gamma_r]_{i,k} &= \frac{4ik}{(2i+2k+1)(2i+1)(2k+1)} \quad \text{for} \quad i, k = 1, \ldots, r. \end{aligned} \tag{3.10}$$

For $r = 0$, let $\mu_r = 0$ and $\Gamma_r = 1$. We show below that the asymptotic variance of $\hat{d}_r$ is proportional to

$$(3.11)\qquad c_r = (1 - \mu_r' \Gamma_r^{-1} \mu_r)^{-1}.$$

For example, $c_0 = 1$, $c_1 = 9/4$, $c_2 = 3.52$, $c_3 = 4.79$, and $c_4 = 6.06$.

Let ξ_r be a column r-vector with kth element $\xi_{r,k}$, where

$$(3.12)\qquad \xi_{r,k} = \frac{2k(3+2r)}{(2r+2k+3)(2k+1)} \quad \text{for} \quad k = 1, \ldots, r.$$

Let

$$(3.13)\qquad \tau_r = -\frac{(2\pi)^{2+2r}(2+2r)c_r}{2(3+2r)!(3+2r)}(1 - \mu_r' \Gamma_r^{-1} \xi_r).$$

For example, $\tau_0 = -2.19$, $\tau_1 = 2.23$, $\tau_2 = -0.793$, $\tau_3 = .146$, and $\tau_4 = -0.0164$.

We now state the asymptotic bias and variance of $\hat{d}_r$.

THEOREM 1: *Suppose Assumptions* 1 *and* 2 *hold*. *Then*:

(*a*) $E\hat{d}_r - d = 1(s \geq 2+2r)\tau_r b_{2+2r} \frac{m^{2+2r}}{n^{2+2r}}(1 + o(1)) + O(\frac{m^q}{n^q}) + O(\frac{\log^3 m}{m})$ *and*

(*b*) $\text{var}(\hat{d}_r) = \frac{\pi^2}{24}\frac{c_r}{m} + o(\frac{1}{m})$.

If s is an integer, part (*a*) *holds with $O(m^q/n^q)$ replaced by $o(m^q/n^q)$. In particular, if $s = 2+2r$, part* (*a*) *holds with $O(m^q/n^q)$ replaced by $o(m^q/n^q) = o(m^{2+2r}/n^{2+2r})$.*

COMMENTS: 1. When $s \geq 2+2r$, the dominant bias term is $\tau_r b_{2+2r} m^{2+2r}/n^{2+2r}$ whenever m grows at rate n^γ for $\gamma > (2+2r)/(3+2r)$ and $\gamma < 1$. As shown below, the MSE-optimal choice of m satisfies this condition. When $s < 2+2r$, the dominant bias term is $O(m^s/n^s)$ whenever m grows at rate n^γ for $\gamma > s/(s+1)$ and $\gamma < 1$.

2. Comparing the results of the theorem with (2.6), one sees that the convergence to zero of the bias of $\hat{d}_r$ is faster than that of $\hat{d}_{GPH}$, whereas its variance differs only by the multiplicative constant c_r.

3. The slope of the bias term $\tau_r b_{2+2r} m^{2+2r}/n^{2+2r}$ as a function of m is $(2+2r)\tau_r b_{2+2r} m^{1+2r}/n^{2+2r}$. Note that the slope converges to zero more quickly when $r > 0$ than when $r = 0$. That is, the bias term is flatter as a function of m for $r > 0$ than for $r = 0$, at least for large sample sizes.

4. Theorem 1 holds when the regressor X_j is replaced by $\widetilde{X}_j$ and $\hat{d}_r$ is -0.5 times the LS coefficient on the regressor $\widetilde{X}_j$ from the regression of $\log I_j$ on $1, \widetilde{X}_j, \lambda_j^2, \lambda_j^4, \ldots, \lambda_j^{2r}$ for $j = 1, \ldots, m$.[4]

5. The proof of Theorem 1 relies on various lemmas given in HDB.

[4] This can be proved using the fact that $\widetilde{X}_j = -(1/2)X_j + (1/2)\log\cos\vartheta_j$, where $0 \leq \vartheta_j \leq \lambda_j$; see Hurvich and Beltrao (1994, p. 299).

We now consider the MSE optimal choice of m for the bias-reduced estimator, i.e., the choice that maximizes the rate of convergence to zero of its MSE. Straightforward calculations show that the MSE optimal choice of m is

$$(3.14)\qquad m \sim n^{2\phi/(2\phi+1)}, \quad \text{where} \quad \phi = \min\{s, 2+2r\}$$

and $m \sim n^{2\phi/(2\phi+1)}$ means that $\lim_{n\to\infty} m/n^{2\phi/(2\phi+1)} \in (0,\infty)$.

For this choice,

$$(3.15)\qquad \mathrm{MSE}(\hat{d}_r) = O\big(n^{-2\phi/(2\phi+1)}\big).$$

Hence, given $s \geq 1$, if $m \sim n^{2s/(2s+1)}$ and $r \geq (s-2)/2$, then $\mathrm{MSE}(\hat{d}_r) = O(n^{-2s/(2s+1)})$. Alternatively, given $r \geq 0$, if $m \sim n^{(4+4r)/(5+4r)}$ and $s \geq 2+2r$, then $\mathrm{MSE}(\hat{d}_r) = O(n^{-(4+4r)/(5+4r)})$.

If s and r are arbitrarily large, then $\mathrm{MSE}(\hat{d}_r)$ converges to zero at a rate arbitrarily close to the rate n^{-1} of parametric estimators. This result is primarily of theoretical interest. Except for extremely large sample sizes one would choose a small value of r, because the variance of $\hat{d}_r$ for fixed m increases fairly quickly as r increases.

The MSE of the GPH estimator satisfies the formula above with $r = 0$. In consequence, when $s \leq 2$, the maximal rates of convergence to zero of the MSE of the GPH estimator and the bias-reduced estimator $\hat{d}_r$ with $r \geq 1$ are the same, viz., $n^{-2s/(2s+1)}$. When $s > 2$, however, the GPH estimator has maximal rate of convergence to zero equal to $n^{-4/5}$, whereas $\hat{d}_r$ with $r \geq 1$ has maximal rate of convergence equal to the faster rate $n^{-(2\phi)/(2\phi+1)}$ (which equals $n^{-(4+4r)/(5+4r)}$ when $s \geq 2r+2$). In fact, when $s > 2$, $\hat{d}_r$ with $r \geq 1$ has a faster rate of convergence of MSE to zero than the GPH estimator whenever $\hat{d}_r$ and $\hat{d}_{GPH}$ are defined with the same value $m \sim n^\gamma$ and $4/5 < \gamma < 1$.

Next, we derive an explicit formula for the MSE optimal choice of m for $\hat{d}_r$ when g is sufficiently smooth that $s \geq 2+2r$. Suppose that $m \sim n^\gamma$ for some $0 < \gamma < 1$. In this case, the results of Theorem 1 and some calculations show that the MSE of $\hat{d}_r$ equals

$$(3.16)\qquad \mathrm{MSE}(\hat{d}_r) = \tau_r^2 b_{2+2r}^2 \frac{m^{4+4r}}{n^{4+4r}}(1+o(1)) + O\left(\frac{m^{1+2r}\log^3 m}{n^{2+2r}}\right) + \frac{\pi^2}{24}\frac{c_r}{m}(1+o(1)).$$

(The second term on the right-hand side comes from the squared bias term $O(m^{2+2r}/n^{2+2r})O(\log^3(m)/m)$. The other remainder terms from the squared bias are dominated by the three terms in (3.16).) If $\gamma > (2+2r)/(3+2r)$, then the $O(\cdot)$ term in (3.16) is of smaller order than the other two terms. Ignoring the $O(\cdot)$ term, straightforward calculations yield the value of m that minimizes the asymptotic MSE:

$$(3.17)\qquad m_{opt,1} = \left[\left(\frac{\pi^2 c_r}{24(4+4r)\tau_r^2 b_{2+2r}^2}\right)^{1/(5+4r)} n^{(4+4r)/(5+4r)}\right],$$

where $[a]$ denotes the integer part of a and the expression for m_{opt} only applies when $\tau_r \neq 0, b_{2+2r} \neq 0$, and $c_r < \infty$. Note that the MSE optimal growth rate of $n^{(4+4r)/(5+4r)}$ allows one to ignore the $O(\cdot)$ term in (3.16).

Simulations show that a better estimator of the variance of $\hat{d}_r$ than $(\pi^2/24)\times(c_r/m)$ is obtained by using the finite sample expression $m(X^{*\prime}M_{Q^*}X^*)^{-1}$ in place of $c_r/4$. Thus, it is more accurate in finite samples to use $(\pi^2/6)(X^{*\prime}M_{Q^*}X^*)^{-1}$ in place of $(\pi^2/24)(c_r/m)$ in (3.16). This motivates an alternative specification of the value of m that minimizes the asymptotic MSE, which we denote $m_{opt,2}$. $m_{opt,2}$ is defined to be the value of m that minimizes

$$\tau_r^2 b_{2+2r}^2 \frac{m^{4+4r}}{n^{4+4r}} + \frac{\pi^2}{6}(X^{*\prime}M_{Q^*}X^*)^{-1} \tag{3.18}$$

over $m \in [m_0, n/2]$ for some small positive integer m_0. We have $m_{opt,1}/m_{opt,2} \to 1$ as $n \to \infty$ because $m(X^{*\prime}M_{Q^*}X^*)^{-1} \to c_r/4$; see Lemma 2(j) in the Appendix.

In Section 5 below, we use $m_{opt,1}$ and $m_{opt,2}$ to specify data-dependent selection rules for m by replacing b_{2+2r} with a consistent estimator $\hat{b}_{2+2r}$.

It is desirable that the MSE of $\hat{d}_r$, as a function of m, be relatively flat, because then a wide range of values of m yield MSE that is close to the value at m_{opt}. We can analyze the slope of the asymptotic MSE function in the region around m_{opt} by letting m grow at the MSE optimal rate, i.e., $m = An^{(4+4r)/(5+4r)}$ for some constant A, and computing the derivative of the dominant terms of the MSE in (3.16) with respect to the tuning parameter A. For $m = An^{(4+4r)/(5+4r)}$, we have

$$\mathrm{MSE}(\hat{d}_r) = \left(A^{4+4r}\tau_r^2 b_{2+2r}^2 + A^{-1}\frac{\pi^2}{24}c_r\right)n^{-(4+4r)/(5+4r)}(1+o(1)). \tag{3.19}$$

Ignoring the $o(1)$ term, the derivative of this expression with respect to A is

$$\left((4+4r)A^{3+4r}\tau_r^2 b_{2+2r}^2 - A^{-2}\frac{\pi^2}{24}c_r\right)n^{-(4+4r)/(5+4r)}. \tag{3.20}$$

Hence, the slope of the asymptotic MSE function as a function of the tuning parameter A converges to zero faster for $r > 0$ than for $r = 0$. This suggests that the MSE of $\hat{d}_r$ is flatter around m_{opt} for $r > 0$ than for $r = 0$, at least in large samples.

3.2. *Asymptotic Normality*

We now show that the bias-reduced estimator $\hat{d}_r$ is asymptotically normal with mean zero provided m increases to infinity at a slower rate than the MSE-optimal rate. We suppose that m is chosen to satisfy the following assumption.

ASSUMPTION 3: $m = o(n^{2\phi/(2\phi+1)})$, *where* $\phi = \min\{s, 2+2r\}$, $s \geq 1$, *and* s *is as in Assumption* 2.

We note that Assumption 3 implies Assumption 1.

THEOREM 2: *Suppose Assumptions* 2 *and* 3 *hold*. *Then*,

$$m^{1/2}(\hat{d}_r - d) \rightarrow_d N\left(0, \frac{\pi^2}{24}c_r\right) \quad as \quad n \rightarrow \infty.$$

COMMENTS: 1. Assumption 3 allows one to take m much larger for $\hat{d}_r$ than for the GPH estimator provided g is sufficiently smooth. In consequence, by appropriate choice of m, one has asymptotic normality of $\hat{d}_r$ with a faster rate of convergence than is possible with $\hat{d}_{GPH}$.

2. Assumption 3 prohibits m from growing at the MSE-optimal rate $n^{(4+4r)/(5+4r)}$ (when $s \geq 2+2r$). However, Theorem 2 can be extended easily to cover this case. Suppose Assumption 2 holds, Assumption 3 holds with $o(\cdot)$ replaced by $O(\cdot)$, and $s \geq 2+2r$. Then,

$$m^{1/2}(\hat{d}_r - d) - \tau_2 b_{2+2r}\frac{m^{(5/2)+2r}}{n^{2+2r}} \rightarrow_d N\left(0, \frac{\pi^2}{24}c_r\right) \quad \text{as} \quad n \rightarrow \infty.^5$$

3. The proof of Theorem 2 relies on the proof of Theorem 2 of HDB, which, in turn, relies on the proofs of Theorems 3 and 4 of Robinson (1995b).

4. OPTIMAL RATE OF CONVERGENCE

In this section, we determine the optimal rate of convergence of a minimax risk criterion for any estimator of d in model (1.1) for stationary Gaussian processes when the true function g is in a class of functions that includes those that are smooth of order s at zero for given $s \geq 1$. The optimal rate is $n^{-s/(2s+1)}$, which is arbitrarily close to $n^{-1/2}$ if s is arbitrarily large. We show that the bias-reduced log-periodogram estimator $\hat{d}_r$ achieves this rate provided $r \geq (s-2)/2$ and m is chosen appropriately.

Our results are obtained by establishing a lower bound for risk via the method of GRS, but we consider least favorable spectral densities that are continuous, rather than discontinuous. Then, we use the asymptotic bias and variance results of the previous section to show that the lower bound is achieved uniformly over the class of densities by the estimator $\hat{d}_r$. This yields the optimal rate of convergence result plus its achievement by the bias-reduced log-periodogram estimator.

Our optimal rate results are essentially the same as those of GRS when $1 \leq s \leq 2$. For $s > 2$, the results differ. Roughly speaking, GRS consider a class of spectral densities of the form $f(\lambda) = |\lambda|^{-2d}g(\lambda)$, where $g(\lambda) = g(0) + O(|\lambda|^s)$. Functions that are smooth of order s at zero only satisfy this condition if all the coefficients of the Taylor expansion of $g(\lambda)$ about $\lambda = 0$ to order $[s]$ are zero. That is, $g^{(k)}(0) = 0$ for all $k = 1, \ldots, [s]$. For this class of spectral densities,

[5] The proof of this result just requires altering the last equality in (7.22) in the proof of Theorem 2. (Note that when $s = 2+2r$, $O(m^{q+0.5}/n^q)$ is actually $o(m^{q+0.5}/n^q) = o(1)$ in (7.22) by the last paragraph of the proof of part (a) of Theorem 1.)

they show that the GPH estimator (with frequencies close to zero trimmed out) attains the optimal rate of convergence.

For $s > 2$, it is restrictive to focus attention only on functions $g(\lambda)$ that have derivatives up to order $[s]$ equal to zero at $\lambda = 0$. For example, an ARFIMA process has nonzero derivatives of all positive even orders at zero. This is true even if the process after differencing is white noise. A fractionally differenced process satisfies $g(\lambda) = g(0) + O(|\lambda|^s)$ only for $s = 2$, even though its $g(\cdot)$ function is smooth of order s at zero for all s finite; see Remark 3.1 on p. 57 of GRS. (Note that if an ARFIMA process is white noise after differencing, then the $g(\lambda)$ function has derivatives of all orders equal to zero if one uses the alternative local specification $f(\lambda) = |2\sin(\lambda/2)|^{-2d} g(\lambda)$.)

When we expand the class of functions to include functions $g(\lambda)$ that are smooth of order s and may have nonzero derivatives of some positive even orders less than or equal to s at $\lambda = 0$, the optimal rate of GRS does not change, but the GPH estimator no longer achieves the optimal rate of convergence. However, the bias-reduced log-periodogram estimator does achieve the optimal rate.

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:

$$\begin{aligned}(4.1)\qquad \mathcal{F}(s,a,\delta,K) = \Big\{ f : f(\lambda) = |\lambda|^{-2d_f} g(\lambda),\ |d_f| \le (1/2) - \delta_1,\ \int_{-\pi}^{\pi} f(\lambda)\,d\lambda \le \\ K_1, \text{ and } g \text{ is an even function on } [-\pi,\pi] \text{ that satisfies} \\ \text{(i) } a_0 \le g(0) \le a_{00},\ \text{(ii) } g(\lambda) = g(0) + \sum_{k=1}^{[s/2]} g_k \lambda^{2k} + \Delta(\lambda) \\ \text{for some constants } g_k \text{ with } |g_k| \le a_k \text{ for } k = 1, \ldots, [s/2] \\ \text{and some function } \Delta(\lambda) \text{ with } |\Delta(\lambda)| \le K_2 \lambda^s \\ \text{for all } 0 \le \lambda \le \delta_2,\ \text{(iii) } |g(\lambda_1) - g(\lambda_2)| \le K_3 |\lambda_1 - \lambda_2| \\ \text{for all } 0 < \lambda_1 < \lambda_2 \le \delta_3 \Big\}.\end{aligned}$$

If g is an even function on $[-\pi, \pi]$ that is smooth of order $s \ge 1$ at zero and $f(\lambda) = |\lambda|^{-2d_f} g(\lambda)$ 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 $g(\lambda)$ about $\lambda = 0$. The constants g_k equal $g^{(2k)}(0)/(2k)!$ 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 the mean value expansion because g has a bounded first derivative in a neighborhood of zero.

Next, we define a sequence of sets of values of m for which the bias-reduced estimator achieves the optimal rate of convergence. For $D_0 > 1$, let

$$(4.2)\qquad J_n(s, D_0) = \{ m : m \text{ is an integer and } D_0^{-1} n^{2s/(2s+1)} \le m \le D_0 n^{2s/(2s+1)} \}.$$

The optimal rate results are given in the following theorem:

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* $s \geq 1$, $a_0 < a_{00}$, $\delta_1 < 1/2$, *and* $K_1 \geq 2\pi a_{00}$. *Then*:

(a) *there is a constant* $C > 0$ *such that*

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

where the inf *is taken over all estimators* $\hat{d}(n)$ *of* d_f; *and*

(b) *for* $r \geq (s-2)/2$ *and any* $D_0 > 1$,

$$\limsup_{n\to\infty} \max_{m\in J_n(s,D_0)} \sup_{f\in\mathcal{F}(s,a,\delta,K)} n^{s/(2s+1)}(E_f(\hat{d}_{r,m} - d_f)^2)^{1/2} < \infty,$$

where $\hat{d}_{r,m}$ *denotes the bias-reduced estimator* $\hat{d}_r$ *calculated using* m *frequencies. Here* P_f *and* E_f *denote probability and expectation*, *respectively*, *when the true spectral density is* f.

COMMENTS: 1. The lower bound for risk stated in part (a) is for the 0-1 loss function $\ell(x) = 1(|x| > C)$. As noted in GRS, the result implies a similar result for any loss function $\ell(\cdot)$ for which $\ell(x) \geq \varepsilon 1(|x| > C)$ for all x for some $\varepsilon > 0$, such as the pth power absolute error loss function $\ell(x) = |x|^p$ for any $p > 0$. The upper bound on the risk of $\hat{d}_r$ given in part (b) is for the quadratic loss function. This result implies a similar result for any loss function $\ell(\cdot)$ for which $E_f\ell(n^{s/(2s+1)}(\hat{d}_{r,m} - d_f)) \leq h(n^{s/(2s+1)}(E_f(\hat{d}_{r,m} - d_f)^2)^{1/2})$ for any monotone positive function $h(\cdot)$, such as $h(x) = x^2$ or $h(x) = x$. In consequence, part (b) holds with the 0-1 loss function of part (a) and the pth power absolute error loss function for any $1 \leq p \leq 2$.

2. The restriction that $s \geq 1$ and condition (iii) of $\mathcal{F}(s, a, \delta, K)$ are used in place of Assumption 2 of Robinson (1995b), which requires g to be differentiable in a neighborhood of zero. The former conditions are used in the proof of part (b) of the theorem. In particular, see Lemma 3 and its proof.

3. The restrictions that $|d_f|$, $\int_{-\pi}^{\pi} f(\lambda)\, d\lambda$, and $g(0)$ are bounded away from $1/2$, ∞, and 0, respectively, in $\mathcal{F}(s, a, \delta, K)$ are imposed to ensure that uniformity over $f \in \mathcal{F}(s, a, \delta, K)$ holds in the theorem. See the proof of Lemma 3 for further discussion. The condition $K_1 \geq 2\pi a_{00}$ in the theorem ensures that the bound on the integral of $f \in \mathcal{F}(s, a, \delta, K)$ is not too severe relative to the scale of f, which is determined by $g(0)$ ($\leq a_{00}$).

5. BANDWIDTH CHOICE

In this section, we briefly discuss the choice of the number of frequencies m to employ in the log-periodogram regression. We refer to m as the bandwidth. There are several approaches in the literature for choosing the bandwidth. First, one can circumvent the problem somewhat by reporting results for a range of bandwidths and showing the extent to which the estimate of d depends on the bandwidth.

Second, if one is interested in constructing a confidence interval, one can try to choose m such that the coverage probability of the confidence interval is as close as possible to the nominal coverage probability. A recent paper by Giraitis and Robinson (2000) proposes a procedure for doing so for the local Whittle estimator.

Third, one may wish to choose the bandwidth to minimize the root mean-squared error of the estimator of d. Methods of doing so have been proposed by many authors, including Hurvich and Deo (1999), Giriatis, Robinson, and Samarov (2000), and Iouditsky, Moulines, and Soulier (2000). The latter two papers choose m to adapt to the unknown smoothness of $g(\lambda)$. In this section, we show that the method of Hurvich and Deo (1999) can be extended to the bias-reduced log-periodogram estimators considered in this paper. However, we note that this method has two drawbacks. First, one has to specify an initial bandwidth L, implying that the method is not fully automatic. Second, the finite sample properties of the procedure can be sensitive to L and can be relatively poor for some $g(\lambda)$ functions. On the other hand, for a variety of other $g(\lambda)$ functions, the method works fairly well. See Section 6 for details.

The method is to replace b_{2+2r} by a consistent estimator $\hat{b}_{2+2r}$ in the formulae for $m_{opt,1}$ and $m_{opt,2}$ in (3.17) and (3.18). This gives the following plug-in selection rules for choosing m:

$$(5.1) \qquad \hat{m}_{opt,1} = \left[\left(\frac{\pi^2 c_r}{24(4+4r)\tau_r^2 \hat{b}_{2+2r}^2} \right)^{1/(5+4r)} n^{(4+4r)/(5+4r)} \right]$$

and $\hat{m}_{opt,2}$ is the value of m that minimizes

$$(5.2) \qquad \tau_r^2 \hat{b}_{2+2r}^2 \frac{m^{4+4r}}{n^{4+4r}} + \frac{\pi^2}{6}(X^{*\prime} M_{Q^*} X^*)^{-1}$$

over $m \in [m_{low}, n/2]$. Clearly, $\hat{m}_{opt,1}/m_{opt,1} \to_p 1$ as $n \to \infty$ provided $\hat{b}_{2+2r} \to_p b_{2+2r}$ as $n \to \infty$. If $\hat{b}_{2+2r} \to_p b_{2+2r}$ and $m_{low} \to \infty$ as $n \to \infty$, then it is straightforward to show that $\hat{m}_{opt,2}/m_{opt,2} \to_p 1$ as $n \to \infty$.

It remains to specify an estimator of b_{2+2r} and show that it is consistent. Such an estimator can be obtained from a log-periodogram regression that includes one more regressor, λ_j^{2+2r}, than the regression used to obtain the estimator of d. That is, one regresses $\log I_j$ on a constant, $-2\log\lambda_j, \lambda_j^2, \ldots, \lambda_j^{2r}$, and λ_j^{2+2r}. Let L denote the number of frequencies used in this regression. The estimator $\hat{b}_{2+2r}$ is $(2+2r)!$ times the LS coefficient estimator on the regressor λ_j^{2+2r}.

To establish conditions under which $\hat{b}_{2+2r}$ is consistent, it is simplest notationally to determine conditions on m under which the LS estimator, $\hat{b}(r)$, of the r-vector $b(r)$ is consistent in the regression (3.5) that includes only r powers of λ_j. Using this notation, we have

$$(5.3) \qquad \hat{b}_{2+2r} = (2+2r)!\hat{b}(r+1)_{r+1},$$

where $\hat{b}(r+1)_{r+1}$ denotes the $(r+1)$st element of $\hat{b}(r+1)$.

The estimator $\hat{b}(r)$ of $b(r)$ is defined by

$$(5.4) \qquad \hat{b}(r) = (Q^{*\prime} M_{X^*} Q^*)^{-1} Q^{*\prime} M_{X^*} \log I, \quad \text{where}$$
$$M_{X^*} = I_m - X^*(X^{*\prime}X^*)^{-1}X^{*\prime}.$$

We define the following quantities:

$$(5.5) \qquad \begin{aligned} D_{n,r} &= \operatorname{diag}\left\{\left(\frac{2\pi m}{n}\right)^2, \dots, \left(\frac{2\pi m}{n}\right)^{2r}\right\} \quad \text{and} \\ \rho_r &= \frac{(2\pi)^{2+2r}(2+2r)}{(3+2r)!(3+2r)}. \end{aligned}$$

The bias and variance of $\hat{b}(r)$, normalized by $D_{n,r}$, are given in the following theorem.

THEOREM 4: *Suppose Assumptions* 1 *and* 2 *hold*. *Then*:

(a) $D_{n,r}(E\hat{b}(r) - b(r)) = 1(s \geq 2+2r)(\Gamma_r - \mu_r\mu_r')^{-1}(\xi_r - \mu_r)\rho_r b_{2+2r}\frac{m^{2+2r}}{n^{2+2r}}(1+o(1)) + O(\frac{m^q}{n^q}) + O(\frac{\log^3 m}{m})$ *and*

(b) $\operatorname{var}(D_{n,r}\hat{b}(r)) = (\Gamma_r - \mu_r\mu_r')^{-1}\frac{\pi^2}{6m}(1+o(1))$.

If s is an integer, part (a) *holds with $O(m^q/n^q)$ replaced by $o(m^q/n^q)$. In particular, if $s = 2+2r$, part* (a) *holds with $O(m^q/n^q)$ replaced by $o(m^q/n^q) = o(m^{2+2r}/n^{2+2r})$.*

We now use the results of Theorem 4 to determine the asymptotic bias and variance of $\hat{b}_{2+2r}$ when L frequencies are employed in the regression used to obtain $\hat{b}_{2+2r}$:

$$(5.6) \qquad \begin{aligned} &E\hat{b}_{2+2r} - b_{2+2r} \\ &\quad = 1(s \geq 4+2r)[(\Gamma_{r+1} - \mu_{r+1}\mu_{r+1}')^{-1}(\xi_{r+1} - \mu_{r+1})]_{r+1}(2+2r)!\rho_{r+1} \\ &\qquad \times b_{4+2r}(2\pi)^{-(2+2r)}\frac{L^2}{n^2}(1+o(1)) + O\left(\frac{L^{q_{r+1}-(2+2r)}}{n^{q_{r+1}-(2+2r)}}\right) \\ &\qquad + O\left(\frac{n^{2+2r}\log^3 L}{L^{3+2r}}\right) \quad \text{and} \\ &\operatorname{var}(\hat{b}_{2+2r}) \\ &\quad = \left[(\Gamma_{r+1} - \mu_{r+1}\mu_{r+1}')^{-1}\right]_{r+1}((2+2r)!)^2(2\pi)^{-(4+4r)}\frac{\pi^2}{6}\frac{n^{4+4r}}{L^{5+4r}}(1+o(1)), \end{aligned}$$

where $q_{r+1} = \min\{s, 6+2r\}$ and $[v]_{r+1}$ denotes the $(r+1)$st element of the $(r+1)$-vector v.

Hence, if $s \geq 4+2r$, the bias and variance of $\hat{b}_{2+2r}$ are $o(1)$ provided $n^{4+4r}/L^{5+4r} \to 0$ as $n \to \infty$. Thus, $\hat{b}_{2+2r}$ is consistent if $L = Cn^\delta$ for $\delta \in ((4+4r)/(5+4r), 1)$. As stated in the Introduction, for better finite sample performance, we recommend using a relatively small value of r, such as one or two, even if $g(\lambda)$ is fairly smooth. In such a case, the condition $s \geq 4+2r$ is not restrictive.

A rate-optimal choice of L in terms of minimizing the asymptotic MSE of $\hat{b}_{2+2r}$ is

$$L = Cn^{(8+4r)/(9+4r)} \text{ for some constant } C > 0.\text{[6]} \tag{5.7}$$

Given this choice of L, the rate of convergence of $\hat{b}_{2+2r}$ is $n^{-2/(9+4r)}$. That is, $n^{2/(9+4r)}(\hat{b}_{2+2r} - b_{2+2r}) = O_p(1)$. This rate is quite slow, especially when $r > 0$. This is due to the fact that estimation of b_{2+2r} is a nonparametric estimation problem. When $r > 0$, b_{2+2r} is a higher-order derivative than when $r = 0$. In consequence, the rate of convergence of $\hat{b}_{2+2r}$ is slower when $r > 0$ than when $r = 0$.

We note that for the case where $r = 0$ and $s = 3$, Hurvich and Deo (1999) show that the bias of $\hat{b}_2$ is $O(L/n)$ provided $L = An^{\delta}$ for $\delta \in (3/4, 1)$ and some constant $A > 0$. Equation (5.6) shows that this rate is not sharp if $s > 3$. In particular, for all $s \geq 4$, the sharp rate is $O(L^2/n^2)$. (Their result is not sharp when $s > 3$ because the coefficient on the cubed frequency in the Taylor expansion of $\log g(\lambda)$, using our notation, in their equation (6) is zero by the symmetry of $g(\lambda)$ about zero.)

Hurvich and Deo (1999) state that the optimal growth rate of L for minimizing the asymptotic mean-squared error of $\hat{b}_2$ is $n^{6/7}$. This is true only if $s = 3$. For $s > 3$ the optimal rate is faster and for $s < 3$ the optimal rate is slower. For all $s \geq 4$, the optimal rate is $n^{8/9}$.

6. MONTE CARLO EXPERIMENT

6.1. *Experimental Design*

In this section, we compare the finite sample behavior of the estimators $\hat{d}_0, \hat{d}_1$, and $\hat{d}_2$. The estimator $\hat{d}_0$ is the standard log-periodogram estimator, whereas $\hat{d}_1$ and $\hat{d}_2$ are bias-reduced log-periodogram estimators. We consider stationary Gaussian ARFIMA$(1, d, 1)$ processes with autoregressive parameter (AR) ϕ and moving average (MA) parameter θ. When $d = 0$, the model is

$$Y_t = \phi Y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1} \quad \text{for} \quad t = 1, \ldots, n, \tag{6.1}$$

where $\{\varepsilon_t : t = 0, \ldots, n\}$ are iid standard normal random variables.

We consider the processes that correspond to all possible combinations of

$$\begin{aligned} &d = 0,\ .4,\ -.4, \\ &\phi = .9,\ .6,\ .3,\ 0,\ -.3,\ -.6,\ -.9, \quad \text{and} \\ &\theta = .9,\ .6,\ .3,\ 0,\ -.3,\ -.6,\ -.9. \end{aligned} \tag{6.2}$$

We consider sample sizes $n = 128, 512$, and 2048. We use 20,000 simulation repetitions.

[6] The optimal constant can be determined straightforwardly from (5.6), but it depends on the unknown b_{4+2r}.

The results are quite similar for a wide variety of parameter combinations, so we only explicitly report results for a small subset of the parameter combinations.

We calculate the biases, standard deviations, and RMSE's of $\hat{d}_0$, $\hat{d}_1$, and $\hat{d}_2$ as functions of m for $m = 4, 5, \ldots, n/2$ for $n = 128$ and 512 and for $m = 10, 11, \ldots, n/2$ for $n = 2048$. For a given parameter combination, we report these quantities in three graphs—one each for the bias, standard deviation, and RMSE. In each graph, values of m are given on the horizontal axis. For ease of comparison, the axes have the same scales in each of the three graphs.

In addition, we calculate the coverage probabilities, as functions of m, of the nominal 90% CI's that are obtained by using the asymptotic normality result of Theorem 2. When constructing these CI's, we estimate the standard error of $m^{1/2}(\hat{d}_r - d)$ using the finite sample expression $(X^{*\prime}M_{Q^*}X^*/m)^{-1}$ rather than its limit $c_r/4$ (see Lemma 2(j) in the Appendix), because it yields better finite sample results for all parameter combinations and estimators considered. In particular, the CI's are

$$\left[\hat{d}_r - z_{.95}\left(\frac{\pi^2}{6X^{*\prime}M_{Q^*}X^*}\right)^{1/2}, \hat{d}_r + z_{.95}\left(\frac{\pi^2}{6X^{*\prime}M_{Q^*}X^*}\right)^{1/2}\right] \quad \text{for} \quad r = 0, 1, 2, \tag{6.3}$$

where $z_{.95}$ is the .95 quantile of the standard normal distribution. We compute the average lengths of the CI's as functions of m. These lengths do not depend on the parameter combination considered and, hence, are only reported for one parameter combination.

We evaluate the performance of the data-dependent choices $\hat{m}_{opt,1}$ and $\hat{m}_{opt,2}$. The number of frequencies L used to obtain the estimator $\hat{b}_{2+2r}$ is given by the rate-optimal formula (5.7). Two values of the constant C are considered, viz., $C = .3$ and $C = .4$. These values were chosen because they perform reasonably well in an overall sense for a wide range of ARFIMA$(1, d, 1)$ processes.

6.2. *Simulation Results*

6.2.1. *Basic Results*

We discuss the results for $d = 0$ and $n = 512$ first. We find that for any given positive AR parameter ϕ the pattern of results does not vary much across MA parameter values $\theta < \phi$. In addition, cases where $\theta > \phi$ are ones in which the first two autocorrelations of the process are negative, which is of relatively low empirical relevance; cases in which $\theta = \phi$ all reduce to the iid case; and cases in which the AR parameter is negative are of relatively low empirical relevance. In consequence, we focus on reporting results for nonnegative values of the AR parameter and these results can be well summarized by considering the parameter combinations in which the MA parameter θ is zero. When $\phi = \theta = 0$, the process is iid, none of the estimators are biased for any value of m, and the results are as expected. Hence, for the case when $\phi = 0$, it is more interesting to consider results for $\theta = -.9$, which yields a MA(1) process with positive autocorrelation. In consequence, we report in Figures 1 and 2 results for $(\phi, \theta) = (.9, 0)$

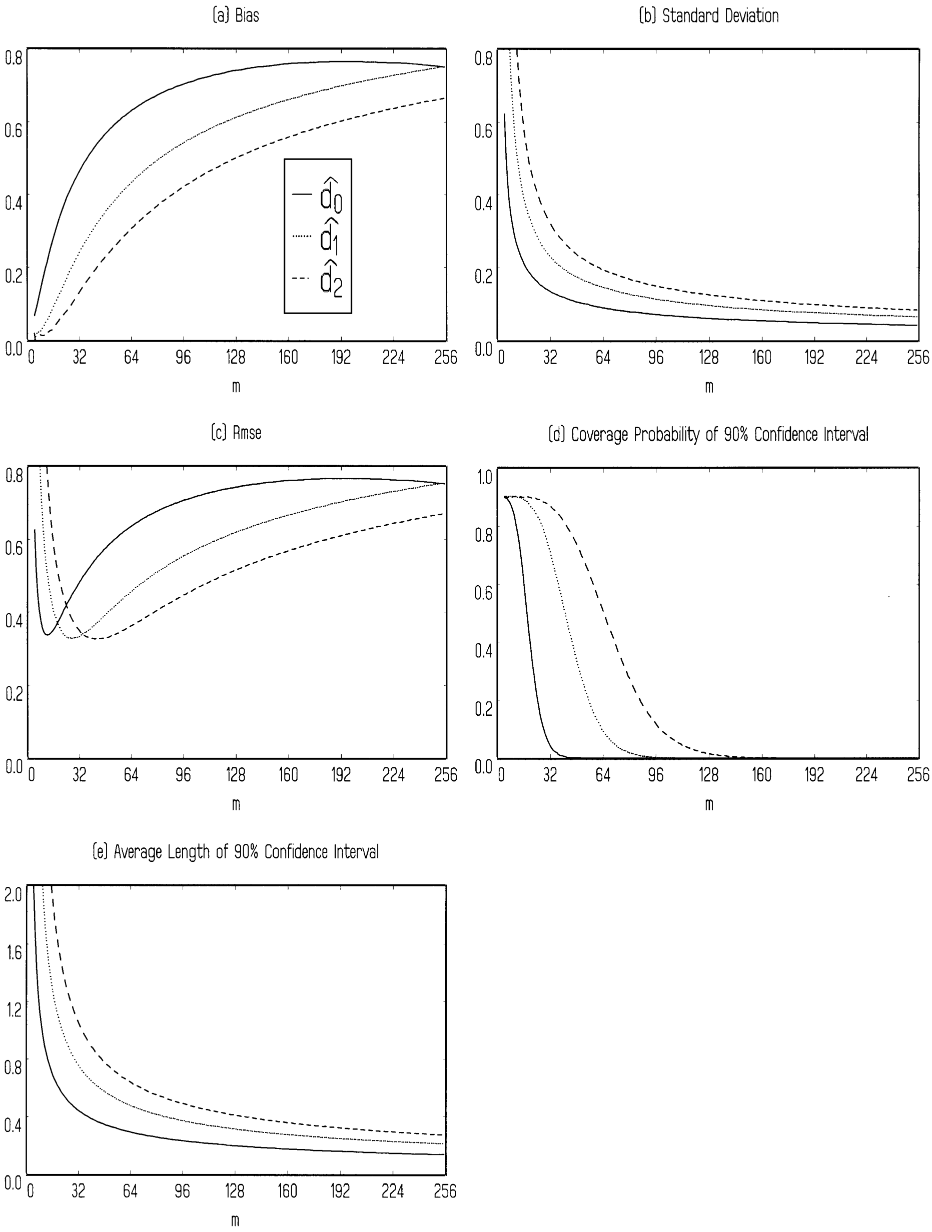

FIGURE 1.—Performance of the log-periodogram regression estimators $\hat{d}_0$, $\hat{d}_1$, and $\hat{d}_2$ for an AR(1) process with AR parameter $\phi = .9$, for sample size $n = 512$, computed using 20,000 simulation repetitions.

and $(.6, 0)$, respectively, where in each case $d = 0$ and $n = 512$. In addition, we describe, but do not present figures for, results for the cases where $(\phi, \theta) = (.3, 0)$ and $(0, -.9)$.

Figure 1 provides results for the AR(1) model with AR parameter .9. The bias graphs in Figure 1(a) show that the bias of the standard log-periodogram estimator $\hat{d}_0$ grows very rapidly as m increases, whereas the biases of $\hat{d}_1$ and especially $\hat{d}_2$ grow much more slowly. It is apparent that the bias-reducing features of $\hat{d}_1$ and $\hat{d}_2$ that are established in the asymptotic results are reflected in this finite sample scenario. The standard deviation graphs in Figure 1(b) show that the standard deviation of $\hat{d}_0$ is less than that of $\hat{d}_1$ and $\hat{d}_2$ for all values of m, as predicted by the asymptotic results. For each estimator, the standard deviation declines at the approximate rate $1/\sqrt{m}$ as m increases because m indexes the effective sample size used to estimate d. We note that the standard deviation graphs for all (ϕ, θ) combinations (including those that are not reported) are essentially the same; whereas the bias graphs and, hence, the RMSE graphs, vary across parameter combinations.

The RMSE graph in Figure 1(c) shows that the minimum RMSE across values of m is somewhat smaller for $\hat{d}_1$ and $\hat{d}_2$ than for $\hat{d}_0$, which is in accord with the asymptotic results. The actual minimum RMSE values for $\hat{d}_0, \hat{d}_1$, and $\hat{d}_2$ are .337, .328, and .327, respectively (see column three of Table II below). In addition, one sees that the RMSE graph for $\hat{d}_0$ rises very steeply from its minimal value, whereas the RMSE graphs for $\hat{d}_1$ and $\hat{d}_2$ rise more slowly. In consequence, $\hat{d}_1$ and $\hat{d}_2$ have low RMSE's over wider ranges of m values and, hence, are not as sensitive to the choice of m as $\hat{d}_0$. This reflects the asymptotic result that the slope of the RMSE function converges to zero more quickly when $r > 0$ than when $r = 0$.

The CI coverage probability graphs in Figure 1(d) show that $\hat{d}_0$ has true coverage probability close to .9 only for very small values of m. This is due to the bias of $\hat{d}_0$ for larger values of m. In contrast, the coverage probabilities of $\hat{d}_1$ and $\hat{d}_2$ are close to .9 for a wider range of values of m, due to their smaller biases. Thus, $\hat{d}_1$ and $\hat{d}_2$ yield CI's that are more robust to the choice of m than does $\hat{d}_0$. On the other hand, the larger standard deviations of $\hat{d}_1$ and $\hat{d}_2$ lead to larger average lengths of their CI's than those of $\hat{d}_0$, as is shown in Figure 1(e).

Figure 2 provides results for the AR(1) model with AR parameter .6. Given the lower level of dependence in the data, the bias of $\hat{d}_0$ increases more slowly as m increases than in Figure 1(a). The biases of $\hat{d}_1$ and $\hat{d}_2$ are reduced quite considerably as well. They are sufficiently small that a wide range of values of m yield good coverage probabilities in Figure 2(d). The RMSE's of $\hat{d}_1$ and $\hat{d}_2$ in Figure 2(c) are slightly lower than those of $\hat{d}_0$ due to their lower biases. In particular, the minimum RMSE values for $\hat{d}_0, \hat{d}_1$, and $\hat{d}_2$ are .150, .143, and .142, respectively. In addition, the RMSE graphs for $\hat{d}_1$ and $\hat{d}_2$ are quite flat in Figure 2(c), which implies that a wide range of values of m yield low RMSE.

For the AR(1) model with AR parameter .3, the biases of all three estimators are reduced further from those reported in Figures 1 and 2. In fact, the biases of $\hat{d}_1$ and $\hat{d}_2$ are quite small across the entire range of m values. As a result,

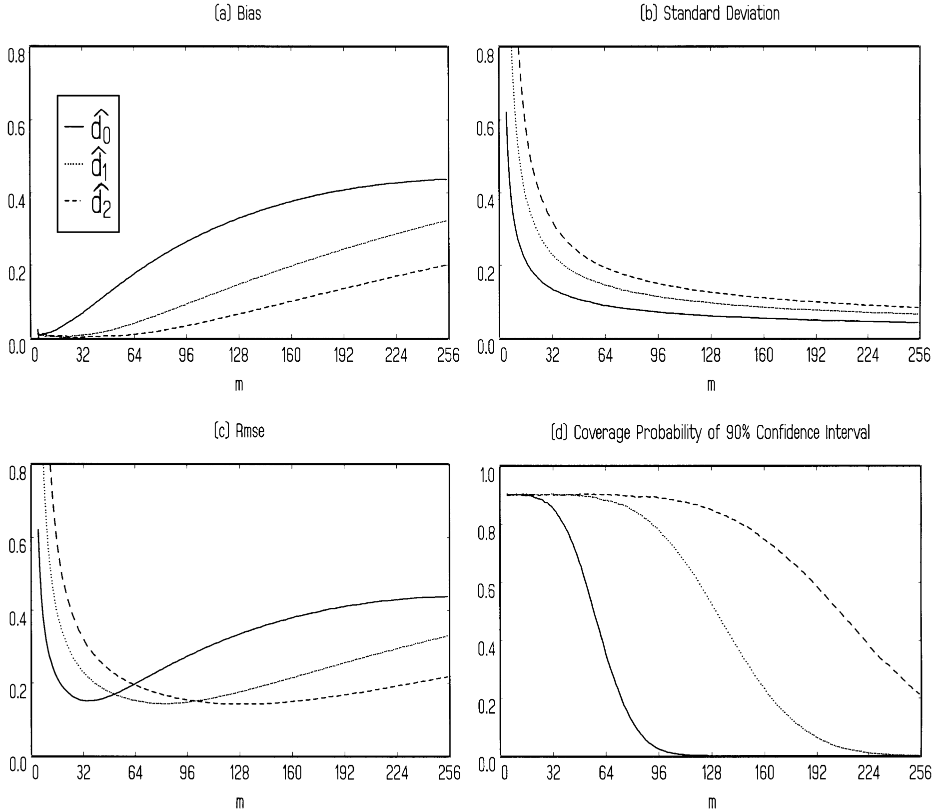

FIGURE 2.—Performance of the log-periodogram regression estimators $\hat{d}_0$, $\hat{d}_1$, and $\hat{d}_2$ for an AR(1) process with AR parameter $\phi = .6$, for sample size $n = 512$, computed using 20,000 simulation repetitions.

the coverage probabilities of the CI's based on $\hat{d}_1$ and $\hat{d}_2$ are quite robust to the choice of m—much more so than $\hat{d}_0$. The minimum RMSE values of $\hat{d}_0$, $\hat{d}_1$, and $\hat{d}_2$ are .099, .094, and .093, respectively. So, again, the estimators $\hat{d}_1$ and $\hat{d}_2$ have lower minimum RMSE's than $\hat{d}_0$. The RMSE's of all three estimators are fairly flat, which indicates that all three are relatively robust to the choice of m.

For the MA(1) model with MA parameter $-.9$, the results are similar to those of the AR(1) model with AR parameter .3 except that the bias of $\hat{d}_1$ is negative and the bias and RMSE of $\hat{d}_0$ rise more sharply for large values of m. The biases of $\hat{d}_1$ and $\hat{d}_2$ are quite close to zero over a wide range of values of m, which yields low RMSE's and CI coverage probabilities that are close to the nominal level .9 for a wide range of values of m. The minimum RMSE values for $\hat{d}_0$, $\hat{d}_1$, and $\hat{d}_2$ are .085, 0.79, and .090, respectively. Thus, in this case, $\hat{d}_1$ has a lower RMSE than $\hat{d}_0$, but $\hat{d}_2$ does not.

Next, we discuss the results for $d = .4$ and $d = -.4$. The results are so similar to those for $d = 0$ that there is no point in presenting graphs of any of these results. For most parameter combinations the differences across values of d are so small that they cannot be detected by the eye. In the few cases where differences can be detected, they are small differences in the magnitudes of the biases for quite large values of m. Additional simulations for $d = .49$ and $d = -.49$ also show very little sensitivity of the results to the value of d.

Finally, we discuss the results for the larger and smaller sample sizes $n = 2048$ and 128. For all sample sizes, the horizontal scaling of the graphs is such that m varies from 0 to $n/2$. The results are relatively easy to describe. For every (ϕ, θ) parameter combination, the bias and coverage probability graphs are quite similar to their $n = 512$ counterparts and the standard deviation, RMSE, and average CI length graphs are quite similar in shape to those for $n = 512$ but are shifted down toward the horizontal axis for $n = 2048$ and are shifted up for $n = 128$. Similarity of the results is very close for $n = 2048$ and somewhat less so for $n = 128$. For brevity, we do not report any figures for $n = 2048$ and $n = 128$.

The similarity of the bias graphs for $n = 128, 512$, and 2048 is due to the horizontal scaling of the graphs in which m varies from 0 to $n/2$ and a given horizontal distance corresponds to the same fraction of the sample size in all graphs. For a given value of m the bias is noticeably smaller when $n = 2048$ and noticeably larger when $n = 128$ than when $n = 512$, but for m equal to a given fraction of the sample size, the bias is found to be almost the same. The similarities of the bias graphs yield similarities of the coverage probability graphs. On the other hand, the standard deviation graphs shift downward when n is increased to 2048 and upward when $n = 128$ because a given fraction of the sample size corresponds to a larger value of m and it is the value of m that primarily determines the standard deviation. In consequence, the RMSE and average CI lengths also shift downward when n is increased to 2048 and upward when n is decreased to 128.

Table I provides a comparison for sample sizes $n = 128, 512$, and 2048 of the minimum RMSE's of $\hat{d}_0$, $\hat{d}_1$, and $\hat{d}_2$ for a variety of ARFIMA$(1, d, 0)$ processes. (For $n = 128$ and 512, the minima are taken over $m \in \{4, \ldots, n/2\}$. For $n = 2048$, the minima are taken over $m \in \{10, \ldots, n/2\}$.) The numbers in parentheses in Table I give the ratios of the RMSE's of $\hat{d}_1$ and $\hat{d}_2$ to that of $\hat{d}_0$ for each (ϕ, n) combination. The results of Table I show that $\hat{d}_0$ has the lowest minimum RMSE for most values of the AR parameter ϕ when $n = 128$ and $\hat{d}_1$ is a close second. When the sample size is increased to 512, the estimator $\hat{d}_1$ has smaller minimum RMSE than $\hat{d}_0$ by 3% to 18% for all values of ϕ except $\phi = 0$. When the sample size is increased further to 2048, both $\hat{d}_1$ and $\hat{d}_2$ have smaller minimum RMSE's than $\hat{d}_0$ by 5% to 19% for all values of ϕ except $\phi = 0$. It is clear from the table that the larger the sample size, the better the performance of $\hat{d}_1$ and $\hat{d}_2$ relative to $\hat{d}_0$ in terms of minimum RMSE for ARFIMA$(1, d, 0)$ processes.

In sum, the Monte Carlo simulation results show that $\hat{d}_1$ and $\hat{d}_2$ have lower biases, higher standard deviations, and slightly lower RMSE's compared to $\hat{d}_0$ for a wide range of stationary Gaussian ARFIMA$(1, d, 1)$ processes and sample

TABLE I

MINIMUM RMSE'S FOR $\hat{d}_0$, $\hat{d}_1$, AND $\hat{d}_2$ FOR SAMPLE SIZES 128, 512, AND 2048 FOR GAUSSIAN ARFIMA$(1, d, 0)$ PROCESSES WITH AR PARAMETER ϕ AND $d = 0$

AR Parameter ϕ	Estimator $\hat{d}_r$ r	Sample Size n 128	512	2048
.3	0	.184*	.099	.054
	1	.186 (1.01)	.094 (.95)	.049 (.91)
	2	.201 (1.09)	.093* (.94)	.048* (.89)
.6	0	.299	.150	.080
	1	.294 (.98)	.143 (.95)	.073 (.91)
	2	.293* (.98)	.142* (.95)	.072* (.90)
.9	0	.680*	.337	.166
	1	.682 (1.00)	.328 (.97)	.157 (.95)
	2	.682 (1.00)	.327* (.97)	.155* (.93)
−.3	0	.148*	.077	.042
	1	.149 (1.01)	.069* (.90)	.034* (.81)
	2	.196 (1.32)	.084 (1.09)	.040 (.95)
−.6	0	.161	.083	.045
	1	.148* (.92)	.068* (.82)	.038* (.84)
	2	.196 (1.22)	.085 (1.02)	.040 (.89)
−.9	0	.164*	.084	.045
	1	.171 (1.04)	.079* (.94)	.040* (.89)
	2	.204 (1.24)	.089 (1.06)	.043 (.96)
0	0	.090*	.042*	.020*
	1	.146 (1.62)	.065 (1.55)	.031 (1.55)
	2	.195 (2.17)	.083 (1.98)	.039 (1.95)

Notes: The numbers in parentheses are the ratios of the minimum RMSE's of $\hat{d}_r$ to those of $\hat{d}_0$ for each (ϕ, n) combination. Asterisks denote the smallest minimum RMSE over $\hat{d}_0$, $\hat{d}_1$, and $\hat{d}_2$ for each (ϕ, n) combination.

sizes 512 or 2048. These results are consistent with the asymptotic results. The lower biases lead to good CI coverage probabilities for $\hat{d}_1$ and $\hat{d}_2$ over a wider range of m values than for $\hat{d}_0$. On the other hand, the lower standard deviation of $\hat{d}_0$ leads to shorter CI intervals than CI's based on $\hat{d}_1$ and $\hat{d}_2$. The RMSE graphs for $\hat{d}_1$ and $\hat{d}_2$ are flatter than those for $\hat{d}_0$, which implies that $\hat{d}_1$ and $\hat{d}_2$ are less sensitive to the choice of m than is $\hat{d}_0$. The results are essentially the same for the three values of d considered: $-4, 0$, and .4. The basic pattern of results is the same for sample sizes $n = 128, 512$, and 2048. However, the minimum RMSE values move in favor of $\hat{d}_1$ and $\hat{d}_2$ as the sample size is increased.

6.2.2. *Plug-in Selection of m Results*

Table II presents results for the plug-in selection estimator $\hat{m}_{opt,2}$ for $n = 512$ and $d = 0$. Results for $\hat{m}_{opt,1}$ are similar, but not quite as good. Table II provides the RMSE's and CI coverage probabilities (in the columns headed 90% CI)

TABLE II

MINIMUM RMSE AND RESULTS FOR PLUG-IN SELECTION OF m FOR $\hat{d}_0$, $\hat{d}_1$, AND $\hat{d}_2$ FOR ARFIMA$(1, d, 0)$ PROCESSES WITH AR PARAMETER ϕ, $d = 0$, AND $n = 512$

	Estimator	Min	Data-dependent m: $C = .3$					Data-dependent m: $C = .4$				
ϕ	r	RMSE	RMSE	90% CI	Bias	Std Dev	Avg m	RMSE	90% CI	Bias	Std Dev	Avg m
.3	0	.099	.119*	.742	.045	.110	77	.117	.678	.059	.101	94
	1	.094	.125	.875	.011	.125	99	.110*	.853	.021	.108	129
	2	.093*	.146	.890*	.003	.146	110	.125	.884*	.008	.124	145
.6	0	.150	.186	.576	.129	.134	58	.198	.460	.159	.119	65
	1	.143	.156	.755	.077	.136	95	.168	.631	.113	.123	119
	2	.142*	.155*	.859*	.041	.149	109	.149*	.791*	.072	.130	143
.9	0	.337	.568	.015	.551	.137	53	.625	.001	.614	.117	69
	1	.328	.507	.046	.486	.147	86	.568	.007	.554	.127	111
	2	.327*	.450*	.138*	.420	.160	104	.513*	.032*	.494	.137	135
−.3	0	.077	.105*	.819	−.020	.103	79	.099*	.769	−.030	.094	101
	1	.069*	.123	.884	−.001	.123	100	.105	.882	−.001	.105	130
	2	.084	.146	.891*	.001	.146	110	.124	.886*	.001	.124	145
−.6	0	.083	.119*	.795	−.031	.115	79	.121	.743	−.045	.113	98
	1	.068*	.124	.883	.001	.124	100	.105*	.882	.003	.105	130
	2	.085	.146	.889*	.001	.146	110	.124	.887*	.001	.124	145
−.9	0	.084	.131	.789	−.036	.126	79	.141	.737	−.052	.131	97
	1	.079*	.125*	.880	.002	.124	100	.107*	.869	.008	.107	130
	2	.089	.146	.888*	.001	.146	110	.124	.885*	.000	.124	145
0	0	.042*	.098*	.875	.001	.098	81	.083*	.874	.001	.083	105
	1	.065	.123	.885	−.001	.123	100	.105	.885	.000	.105	130
	2	.083	.146	.891*	.001	.146	110	.124	.887*	.001	.124	145

obtained using $\hat{m}_{opt,2}$ (for two values of the constant C that is used to define the number of frequencies L employed in obtaining $\hat{b}_{2+2r}$). Asymptotically, $\hat{m}_{opt,2}$ selects m values that are too large to yield CI's with coverage probability 90%. Nevertheless, it is of interest to see what coverage probabilities are obtained in finite samples. Table II also provides information on the bias, standard deviation, and average m values obtained by $\hat{m}_{opt,2}$.

In Table II, the estimator with the lowest RMSE varies with ϕ. However, it seems that $\hat{d}_1$ with $C = .4$ is the best estimator overall in terms of RMSE. It is better than $\hat{d}_0$ with $C = .4$ for $\phi = .3, .6, .9, -.6$, and $-.9$. It is as good as or better than $\hat{d}_0$ with $C = .3$ for $\phi = .3, .6, .9, -.3, -.6$, and $-.9$. It is better than $\hat{d}_2$ with $C = .3$ or $.4$ for $\phi = .3, -.3, -.6, -.9$, and 0. Nevertheless, the differences between the estimators in terms of RMSE is not large and each estimator is best for some values of ϕ.

One can compare the finite sample minimum RMSE to the RMSE delivered by the plug-in selection method $\hat{m}_{opt,2}$ by comparing the third column of Table II to the fourth and ninth columns. For example, the percentage increases in RMSE due to estimating b_4 using $C = .4$ for use with $\hat{d}_1$ for $\phi = .3, .6, .9, -.3, -.6, -.9, 0$ are $17, 17, 73, 52, 54, 35, 62$, respectively. One obtains similar percentage increases for $\hat{d}_0$ with $C = .3$ or .4—some smaller and some larger. The percentage increases for $\hat{d}_2$ with $C = .4$ are somewhat lower. Given that estimation of b_{2+2r} is a nonparametric estimation problem requiring estimation of a higher-order derivative, increases in RMSE of this magnitude should not be unexpected.

The CI coverage probability results of Table II show that $\hat{d}_2$ performs best and $\hat{d}_1$ is a close second. $\hat{d}_0$ performs noticeably worse. The coverage probabilities of $\hat{d}_1$ and $\hat{d}_2$ are pretty good for all ϕ values except $\phi = .6$ and .9. For $\phi = .9$, all estimators have coverage probabilities that are far too low.

In Table II, the biases of $\hat{d}_1$ and $\hat{d}_2$ using $\hat{m}_{opt,2}$ are smaller than those of $\hat{d}_0$ for all values of ϕ except $\phi = 0$. The absolute bias decreases with r. The differences in bias are appreciable. For example, the ratios of the bias of $\hat{d}_0$ to that of $\hat{d}_1$ with $C = .4$ for $\phi = .3, .6, .9, -.3, -.6, -.9$ are $2.8, 1.4, 1.1, 30.0, -15.0, -6.5$, respectively.

The opposite is true for the standard deviations. The standard deviations increase with r in most cases in Table II. The ratios of standard deviations of $\hat{d}_1$ to $\hat{d}_0$ with $C = .4$ for $\phi = .3, .6, .9, -.3, -.6, -.9, 0$ are $1.07, 1.03, 1.09, 1.12, .93, .82, 1.27$, respectively. It is apparent that the estimators $\hat{d}_1$ and $\hat{d}_2$ with m chosen via $\hat{m}_{opt,2}$ provide different ratios of bias to standard deviation than $\hat{d}_0$. It is the smaller biases for $\hat{d}_1$ and $\hat{d}_2$ that yield better CI coverage probabilities than for $\hat{d}_0$.

The average m values selected by $\hat{m}_{opt,2}$ increase with r in all cases. For any given estimator, the average m decreases as ϕ increases from 0 to .9 (except for $\hat{d}_2$ with $C = .4$).

The conclusions from Table II are as follows. The estimators $\hat{d}_1$ and $\hat{d}_2$ using $\hat{m}_{opt,2}$ are certainly competitive with $\hat{d}_0$. Of the estimators considered, the best estimator in terms of overall RMSE and CI coverage probability performance is $\hat{d}_1$ with $C = .4$. The date-dependent choice of m significantly increases RMSE's over the finite sample minimum RMSE's. This is not too surprising because the plug-in choice of m relies on estimation of b_{2+2r}, which is a nonparametric estimation problem. Nevertheless, it may be possible to derive data-dependent choices of m that outperform the method $\hat{m}_{opt,2}$ that is considered here.

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

and

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

Manuscript received May, 2000; final revision received May, 2002.

APPENDIX: PROOFS

The estimator $\hat{d}_r$ is unchanged if we replace the regressor matrix Q^* by a matrix that spans the same column space. To simplify some of the expressions below, we replace Q^* by such a matrix Z^*, which rescales the columns of Q^*. Let

$$Z_{k,j} = (j/m)^k. \tag{7.1}$$

Let Z_k denote the m-vector whose jth element is $Z_{k,j}$ for $k = 1, 2, \ldots$. Let Z denote the $m \times r$ matrix whose kth column is Z_{2k} for $k = 1, \ldots, r$. Let Z^* denote the $m \times r$ deviation from column mean matrix defined by

$$Z^* = Z - 1_m \overline{Z}', \quad \text{where} \quad \overline{Z} = \frac{1}{m} Z' 1_m. \tag{7.2}$$

Similarly, let $Z_k^* = Z_k - 1_m \overline{Z}_k$, where $\overline{Z}_k = (1/m) Z_k' 1_m$. Let $M_{Z^*} = I_m - Z^*(Z^{*\prime} Z^*)^{-1} Z^{*\prime}$. (If $r = 0$, we take $M_{Z^*} = I_m$.)

The proof of Theorem 1 uses the following lemmas:

LEMMA 1: *Suppose Assumptions* 1 *and* 2 *hold. Then*:
(a) $1_m' X = 2m(\log n - \log m + 1 - \log(2\pi)) + O(\log m)$,
(b) $1_m' Z_k = \frac{1}{k+1} m + O(1)$,
(c) $Z_i' Z_k = \frac{1}{i+k+1} m + O(1)$,
(d) $(X + 1_m 2\log(2\pi/n))' Z_k = -\frac{2}{k+1} m \log m + \frac{2}{(k+1)^2} m(1 + o(1))$,
(e) $1_m' R = 1(s \geq 2 + 2r) \frac{(2\pi)^{2+2r} b_{2+2r}}{(3+2r)!} \frac{m^{3+2r}}{n^{2+2r}} (1 + O(\frac{1}{m})) + O(\frac{m^{q+1}}{n^q})$,
(f) $Z_k' R = 1(s \geq 2 + 2r) \frac{(2\pi)^{2+2r} b_{2+2r}}{(2+2r)!(3+2r+k)} \frac{m^{3+2r}}{n^{2+2r}} (1 + o(1)) + O(\frac{m^{q+1}}{n^q})$,
(g) $1_m' |E\varepsilon| = O(\log^2 m)$, *and*
(h) $Z_k' E\varepsilon = O(\log m)$,
for $i, k = 1, 2, \ldots$, *where* $|E\varepsilon|$ *denotes the m-vector of absolute values of the elements of* $E\varepsilon$. *When s is an integer, the results in parts* (e) *and* (f) *hold with* $O(m^{q+1}/n^q)$ *replaced by* $o(m^{q+1}/n^q)$.

LEMMA 2: *Suppose Assumptions* 1 *and* 2 *hold. Then*:
(a) $X^{*\prime} X^* = 4m(1 + o(1))$,
(b) $Z_i^{*\prime} Z_k^* = \frac{ik}{(i+k+1)(i+1)(k+1)} m(1 + o(1))$,
(c) $X^{*\prime} Z_k^* = -\frac{2k}{(k+1)^2} m(1 + o(1))$,
(d) $X^{*\prime} R = -1(s \geq 2 + 2r) \frac{2(2\pi)^{2+2r}(2+2r) b_{2+2r}}{(3+2r)!(3+2r)} \frac{m^{3+2r}}{n^{2+2r}} (1 + o(1)) + O(\frac{m^{q+1}}{n^q})$,
(e) $Z_k^{*\prime} R = 1(s \geq 2 + 2r) \frac{(2\pi)^{2+2r}(2+2r) k b_{2+2r}}{(3+2r)!(3+2r+k)(k+1)} \frac{m^{3+2r}}{n^{2+2r}} (1 + o(1)) + O(\frac{m^{q+1}}{n^q})$,
(f) $X^{*\prime} E\varepsilon = O(\log^3 m)$,
(g) $Z_k^{*\prime} E\varepsilon = O(\log^2 m)$,
(h) $Z^{*\prime} Z^* = \Gamma_r m(1 + o(1))$,
(i) $Z^{*\prime} X^* = -2\mu_r m(1 + o(1))$,
(j) $X^{*\prime} M_{Z^*} X^* = (4m/c_r)(1 + o(1))$, *and*
(k) $\max_{1 \leq j \leq m} |[M_{Z^*} X^*]_j| = O(\log m)$,
for $i, k = 1, 2, \ldots$, *provided* Γ_r *is nonsingular in parts* (j) *and* (k). *When s is an integer, the results in parts* (d) *and* (e) *hold with* $O(m^{q+1}/n^q)$ *replaced by* $o(m^{q+1}/n^q)$.

PROOF OF THEOREM 1: We prove part (a) first. Using (3.9), we just have to approximate the following three terms: (i) $X^{*\prime} M_{Z^*} X^*$, (ii) $X^{*\prime} M_{Z^*} R$, and (iii) $X^{*\prime} M_{Z^*} E\varepsilon$. The term in (i) equals $(4m/c_r)(1 + o(1))$ by Lemma 2(j).

Suppose $s \geq 2 + 2r$. Then, by Lemma 2(e) and the definition of ξ_r,

$$Z^{*\prime} R = \frac{(2\pi)^{2+2r}(2+2r) b_{2+2r}}{(3+2r)!(3+2r)} \xi_r \frac{m^{3+2r}}{n^{2+2r}} (1 + o(1)) + O\left(\frac{m^{q+1}}{n^q}\right). \tag{7.3}$$

Combining this with Lemma 2(d), (h), and (i) gives

$$X^{*\prime} M_{Z^*} R = -\frac{2(2\pi)^{2+2r}(2+2r)}{(3+2r)!(3+2r)}\left(1-\mu_r' \Gamma_r^{-1}\xi_r\right) b_{2+2r}\frac{m^{3+2r}}{n^{2+2r}}(1+o(1)) + O\left(\frac{m^{q+1}}{n^q}\right). \tag{7.4}$$

Next, suppose $s < 2+2r$. Then, by Lemma 2(d), (e), (h), and (i),

$$X^{*\prime} M_{Z^*} R = X^{*\prime} R - X^{*\prime} Z^* (Z^{*\prime} Z^*)^{-1} Z^{*\prime} R = O\left(\frac{m^{s+1}}{n^s}\right). \tag{7.5}$$

Finally, by Lemma 2(f)–(i),

$$X^{*\prime} M_{Z^*} E\varepsilon = O(\log^3 m) + 2\mu_r' \Gamma_r^{-1} 1_r (1+o(1)) O(\log^2 m) = O(\log^3 m). \tag{7.6}$$

Equations (3.9), (7.4), and (7.6), Lemma 2(j), and the definition of τ_r combine to establish part (a) when $s \geq 2+2r$. Equations (3.9), (7.5), and (7.6), and Lemma 2(j) combine to establish part (a) when $s < 2+2r$. When s is an integer, the $O(\cdot)$ terms are $o(\cdot)$ in (7.3)–(7.5) because the same holds in the definition of R_j in (3.3).

To prove part (b), we use (3.8) and the proof of Theorem 1 of HDB. We replace their $4S_{xx}^2$ by $(X^{*\prime} M_{Z^*} X^*)^2 = (4m/c_r)^2(1+o(1))$ and note that their proof goes through with the variance of their term $T_2 = \sum_{j=1+\log^6 m}^{m} a_j \varepsilon_j$ equal to $K_0 \pi^2 m/6 + o(m)$ for any triangular array $\{a_j : j = 1, \ldots, m;\ m \geq 1\}$ for which

$$\max_{1\leq j\leq m} |a_j| = O(\log m) \quad \text{and} \quad \sum_{j=1}^{m} a_j^2 = K_0 m(1+o(1)) \quad \text{for some } K_0 > 0. \tag{7.7}$$

In our case, $a_j = [M_{Z^*} X^*]_j$ and $\sum_{j=1}^m a_j^2 = X^{*\prime} M_{Z^*} X^* = (4m/c_r)(1+o(1))$ by Lemma 2(j). The first condition of (7.7) holds by Lemma 2(k). The second holds with $K_0 = 4/c_r$. From the proof of Theorem 1 of HDB, we have

$$\operatorname{var}(\hat{d}_r) = \frac{1}{(X^{*\prime} M_{Z^*} X^*)^2}\frac{4m}{c_r}\frac{\pi^2}{6}(1+o(1)) = \frac{\pi^2}{24}\frac{c_r}{m}(1+o(1)). \tag{7.8}$$

We note that the Lemmas of HDB, which are relied on here and in the proof of Lemma 1(g) and (h), utilize Theorem 2 of Robinson (1995b). The latter uses Assumption 2 of Robinson (1995b) that $f^{(1)}(\lambda) = O(|\lambda|^{-1-2d})$ as $\lambda \to 0$. This assumption is implied by our Assumption 2. *Q.E.D.*

PROOF OF LEMMA 1: Part (a) holds because $X_j = -2\log\lambda_j$, $\sum_{j=1}^m \log\lambda_j = -m\log n + m\log(2\pi) + \sum_{j=1}^m \log j$ by the definition of λ_j, and $\sum_{j=1}^m \log j = m\log m - m + O(\log m)$ by estimating the sum by integrals. Parts (b), (c), and (d) are established by estimating sums by integrals. In particular, for part (d) we use

$$\sum_{j=1}^{m} j^k \log j = \frac{1}{k+1} m^{k+1}\log m - \frac{1}{(k+1)^2} m^{k+1} + o(m^{k+1}). \tag{7.9}$$

The proofs of parts (e) and (f) are straightforward using the definition of R_j in (3.3) and the fact that $\sum_{j=1}^m j^k = m^{k+1}/(k+1) + O(m^k)$ for any integer $k \geq 1$.

We now prove part (g). By Lemma 5 of HDB, $\limsup_{n\to\infty} \sup_{1\leq j\leq m} E\log^2(I_j/f_j) < \infty$. By Jensen's inequality, $\{E\log(I_j/f_j)\}^2 \leq E\log^2(I_j/f_j)$. Thus,

$$\max_{1\leq j\leq m} |E\varepsilon_j| = \max_{1\leq j\leq m} |E\log(I_j/f_j) + C| = O(1). \tag{7.10}$$

Furthermore, by Lemma 6 of HDB, there exists a constant $C < \infty$ such that

$$|E\varepsilon_j| \leq C\frac{\log j}{j} \quad \text{for all } \log^2 m \leq j \leq m \text{ and } n \text{ sufficiently large.} \tag{7.11}$$

We have $\sum_{j=[\log^2 m]+1}^{m} \log(j)/j = O(\log^2 m)$ by bounding the sum by an integral. Combining these results gives

$$\text{(7.12)} \qquad \sum_{j=1}^{m} |E\varepsilon_j| = \sum_{j=1}^{[\log^2 m]} |E\varepsilon_j| + \sum_{j=[\log^2 m]+1}^{m} |E\varepsilon_j| = O(\log^2 m).$$

Part (h) is proved using (7.9)–(7.11):

$$\text{(7.13)} \qquad \left|Z_k' E\varepsilon\right| \le \left|\sum_{j=1}^{[\log^2 m]} (j/m)^k E\varepsilon_j\right| + \left|\sum_{j=[\log^2 m]+1}^{m} (j/m)^k E\varepsilon_j\right|$$
$$= O(1) \sum_{j=1}^{[\log^2 m]} (j/m)^k + O\left(\frac{1}{m^k}\right) \sum_{j=[\log^2 m]+1}^{m} j^{k-1} \log j = O(\log m). \qquad \text{Q.E.D.}$$

PROOF OF LEMMA 2: Part (a) is established as follows:

$$\text{(7.14)} \qquad X^{*\prime} X^* = (X + 1_m 2\log(2\pi/n))'(X + 1_m 2\log(2\pi/n)) - (1_m'(X + 1_m 2\log(2\pi/n)))^2/m$$
$$= 4\sum_{j=1}^{m} \log^2 j - (-2m\log m + 2m + O(\log m))^2/m,$$

using Lemma 1(a). In addition, $\sum_{j=1}^{m} \log^2 j = m\log^2 m - 2m\log m + 2m + O(\log^2 m)$ by estimating the sum by an integral. These results combine to give part (a).

Part (b) follows from Lemma 1(b) and (c). Part (c) follows from Lemma 1(a), (b), and (d).

For part (d), by the definition of R_j in (3.3) and Lemma 2(c), we have

$$\text{(7.15)} \qquad X^{*\prime} R = 1(s \ge 2+2r) \frac{b_{2+2r}}{(2+2r)!} X^{*\prime} Z_{2+2r}^* (2\pi m/n)^{2+2r} + \sum_{j=1}^{m} (X_j - \overline{X}) Rem_j^*$$
$$= -1(s \ge 2+2r) \frac{2(2\pi)^{2+2r}(2+2r) b_{2+2r}}{(3+2r)!(3+2r)} \frac{m^{3+2r}}{n^{2+2r}} (1+o(1)) + \sum_{j=1}^{m} (X_j - \overline{X}) Rem_j^*.$$

Next, we have

$$\text{(7.16)} \qquad \left|\frac{1}{2}\sum_{j=1}^{m} (X_j - \overline{X}) Rem_j^*\right| = \left|\sum_{j=1}^{m} \left(\log m - \log j - 1 + O\left(\frac{\log m}{m}\right)\right) Rem_j^*\right|$$
$$\le O(1) \sum_{j=1}^{m} \lambda_j^q \left(\log m - \log j + \left|1 + O\left(\frac{\log m}{m}\right)\right|\right) = O\left(\frac{m^{q+1}}{n^q}\right),$$

using Lemma 1(a), the triangle inequality, (7.9), $\max_{1\le j\le m} |Rem_j^*/\lambda_j^q| = O(1)$, and $\sum_{j=1}^{m} j^q = m^{q+1}/(q+1) + O(m^q)$. Equations (7.15) and (7.16) combine to establish part (d). When s is an integer, $O(1)$ and $O(m^{q+1}/n^q)$ are $o(1)$ and $o(m^{q+1}/n^q)$, respectively, in (7.16) because $\max_{1\le j\le m} |Rem_j^*/\lambda_j^q| = o(1)$ in (3.3).

Part (e) is established by applying Lemma 1(b), (e), and (f) to $Z_k^{*\prime} R = Z_k' R - (1/m) 1_m' Z_k 1_m' R$.

To prove part (f), we write $|X_j^*| \le |X_j + 2\log(2\pi/n)| + |\overline{X} + 2\log(2\pi/n)|$. Then, by Lemma 1(a), we obtain $\max_{j=1,\dots,m} |X_j^*| = O(\log m)$. This and Lemma 1(g) give the desired result: $|X^{*\prime} E\varepsilon| \le O(\log m) 1_m' |E\varepsilon| = O(\log^3 m)$.

Part (g) is established using Lemma 1(b), (g), and (h): $Z_k^{*\prime} E\varepsilon = Z_k' E\varepsilon - (1/m) 1_m' Z_k 1_m' E\varepsilon = O(\log^2 m)$. Parts (h) and (i) hold by Lemma 2(b) and (c) and the definitions of Γ_r and μ_r. Part (j) holds by Lemma 2(a), (h), and (i) using the definition of c_r.

Lastly, we establish part (k). Using $\max_{1\le j\le m} |X_j^*| = O(\log m)$ (proved above) and Lemma 2(h) and (i), we obtain

$$\text{(7.17)} \qquad M_{Z^*} X^* = X^* - Z^*(Z^{*\prime} Z^*)^{-1} Z^{*\prime} X^* = O(\log m) 1_m + 2Z^* \Gamma_r^{-1} \mu_r (1+o(1)).$$

Thus, $\max_{1\le j\le m} |[M_{Z^*} X^*]_j| = O(\log m) + O(1) = O(\log m)$. Q.E.D.

The proofs of Theorems 2 and 3 use the following lemma, of which part (a) is a variant of Theorem 2 of Robinson (1995b). Part (a) is also a variant of Lemma 3 and other results stated on p. 23 of HDB. Part (b) is a variant of (3.5) and (3.6) of GRS. Define

$$\omega(\lambda) = \frac{1}{(2\pi n)^{1/2}} \sum_{t=1}^{n} Y_t \exp(i\lambda_j t) \quad \text{and} \quad u(\lambda) = \omega(\lambda)/f^{1/2}(\lambda). \tag{7.18}$$

LEMMA 3: (a) *Suppose Assumption 2 holds. Then, there exist constants* $0 < \alpha < 1$ *and* $C_\alpha < \infty$ *such that for all* $1 \leq k < j \leq n^\alpha$ *and all* $n \geq 1$,

(i) $|Eu(\lambda_k)\overline{u(\lambda_k)} - 1| \leq C_\alpha \frac{\log(k+1)}{k}$,
(ii) $|Eu(\lambda_k)u(\lambda_k)| \leq C_\alpha \frac{\log(k+1)}{k}$,
(iii) $|Eu(\lambda_j)\overline{u(\lambda_k)}| \leq C_\alpha \frac{\log(j+1)}{k}$, *and*
(iv) $|Eu(\lambda_j)u(\lambda_k)| \leq C_\alpha \frac{\log(j+1)}{k}$.

(b) *The results* (*i*)–(*iv*) *of part* (*a*) *hold uniformly over* $f \in \mathcal{F}(s, a, \delta, K)$.

PROOF OF LEMMA 3: A density f that satisfies our Assumption 2 satisfies Assumptions 1 and 2 of Robinson (1995b). In consequence, results (i)–(iv) of part (a) follow from Theorem 2 of Robinson (1995b) using the normalization of $\omega(\lambda)$ by $f^{1/2}(\lambda)$ rather than $g(0)^{1/2}|\lambda|^{-d}$. The remainder term in (i) is different from that in Robinson (1995b) because the proof only requires (4.1), and not (4.2), of Robinson (1995b) to hold. Also, the results of part (a) are stronger than those stated in Theorem 2 of Robinson (1995b) because of the uniformity of the bounds over all $1 \leq k < j \leq n^\alpha$, but Robinson's proof still gives the desired results.

Part (b) follows by inspection of the proof of Theorem 2 of Robinson (1995b) using the following condition in place of his Assumption 2: For all $0 < \lambda_1 < \lambda_2 \leq \tilde{\delta} = \min\{\delta_2, \delta_3\}$,

$$|f(\lambda_1) - f(\lambda_2)| \leq C|\lambda_1|^{-1-2d_f}|\lambda_1 - \lambda_2| \tag{7.19}$$

for some constant $C < \infty$ that is independent of $f \in \mathcal{F}(s, a, \delta, K)$. (This condition is used to show that the left-hand side of (4.6) of Robinson (1995b) is $O(\varepsilon_{jj})$. It is also used for similar calculations in the proofs of parts (b)–(d) of Theorem 2 of Robinson (1995b).)

The condition in (7.19) holds for any $f \in \mathcal{F}(s, a, \delta, K)$ by the following calculation:

$$\begin{aligned} |f(\lambda_1) - f(\lambda_2)| &= |\lambda_1^{-2d_f} g(\lambda_1) - \lambda_2^{-2d_f} g(\lambda_2)| \\ &\leq \lambda_1^{-2d_f}|g(\lambda_1) - g(\lambda_2)| + |g(\lambda_2)| \cdot |\lambda_1^{-2d_f} - \lambda_2^{-2d_f}| \\ &\leq K_3 \lambda_1^{-2d_f}|\lambda_1 - \lambda_2| + 2|d_f| C_5 \lambda_1^{-1-2d_f}|\lambda_1 - \lambda_2| \end{aligned} \tag{7.20}$$

for all $0 < \lambda_1 < \lambda_2 \leq \tilde{\delta}$, where the second inequality holds using condition (iii) of $\mathcal{F}(s, a, \delta, K)$, a mean value expansion of $\lambda_1^{-2d_f}$ about λ_2, and the fact that $\sup_{0 \leq \lambda \leq \delta_2} |g(\lambda)| \leq C_5$ for some constant $C_5 < \infty$ by condition (ii) of $\mathcal{F}(s, a, \delta, K)$.

In the division of the domain of integration of the integrals in the proof of Theorem 2 of Robinson (1995b), we replace his ε by $\tilde{\delta}$.

Note that we impose the restriction $|d_f| \leq 1/2 - \delta_1$ in the definition of $\mathcal{F}(s, a, \delta, K)$, whereas GRS allow $|d_f| < 1/2$, because in the proof of part (b) of the theorem we use it to obtain uniformity of the convergence results. In particular, we were not able to verify that Theorem 2 of Robinson (1995b) holds uniformly without this restriction, although GRS state that it does. Our difficulty came in verifying that the stated order of the integral $\int_{\lambda_j/2}^{\infty}$ in the eleventh to last line on p. 1061 and the integrals $\int_{-\lambda_j/2}^{\lambda_j/2}$ in the first and last equations on p. 1062 hold uniformly over $|d_f| < 1/2$. The difficulty in the former case is bounding

$$\sup_{-1/2 < d_f < 0} n^{-1} \lambda_j^{d-1/2} \int_{\lambda_j/2}^{\gamma} \lambda^{-((3/2)+d_f)} \, d\lambda$$

by $C\log(j+1)/j$ for arbitrary $\gamma > 0$. In the latter case the difficulty is that $\sup_{0<d_f<1/2}\int_0^{\lambda_j/2}\lambda^{-2d_f}\,d\lambda = \infty$.

We impose the condition $\int_{-\pi}^{\pi} f(\lambda)\,d\lambda \le K_1$ in $\mathcal{F}(s,a,\delta,K)$ because it is needed on line 7 of p. 1061 of Robinson's (1995b) proof of Theorem 2 to obtain uniformity of the results over $f \in \mathcal{F}(s,a,\delta,K)$. *Q.E.D.*

PROOF OF THEOREM 2: Part of the proof of Theorem 2 is analogous to the proof of Theorem 2 in HDB and part uses an alteration of the asymptotic normality proof of (5.14) in Robinson (1995b). The quantities $S := X^{*\prime}M_{Z^*}X^*$ and $A_j := [M_{Z^*}X^*]_j$ play the roles of $4S_{xx}$ and $-2a_j$ in HDB, respectively, and ε_j is the same as in HDB.

Using (3.7), (3.8), $1_m'X^* = 0$, $1_m'Z^* = 0$, and Lemma 2(j), we have

$$(7.21)\qquad m^{1/2}(\hat{d}_r - d) = m^{1/2}(X^{*\prime}M_{Z^*}X^*)^{-1}X^{*\prime}M_{Z^*}(R+\varepsilon) = m^{1/2}(X^{*\prime}M_{Z^*}X^*)^{-1}X^{*\prime}M_{Z^*}R + (1+o(1))\frac{c_r}{4m^{1/2}}\sum_{j=1}^{m}[X^{*\prime}M_{Z^*}]_j\varepsilon_j.$$

By Lemma 2(j) and (7.4), when $s \ge 2+2r$,

$$(7.22)\qquad m^{1/2}(X^{*\prime}M_{Z^*}X^*)^{-1}X^{*\prime}M_{Z^*}R = O\left(\frac{m^{2.5+2r}}{n^{2+2r}}\right) + O\left(\frac{m^{q+0.5}}{n^q}\right) = o(1),$$

using Assumption 3 to obtain the second equality. By Lemma 2(j) and (7.5), when $s < 2+2r$,

$$(7.23)\qquad m^{1/2}(X^{*\prime}M_{Z^*}X^*)^{-1}X^{*\prime}M_{Z^*}R = O\left(\frac{m^{s+0.5}}{n^s}\right) = o(1),$$

using Assumption 3 to obtain the second equality.

Hence, it suffices to show that

$$(7.24)\qquad m^{-1/2}\sum_{j=1}^{m}A_j\varepsilon_j \to_d N\left(0, \frac{4}{c_r}\frac{\pi^2}{6}\right).$$

We write

$$(7.25)\qquad m^{-1/2}\sum_{j=1}^{m}A_j\varepsilon_j = T_1+T_2+T_3, \quad \text{where} \quad T_1 := m^{-1/2}\sum_{j=1}^{\log^8 m}A_j\varepsilon_j,$$
$$T_2 := m^{-1/2}\sum_{j=1+\log^8 m}^{m^{0.5+\delta}}A_j\varepsilon_j, \quad \text{and} \quad T_3 := m^{-1/2}\sum_{j=1+m^{0.5+\delta}}^{m}A_j\varepsilon_j$$

for some $0 < \delta < 0.5$. (Here and below, for notational simplicity, we let $\log^8 m$ and $m^{0.5+\delta}$ denote the integer parts of these expressions.) The proofs in HDB that $T_1 = o_p(1)$ and $T_2 = o_p(1)$ also are valid in our case, because $\max_{1\le j\le m}|A_j| = O(\log m)$ by Lemma 2(k).

The remainder of the proof (i.e., showing the asymptotic normality of T_3) differs from that in HDB, because following the line of argument in HDB leads to a restriction on the growth rate of m that is excessive for our purposes, although not for theirs. In particular, HDB's proof relies on Robinson's (1995b) asymptotic normality result (5.14), which uses the third part of his Assumption 6 with $\alpha=\beta=2$ in the proof of his (5.14) and this requires $m = o(n^{4/5})$. (The third part of Robinson's Assumption 6 is used in the first two equations on p. 1068 of Robinson (1995b), which are part of the proof of (5.14).) Note that Robinson's (1995b) α equals $\min\{s,2\}$ in our notation, so that a large value of s does not increase α above 2.

Instead of using the method of HDB, we show that $T_3c_r^{1/2}/2 \to_d N(0,\pi^2/6)$ by altering the asymptotic normality result given in (5.14) of Robinson (1995b). (An alternative method of establishing this result would be to utilize results in Soulier (2001).) Robinson's (5.14) states that

$$(7.26)\qquad m^{-1/2}\sum_{j=1+m^{0.5+\delta}}^{m}a_jU_j \to_d N(0,\Omega_U) \quad \text{as} \quad n\to\infty,$$

where $\{a_j : j = 1, \ldots, m; m \geq 1\}$ is a triangular array of constants that satisfies (i) $\max_{j\in\{1+m^{0.5+\delta},\ldots,m\}} |a_j| = o(m)$, (ii) $m^{-1}\sum_{j=1+m^{0.5+\delta}}^{m} a_j^2 = 1 + o(1)$, (iii) $m^{-1}\sum_{j=1+m^{0.5+\delta}}^{m} |a_j|^p = O(1)$ for all $p \geq 1$, $\{U_j : j = 1, \ldots, m\}$ are the random variables defined by

$$U_j := \varepsilon_j + \log\{g(\lambda_j)/g(0)\}, \tag{7.27}$$

using our notation, and Ω_U is defined by $m^{-1}\sum_{j=1+m^{0.5+\delta}}^{m} U_j^2 \to_p \Omega_U$. (This is actually a special case of Robinson's result when his $J = 1$, $l = m^{0.5+\delta}$, and $\nu = 1$.)

We alter this result by replacing the U_j by ε_j and prove the altered result using an alteration of the proof given by Robinson. Specifically, the altered result is

$$m^{-1/2} \sum_{j=1+m^{0.5+\delta}}^{m} a_j \varepsilon_j \to_d N(0, \Omega_\varepsilon) \quad \text{as} \quad n \to \infty, \tag{7.28}$$

where Ω_ε is defined by $m^{-1}\sum_{j=1+m^{0.5+\delta}}^{m} \varepsilon_j^2 \to_p \Omega_\varepsilon$, provided our Assumptions 2 and 3 hold and the constants $\{a_j : j = 1, \ldots, m; m \geq 1\}$ satisfy conditions (i)–(iii).

Robinson's proof of (7.26) relies on writing U_j as a function of $v_g(\lambda)$ (using his notation), which is the discrete Fourier transform (his $w_g(\lambda)$) divided by $C_g|\lambda|^{-d_g}$ $(=g^{1/2}(0)|\lambda|^{-d}$ in our notation). That is, $U_j = h(v_g(\lambda))$ for some function $h(\cdot)$. Simple calculations show that ε_j is the same function $h(\cdot)$ of the discrete Fourier transform $w_g(\lambda)$ divided by $f_{gg}^{1/2}(\lambda)$ $(=f^{1/2}(\lambda) = g^{1/2}(\lambda)|\lambda|^{-d}$ in our notation). In consequence, if we alter Robinson's definition of $v_g(\lambda)$ to be $w_g(\lambda)/f_{gg}^{1/2}(\lambda)$, then all of his proof goes through as stated except the first two equations on p. 1068, which depend on the properties of $v_g(\lambda)$ (through the second moment matrix $Ev_j^* v_k^{*\prime}$). The requisite properties of $v_g(\lambda)$ for these two equations are established in Theorem 2 of Robinson (1995b). For our altered definition of $v_g(\lambda)$, the results hold by Lemma 3(a). In consequence, the properties of the altered $v_g(\lambda)$ needed in the first two equations of p. 1068 hold without imposing the third part of Robinson's Assumption 6. (We also note that the second part of Robinson's Assumption 6 is satisfied when $\ell = m^{0.5+\delta}$, although this part of his Assumption 6 is not used in his proof of (7.26).) This completes the proof of (7.28).

Now, we apply (7.28) with $a_j := A_j c_r^{1/2}/2$. Condition (i) on $\{a_j\}$ holds by Lemma 2(k). Condition (ii) on $\{a_j\}$ holds because

$$\sum_{j=1+m^{0.5+\delta}}^{m} A_j^2 = \sum_{j=1}^{m} A_j^2 - \sum_{j=1}^{m^{0.5+\delta}} A_j^2 = 4m/c_r + o(m), \tag{7.29}$$

using Lemma 2(j) and (k). Condition (iii) on $\{a_j\}$ holds because

$$\begin{aligned} \sum_{j=1+m^{0.5+\delta}}^{m} |A_j|^p &\leq 2^{p-1} \sum_{j=1+m^{0.5+\delta}}^{m} |X_j^*|^p + 2^{p-1} \sum_{j=1+m^{0.5+\delta}}^{m} |\widetilde{Z}_j^{*\prime}(Z^{*\prime}Z^*)^{-1}Z^{*\prime}X^*|^p \\ &= O(m) \end{aligned} \tag{7.30}$$

for all $p \geq 1$, where $\widetilde{Z}_j^*$ denotes the jth row of Z^* written as a column vector of dimension r. The equality in (7.30) uses (A18) of HDB for the term involving X_j^* and Lemma 2(h) and (i), plus the fact that the absolute values of the elements of $\widetilde{Z}_j^*$ are bounded by one for all j, for the other term. Hence, conditions (i)–(iii) on $\{a_j\}$ hold. Using Lemmas 6 and 7 of HDB, we find that the asymptotic covariance matrix Ω_ε is $\pi^2/6$. This completes the proof. *Q.E.D.*

PROOF OF THEOREM 3: The proof of part (a) is a variant of the proof of Theorem 1 of GRS. We take their least favorable densities $\{f_n : n \geq 1\}$, which are discontinuous, and adjust them to be continuous and satisfy a Lipschitz condition. Call the adjusted densities $\{f_n^* : n \geq 1\}$. We show that (1) $f_n^* \in \mathcal{F}(s, a, \delta, K)$ for n sufficiently large and (2) the result of Lemma 1(ii) of GRS holds for $\{f_n^* : n \geq 1\}$. The latter result is

$$\int_{-\pi}^{\pi} (f_n^*(\lambda) - f_0(\lambda))^2 \, d\lambda \leq C_0 n^{-1} \quad \text{for some constant} \quad C_0 < \infty, \tag{7.31}$$

where $f_0(\lambda) = 1$ for $\lambda \in [-\pi, \pi]$. In consequence, the results of Lemma 2 of GRS hold with f_n replaced by f_n^*, because the proof of their Lemma 2 holds for any sequence of densities that satisfies the result of their Lemma 1(ii), is bounded away from zero, and is in L^2 (which implies square summability of the corresponding covariances). The densities f_n^* satisfy these conditions.

Next, GRS's proof of their Theorem 1 holds for any sequence of densities $\tilde{f}_n$ that satisfies the results of their Lemma 2 and is of the form

$$\tilde{f}_n(\lambda) = \tilde{c}_n|\lambda|^{-h_n}(1+\widetilde{\Delta}_n(\lambda)), \tag{7.32}$$

where $\{\tilde{c}_n : n \geq 1\}$ are bounded constants, $h_n = \kappa n^{-s/(2s+1)}$ for some $\kappa > 0$, and $|\widetilde{\Delta}_n(\lambda)| \leq \widetilde{K}|\lambda|^s$ for all λ in a neighborhood of zero for some $\widetilde{K} < \infty$. We show below that the densities f_n^* are of this form and, hence, the result of Theorem 1 of GRS holds for the class of functions $\mathscr{F}(s, a, \delta, K)$, which establishes part (a) of our Theorem 3.

The reason that we consider the densities f_n^*, rather than f_n as in GRS, is that the functions f_n do not satisfy Robinson's (1995b) Assumption 2 nor condition (iii) of $\mathscr{F}(s, a, \delta, K)$. This is relevant because the proof of part (b) of our Theorem 3 relies on the lemmas of HDB, which in turn rely on the proof of Theorem 2 of Robinson (1995b). The latter utilizes his Assumption 2 that the spectral density f is differentiable in a neighborhood of zero with derivative that is $O(|\lambda|^{-1-2d_f})$ as $\lambda \to 0$. Robinson's Assumption 2 can be avoided if one takes the remainder in each part of his Theorem 2 to have the additional term $O((j/n)^s)$; see GRS and Giraitis, Robinson, and Samarov (2000). This additional term, however, is too large for our purposes. Instead, we replace Robinson's Assumption 2 with the Lipschitz condition (iii) of $\mathscr{F}(s, a, \delta, K)$. Lemma 3 shows that this condition is sufficient to obtain the desired analogues of Robinson's Theorem 2. The least favorable functions f_n used in GRS do not satisfy condition (iii) of $\mathscr{F}(s, a, \delta, K)$, so we replace them by the functions f_n^*, which do.

GRS's function f_n is an even function that is defined as follows:

$$\begin{aligned} f_n(\lambda) &= \begin{cases} c_n\lambda^{-h_n} & \text{for } \lambda \in (0, \delta_n], \\ 1 & \text{for } \lambda \in (\delta_n, \pi], \end{cases} \\ &= c_n\lambda^{-h_n}(1+\Delta_n(\lambda)), \quad \text{where} \\ h_n &= \kappa\delta_n^s \quad \text{for some} \quad \kappa > 0, \quad \delta_n = n^{-1/(2s+1)}, \quad c_n = 1+\log\delta_n^{h_n}, \quad \text{and} \\ \Delta_n(\lambda) &= \begin{cases} 0 & \text{for } \lambda \in (0, \delta_n], \\ c_n^{-1}\lambda^{h_n} - 1 & \text{for } \lambda \in (\delta_n, \pi]. \end{cases} \end{aligned} \tag{7.33}$$

(Note that in the notation of GRS s is β.)

We define the function f_n^* to be an even function that equals f_n on $(0, \delta_n]$ and equals the constant $f_n(\delta_n) = c_n\delta_n^{-h_n}$ on $(\delta_n, \pi]$. In consequence, f_n^* is continuous on $(0, \pi]$. We can write f_n^* as follows:

$$\begin{aligned} f_n^*(\lambda) &= \begin{cases} c_n\lambda^{-h_n} & \text{for } \lambda \in (0, \delta_n], \\ c_n\delta_n^{-h_n} & \text{for } \lambda \in (\delta_n, \pi], \end{cases} \\ &= c_n\lambda^{-h_n}(1+\Delta_n^*(\lambda)) \\ &= \lambda^{-h_n}g_n^*(\lambda), \quad \text{where} \\ \Delta_n^*(\lambda) &= \begin{cases} 0 & \text{for } \lambda \in [0, \delta_n], \\ (\lambda/\delta_n)^{h_n} - 1 & \text{for } \lambda \in (\delta_n, \pi], \end{cases} \quad \text{and} \\ g_n^* &= c_n(1+\Delta_n^*(\lambda)). \end{aligned} \tag{7.34}$$

Now we show that $f_n^* \in \mathscr{F}(s, a, \delta, K)$ for n large and κ small (where κ appears in the definition of h_n). First, $|h_n/2| \leq 1/2 - \delta_1$ for n large, because $h_n \to 0$ as $n \to \infty$. Second, we have $|\Delta_n^*(\lambda)| \leq |\Delta_n(\lambda)| \leq K_2\lambda^s$ for all $\lambda \in (0, \pi]$ for κ sufficiently small. The first inequality holds because $c_n\delta_n^{-h_n} \leq 1$, as shown below in (7.41). The second inequality holds by Lemma 1(i) of GRS. Thus, f_n^* is of the form (7.32) and g_n^* satisfies condition (ii) of $\mathscr{F}(s, a, \delta, K)$ with $g_k = 0$ for $k = 1, \dots, [s/2]$ for any $\delta_2 \in (0, \pi]$.

Third, we have $g_n^*(0) = c_n$. Below we show that $c_n \to 1$ as $n \to \infty$, so condition (i) of $\mathscr{F}(s, a, \delta, K)$ is satisfied for n sufficiently large provided $a_0 < 1 < a_{00}$. If the latter condition is not satisfied, then f_n^* can be rescaled by multiplication by $(a_0 + a_{00})/2$ so that $g_n^*(0) = c_n(a_0 + a_{00})/2 \to (a_0 + a_{00})/2 \in (a_0, a_{00})$.

Fourth, we show that g_n^* satisfies condition (iii) of $\mathscr{F}(s, a, \delta, K)$ for n large. If $\lambda_1, \lambda_2 \in (0, \delta_n]$, then $g_n^*(\lambda_1) - g_n^*(\lambda_2) = 0$, so condition (iii) holds. If $\lambda_1, \lambda_2 \in [\delta_n, \pi]$, then

$$\begin{aligned}(7.35)\qquad |g_n^*(\lambda_1) - g_n^*(\lambda_2)| &= |\lambda_1^{h_n} f_n^*(\lambda_1) - \lambda_2^{h_n} f_n^*(\lambda_2)| \\ &\le |\lambda_1^{h_n}| \cdot |f_n^*(\lambda_1) - f_n^*(\lambda_2)| + |f_n^*(\lambda_2)| \cdot |\lambda_1^{h_n} - \lambda_2^{h_n}| \\ &\le \zeta h_n \lambda_*^{h_n - 1} |\lambda_1 - \lambda_2|,\end{aligned}$$

where λ_* lies between λ_1 and λ_2, the first inequality holds by the triangle inequality, and the second inequality holds using $f_n^*(\lambda_1) = f_n^*(\lambda_2)$, $|f_n^*(\lambda_2)| \le \zeta$ for all n for some constant $\zeta \ge 1$, which follows from $\delta_n^{h_n} \to 1$ and $c_n \to 1$ as $n \to \infty$, as shown below, and a mean value expansion of $\lambda_1^{h_n}$ about λ_2. We have

$$\begin{aligned}(7.36)\qquad h_n \lambda_*^{h_n - 1} &\le h_n \delta_n^{h_n - 1} = (h_n/\delta_n)(1 + o(1)) \\ &= \kappa n^{-(s-1)/(2s+1)}(1 + o(1)) = \kappa O(1) \le K_3/\zeta\end{aligned}$$

for n large and κ small, where the first equality holds because $\delta_n^{h_n} \to 1$ as $n \to \infty$, as shown below. Equations (7.35) and (7.36) combine to yield $|g_n^*(\lambda_1) - g_n^*(\lambda_2)| \le K_3 |\lambda_1 - \lambda_2|$ for n sufficiently large. If $\lambda_1 \in (0, \delta_n]$ and $\lambda_2 \in (\delta_n, \pi]$, then

$$\begin{aligned}(7.37)\qquad |g_n^*(\lambda_1) - g_n^*(\lambda_2)| &\le |g_n^*(\lambda_1) - g_n^*(\delta_n)| + |g_n^*(\delta_n) - g_n^*(\lambda_2)| \\ &\le K_3 |\lambda_1 - \lambda_2|,\end{aligned}$$

where the second inequality holds by the previous results for $\lambda_1, \lambda_2 \in (0, \delta_n]$ and $\lambda_1, \lambda_2 \in [\delta_n, \pi]$. Thus, g_n^* satisfies condition (iii) of $\mathscr{F}(s, a, \delta, K)$ for n large.

Fifth, we have

$$\begin{aligned}(7.38)\qquad \int_{-\pi}^{\pi} f_n^*(\lambda)\, d\lambda/2 &= \int_0^{\delta_n} c_n \lambda^{-h_n}\, d\lambda + \int_{\delta_n}^{\pi} c_n \delta_n^{-h_n}\, d\lambda \\ &= \frac{c_n \delta_n^{1-h_n}}{1 - h_n} + c_n \delta_n^{-h_n}(\pi - \delta_n) = \pi + o(1),\end{aligned}$$

because $c_n \to 1$, $\delta_n \to 0$, $h_n \to 0$, and $\delta_n^{h_n} \to 1$ as $n \to \infty$, as shown below. If multiplication of f_n^* by $(a_0 + a_{00})/2$ is necessary for the third point above, then the right-hand side in (7.38) is $\pi(a_0 + a_{00})/2 + o(1)$, which is less than or equal to $K_1/2$ for n large because $2\pi a_{00} \le K_1$ by assumption.

Next, we show that (7.31) holds. Let $\xi_n = 1 - c_n \delta_n^{-h_n}$. We show below that $\xi_n = O(n^{-2s/(2s+1)} \log^2 n)$. We have

$$\begin{aligned}(7.39)\qquad \int_{-\pi}^{\pi} (f_n^*(\lambda) - f_0(\lambda))^2\, d\lambda &= 2\int_0^{\delta_n} (f_n(\lambda) - f_0(\lambda))^2\, d\lambda + 2\int_{\delta_n}^{\pi} \xi_n^2\, d\lambda \\ &\le C_0 n^{-1} + O(n^{-4s/(2s+1)} \log^4 n)\end{aligned}$$

for some constant $C_0 > 0$, where the inequality uses Lemma 1(ii) of GRS for the bound on the first term. Thus, (7.31) holds provided $s > 1/2$, which is assumed.

We now show that $\delta_n^{h_n} \to 1$ and $c_n \to 1$ as $n \to \infty$ and $\xi_n = O(n^{-2s/(2s+1)} \log^2 n)$. Because $\log(n^{\kappa_1 n^{-\gamma}}) = \kappa_1 n^{-\gamma} \log n \to 0$ for any $\kappa_1 \in R$ and $\gamma > 0$, we have $n^{\kappa_1 n^{-\gamma}} \to 1$ as $n \to \infty$. Taking $\kappa_1 = -\kappa/(2s+1)$ and $\gamma = s/(2s+1)$, this gives $\delta_n^{h_n} \to 1$ as $n \to \infty$. In turn, this implies that

$c_n = 1 + \log \delta_n^{h_n} \to 1$ as $n \to \infty$. Next, we can write ξ_n as a function of $\delta_n^{h_n}$: $\xi_n = 1 - c_n \delta_n^{-h_n} = (\delta_n^{h_n} - 1 - \log \delta_n^{h_n})/\delta_n^{h_n}$. By a mean value expansion about $x = 0$, $n^x = 1 + (n^{x_1} \log n)x$ for x_1 between 0 and x. Taking $x = \kappa_1 n^{-\gamma}$ for κ_1 and γ as above, this gives

$$\delta_n^{h_n} = n^{\kappa_1 n^{-\gamma}} = 1 + (n^{x_{1n}} \log n)\kappa_1 n^{-\gamma} = 1 + O(n^{-s/(2s+1)} \log n), \tag{7.40}$$

where x_{1n} lies between 0 and $\kappa_1 n^{-\gamma}$.

A Taylor expansion of $\log x$ about $x = 1$ gives $\log x = x - 1 - (x-1)^2/(2x_*^2)$, where x_* lies between x and 1. Thus, $0 < (x - 1 - \log x)/x = O((x-1)^2)$ as $x \to 1$. Taking $x = \delta_n^{h_n}$ and using (7.40) gives

$$0 < \xi_n = \big(\delta_n^{h_n} - 1 - \log \delta_n^{h_n}\big)/\delta_n^{h_n} = O\big((\delta_n^{h_n} - 1)^2\big) = O\big(n^{-2s/(2s+1)} \log^2 n\big). \tag{7.41}$$

Next, we prove part (b) of the theorem. The function $\log(1+x)$ satisfies

$$\log(1+x) = \sum_{w=1}^{[s/2]} \frac{(-1)^{w-1}}{w} x^w + \gamma(x), \tag{7.42}$$

where $|\gamma(x)| \le C_1 x^{[s/2]+1}$ for all $|x| \le x_0$ for some fixed $x_0 > 0$ and $C_1 < \infty$. Let g be a function that satisfies conditions (i) and (ii) of $\mathscr{F}(s, a, \delta, K)$. Then,

$$\log(g(\lambda)/g(0)) = \log(1 + h(\lambda)) = \sum_{w=1}^{[s/2]} \frac{(-1)^{w-1}}{w} h(\lambda)^w + \gamma(h(\lambda)), \quad \text{where} \tag{7.43}$$

$$h(\lambda) = \sum_{k=1}^{[s/2]} (g_k/g(0))\lambda^{2k} + \Delta(\lambda)/g(0).$$

Multiplying out the $h(\lambda)^w$ terms and rearranging gives

$$\log(g(\lambda)/g(0)) = \sum_{k=1}^{[s/2]} \tilde{g}_k \lambda^{2k} + \zeta(\lambda), \tag{7.44}$$

where the real numbers $\tilde{g}_k$ and $\zeta(\lambda)$ are defined implicitly and $\zeta(\lambda)$ satisfies $|\zeta(\lambda)| \le C_2 \lambda^s$ for all $0 < \lambda \le \lambda_0$ for some constants $C_2 < \infty$ and $\lambda_0 > 0$ that do not depend on q, but may depend on (s, a, δ, K). Hence, an expansion of $\log(g_j/g_0)$ of the form (3.2) holds with remainder $\zeta(\lambda_j)$ that satisfies $|\zeta(\lambda_j)| \le C_2 \lambda_j^s$ uniformly over functions g that satisfy conditions (i)–(iii) of $\mathscr{F}(s, a, \delta, K)$.

Now, the result of part (b) holds for the estimator $\hat{d}_{r, m_n}$ for a single density $f \in \mathscr{F}(s, a, \delta, K)$ and a single sequence $\{m_n : m_n \in J_n(s, D_0), n \ge 1\}$ by Theorem 1, because (i) Assumption 1 holds, (ii) Assumption 2 holds except that g is not necessarily smooth of order $s \ge 1$, (iii) the expansion (7.44) holds, which is of the form of (3.2), (iv) the proof of Theorem 1 goes through with Assumption 2 replaced by (3.2) or by an expansion of this form, such as (7.44), (v) Theorem 2 of Robinson (1995b), which is utilized in the proof of Theorem 1, can be replaced by Lemma 3(b) in the proof of Theorem 1, and (vi) the restriction $r \ge (s-2)/2$ implies that $E_f \hat{d}_{r, m_n} - d_f = O(m^s/n^s) = O(n^{-s/(2s+1)})$ and $\mathrm{var}_f(\hat{d}_{r, m_n}) = O(m^{-1}) = O(n^{-2s/(2s+1)})$. Hence, it suffices to show that the results of Theorem 1 hold uniformly over $f \in \mathscr{F}(s, a, \delta, K)$ and $m \in J_n(s, D_0)$. This can be seen by inspection of the proof of Theorem 1 plus the proofs of Lemmas 5 and 6 and Theorem 1 of HDB, using Lemma 3(b) in place of Theorem 2 of Robinson (1995b) in the proof of Theorem 1 and using the uniformity of (7.44) over functions g that satisfy conditions (i)–(iii) of $\mathscr{F}(s, a, \delta, K)$.

Note that we impose the condition $g(0) \ge a_0$ in $\mathscr{F}(s, a, \delta, K)$ so that $g_k/g(0)$ in (7.44) is uniformly bounded and $\zeta(\lambda)$ in (7.44) satisfies $|\zeta(\lambda)| \le C_2 \lambda^s$. *Q.E.D.*

PROOF OF THEOREM 4: We prove part (a) first. We have

$$\begin{aligned} D_{n,r}(\hat{b}(r) - b(r)) &= (D_{n,r}^{-1} Q^{*\prime} M_{X^*} Q^* D_{n,r}^{-1})^{-1} D_{n,r}^{-1} Q^{*\prime} M_{X^*} \log I - D_{n,r} b(r) \\ &= (Z^{*\prime} M_{X^*} Z^*)^{-1} Z^{*\prime} M_{X^*} (K 1_m + X^* d + Q^* b(r) + R + \varepsilon) \\ &\quad - D_{n,r} b(r) \\ &= (Z^{*\prime} M_{X^*} Z^*)^{-1} Z^{*\prime} M_{X^*} (R + \varepsilon), \end{aligned} \tag{7.45}$$

where the second equality holds using $Q^* D_{n,r}^{-1} = Z^*$ and (3.7) and the third equality holds using $M_{X^*} 1_m = 0$ and $Q^* = Z^* D_{n,r}$.

Next, we have

$$Z^{*\prime} M_{X^*} Z^* = (\Gamma_r - \mu_r \mu_r') m(1 + o(1)) \tag{7.46}$$

by Lemma 2(a), (h), and (i). In addition,

$$Z^{*\prime} X^* (X^{*\prime} X^*)^{-1} X^{*\prime} R = 1(s \geq 2 + 2r) \rho_r \mu_r b_{2+2r} \frac{m^{3+2r}}{n^{2+2r}} (1 + o(1)) + O\left(\frac{m^{q+1}}{n^q}\right) \tag{7.47}$$

by Lemma 2(a), (d), and (i). Combining (7.3) and (7.47) gives

$$Z^{*\prime} M_{X^*} R = 1(s \geq 2 + 2r) \rho_r (\xi_r - \mu_r) b_{2+2r} \frac{m^{3+2r}}{n^{2+2r}} (1 + o(1)) + O\left(\frac{m^{q+1}}{n^q}\right). \tag{7.48}$$

In the previous two equations, the $O(\cdot)$ term is actually $o(\cdot)$ if s is an integer.

By Lemma 2(a), (f), (g), and (i),

$$Z^{*\prime} M_{X^*} E\varepsilon = O(\log^3 m). \tag{7.49}$$

Combining (7.45)–(7.49) gives part (a).

To establish part (b), we calculate the asymptotic variance of $c' D_{n,r} \hat{b}(r)$ for arbitrary $c \in R^r$. By (7.45),

$$\text{var}(c' D_{n,r} \hat{b}(r)) = \text{var}(c' (Z^{*\prime} M_{X^*} Z^*)^{-1} Z^{*\prime} M_{X^*} \varepsilon). \tag{7.50}$$

As in the proof of Theorem 1, we use the proof of Theorem 1 of HDB. We replace their $4S_{xx}^2$ by m^2 and a_j by $[mc' (Z^{*\prime} M_{X^*} Z^*)^{-1} Z^{*\prime} M_{X^*}]_j$, where $[v]_j$ denotes the jth element of the vector v. It suffices to show that the two conditions of (7.7) hold. We have

$$\sum_{j=1}^m a_j^2 = m^2 c' (Z^{*\prime} M_{X^*} Z^*)^{-1} c = mc' (\Gamma_r - \mu_r \mu_r')^{-1} c (1 + o(1)) \tag{7.51}$$

using (7.46). So, the second condition of (7.7) holds with $K_0 = c' (\Gamma_r - \mu_r \mu_r')^{-1} c$.

To show that the first condition of (7.7) holds, note that $mc' (Z^{*\prime} M_{X^*} Z^*)^{-1} = O(1)$ by (7.46). In consequence, it suffices to show that

$$\max_{1 \leq j \leq m} \left| [Z_{2k}^{*\prime} M_{X^*}]_j \right| = O(\log m) \quad \text{for} \quad k = 1, \ldots, r. \tag{7.52}$$

We have $|[Z_{2k}^*]_j| = (j/m)^{2k} \leq 1$. In addition,

$$\max_{1 \leq j \leq m} \left| [Z_{2k}^{*\prime} X^* (X^{*\prime} X^*)^{-1} X^{*\prime}]_j \right| \leq \left| Z_{2k}^{*\prime} X^* (X^{*\prime} X^*)^{-1} \right| \max_{1 \leq j \leq m} |X_j^*| = O(\log m), \tag{7.53}$$

where the second equality holds because $\left| Z_{2k}^{*\prime} X^* (X^{*\prime} X^*)^{-1} \right| = O(1)$ by Lemma 2(a) and (c) and $\max_{1 \leq j \leq m} |[X^*]_j| = O(\log m)$ by the proof of Lemma 2(f). These results combine to establish (7.52). *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 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.

BHANSALI, R. J., AND P. S. KOKOSZKA (1997): "Estimation of the Long Memory Parameter by Fitting Fractional Autoregressive Models," Working Paper.

DELGADO, M., AND P. M. ROBINSON (1996): "Optimal Spectral Bandwidth for Long Memory," *Statistica Sinica*, 6, 97–112.

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

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.

GIRAITIS, L., AND P. M. ROBINSON (2000): "Edgeworth Expansions for Semiparametric Whittle Estimation of Long Memory," Working Paper, London School of Economics.

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," Ch. 38 in *Handbook of Econometrics, Volume 4*, ed. by R. F. Engle and D. McFadden. New York: Elsevier.

HENRY, M. (2000): "Robust Automatic Bandwidth for Long Memory," *Journal of Time Series Analysis*, 22, 293–316.

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 Analysis, Volume II: Time Series Analysis in Memory of E. J. Hannan*, ed. by P. M. Robinson and M. Rosenblatt. New York: Springer-Verlag, 220–232.

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 K. I. BELTRAO (1994): "Automatic Semiparametric Estimation of the Memory Parameter of a Long-Memory Time Series," *Journal of Time Series Analysis*, 15, 285–302.

HURVICH, C. M., AND J. BRODSKY (2001): "Broadband Semiparametric Estimation of the Memory Parameter of a Long-memory Time Series Using Fractional Exponential Models," *Journal of Time Series Analysis*, 22, 221–249.

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., R. S. DEO, AND J. BRODSKY (1998): "The Mean Squared Error of Geweke and Porter-Hudak's Estimator of the Memory Parameter of a Long-Memory Time Series," *Journal of Time Series Analysis*, 19, 19–46.

IOUDITSKY, A., E. MOULINES, AND P. SOULIER (2001): "Adaptive Estimation of the Fractional Differencing Coefficient," *Bernoulli*, 7, 699–731.

KIM, C. S., AND P. C. B. PHILLIPS (1999a): "Log Periodogram Regression: The Non-stationary Case," Working Paper, Cowles Foundation, Yale University.

——— (1999b): "Modified Log Periodogram Regression," Working Paper, Cowles Foundation, Yale University.

MOULINES, E., AND P. SOULIER (1999): "Broad Band Log-Periodogram Regression of Time Series with Long Range Dependence," *Annals of Statistics*, 27, 1415–1439.

ROBINSON, P. M. (1994): "Semiparametric Analysis of Long-memory Time Series," *Annals of Statistics*, 22, 515–539.

——— (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 (2003): "Higher-order Kernel Semiparametric M-estimation of Long Memory," *Journal of Econometrics*, forthcoming.

SOULIER, P. (2001): "Moment Bounds and Central Limit Theorem for Functions of Gaussian Vectors," *Statistics and Probability Letters*, 54, 193–203.

VELASCO, C. (1999): "Non-stationary Log-periodogram Regression," *Journal of Econometrics*, 91, 325–271.

——— (2000): "Non-Gaussian Log-Periodogram Regression," *Econometric Theory*, 16, 44–79.