COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1263

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

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

Donald W. K. Andrews and Patrik Guggenberger

June 2000

# A Bias-reduced Log-periodogram Regression Estimator for the Long-memory Parameter

Donald W. K. Andrews and Patrik Guggenberger
*Cowles Foundation for Research in Economics*
*Yale University*

September 1999
Revised: May 2000

**Abstract**

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 propose a simple bias-reduced log-periodogram regression estimator, $\widehat{d}_r$, that eliminates the first- and higher-order biases of the GPH estimator. The bias-reduced estimator is the same as the GPH estimator except that one includes frequencies to the power $2k$ for $k = 1, ..., r$, 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, which is consistent with the semiparametric nature of the long-memory model under consideration.

Following the work of Robinson (1995b) and Hurvich, Deo, and Brodsky (1998), we establish the asymptotic bias, variance, and mean-squared error (MSE) of $\widehat{d}_r$, determine the MSE optimal choice of the number of frequencies, $m$, to include in the regression, and establish the asymptotic normality of $\widehat{d}_r$. These results show that the bias of $\widehat{d}_r$ 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. In consequence, the optimal rate of convergence to zero of the MSE of $\widehat{d}_r$ is faster than that of the GPH estimator.

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

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

*Keywords*: Asymptotic bias, asymptotic normality, bias reduction, frequency domain, long-range dependence, optimal rate, rate of convergence, strongly dependent time series.

*JEL Classification Numbers*: C13, C14, C22.

## 1 Introduction

We consider a semiparametric model for a stationary Gaussian long-memory time series $\{Y_t : t = 1, ..., n\}$. 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 $(-\pi, \pi)$. 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.

In this paper, we investigate the asymptotic and finite sample properties of a new bias-reduced log-periodogram estimator $\widehat{d}_r$ of the long memory parameter $d$. Let $\lambda_j = 2\pi j/n$ for $j = 1, ..., [n/2]$ denote the fundamental frequencies for a sample of size $n$. The estimator $\widehat{d}_r$ 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 $\lambda_j$ on a constant, $-2\log\lambda_j$, and $\lambda_j^2, \lambda_j^4, ..., \lambda_j^{2r}$ for $j = 1, ..., m$, where $r$ is a (fixed) non-negative integer. We take $m$ such that $m \to \infty$ and $m/n \to 0$ as $n \to \infty$. When $r = 0$, $\widehat{d}_r$ is asymptotically equivalent to the well-known Geweke and Porter-Hudak (1983) (GPH) estimator $\widehat{d}_{GPH}$.

The motivation for the estimator $\widehat{d}_r$ 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 $r = 2$.

We determine the asymptotic bias, variance, and MSE of $\widehat{d}_r$, calculate the MSE optimal choice of $m$ for $\widehat{d}_r$, and establish the asymptotic normality of $\widehat{d}_r$. 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 $\widehat{d}_r$ is of order $m^{2+2r}/n^{2+2r}$ provided $g$ is sufficiently smooth, whereas that of $\widehat{d}_{GPH}$ is of the larger order $m^2/n^2$. The asymptotic variances of $\widehat{d}_r$ and $\widehat{d}_{GPH}$ are both of order $m^{-1}$. In consequence, the optimal rate of convergence to zero of the MSE of $\widehat{d}_r$ is of order $n^{-(4+4r)/(5+4r)}$ whereas that of $\widehat{d}_{GPH}$ is of the larger order $n^{-4/5}$. For example, for $r = 2$, $n^{-(4+4r)/(5+4r)} = n^{-12/13}$. The rate of convergence of $\widehat{d}_r$ also exceeds that of the local Whittle estimator, see Robinson (1995a), and the average periodogram estimator, see Robinson (1994) and Lobato and Robinson (1996), provided $g$ is sufficiently smooth.

We find that $m^{1/2}(\widehat{d}_r - d)$ is asymptotically normal with mean zero provided $m = o(n^{(4+4r)/(5+4r)})$. In contrast, $\widehat{d}_{GPH}$ 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 $\widehat{d}_r$ 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 utilize results of Giraitis, Robinson, and Samarov (1997) (GRS).

Some Monte Carlo simulations show that the bias-reduced estimators $\widehat{d}_1$ and $\widehat{d}_2$ have lower biases, higher standard deviations, and lower root mean squared errors (rmse's) compared to the standard log-periodogram estimator $\widehat{d}_0$ for a variety of stationary Gaussian ARFIMA(1, d, 1) processes, as the asymptotic results suggest. The lower biases lead to good confidence interval (CI) coverage probabilities for CI's based on $\widehat{d}_1$ and $\widehat{d}_2$ over a wider range of $m$ values than for $\widehat{d}_0$. On the other hand, the lower standard deviation of $\widehat{d}_0$ leads to shorter CI intervals than CI's based on $\widehat{d}_1$ and $\widehat{d}_2$. The rmse graphs for $\widehat{d}_1$ and $\widehat{d}_2$ are flatter than those for $\widehat{d}_0$, which implies that $\widehat{d}_1$ and $\widehat{d}_2$ are less sensitive to the choice of $m$ than is $\widehat{d}_0$. The results are essentially the same for the three values of $d$ considered: $-.4$, 0, and .4. The basic pattern of results are the same for sample sizes $n = 512$ and $n = 2,048$. We conclude that for stationary Gaussian ARFIMA(1, d, 1) processes the estimators $\widehat{d}_1$ and $\widehat{d}_2$ deliver bias and rmse reductions in finite samples that reflect the asymptotic results.

Some papers in the literature that are related to this paper include Hurvich and Brodsky (1997), Bhansali and Kokoszka (1997), Moulines and Soulier (1999, 2000), and Hurvich (2000). 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 $\widehat{d}_r$, 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.

Other papers in the literature that are related to this paper include Delgado and Robinson (1996), Henry and Robinson (1996), Hurvich and Deo (1998), and Henry (1999). 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 differ from the present paper because they do not consider bias-reduced estimation of $d$ based on these regressions.

The results of this paper can be extended in a number of different directions. First, the bias-reduction method utilized here can be extended to a number of other procedures. Andrews and Sun (1999) consider a bias-reduced local polynomial Whittle estimator. The bias-reduction method also could be applied to the average-periodogram estimator of Robinson (1994) and Lobato and Robinson (1996), the pooled, trimmed, 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 (1998a), the pooled log-periodogram estimator of Shimotsu and Phillips (1999), the modified log-periodogram estimator of Kim and

Phillips (1999b) for non-stationary time series, the tapered log-periodogram estimator of Velasco (1998b) for non-stationary 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 non-stationary time series, along the lines of Kim and Phillips (1999a), and with stochastic volatility models, as in Deo and Hurvich (1999).

The remainder of this paper is organized as follows. Section 2 motivates the bias-reduced estimator by briefly reviewing results in the literature for the GPH estimator. Section 3 introduces the bias-reduced log-periodogram estimator, establishes its asymptotic bias, variance, and MSE, and shows that it is asymptotically normal. Section 4 gives the optimal rate of convergence results. Section 5 describes the Monte Carlo simulation results. An Appendix provides proofs of results given in the paper.

## 2 The GPH Estimator

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

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

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

$$I_j = \frac{1}{2\pi n}|\sum_{t=1}^{n} Y_t \exp(i\lambda_j t)|^2 \text{ for } j = 1, ..., m, \tag{2.2}$$

where $\lambda_j = 2\pi j/n$ and $m$ is a positive integer smaller than $n$. Their estimator, which we refer to as the GPH estimator, is given by $-1/2$ times the LS estimator of the slope parameter in a regression of $\{\log I_j : j = 1, ..., m\}$ 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

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

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

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

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

where $\varepsilon_j = \log(I_j/f_j) + C$, $f_j = f(\lambda_j)$, $f_j^* = f^*(\lambda_j)$, $f_0^* = f^*(0)$, and $C = 0.577216...$ is the Euler constant. Equation (2.4) is a pseudo-regression model. If the pseudo-errors

$\{\log(f_j^*/f_0^*) + \varepsilon_j : j = 1, ..., m\}$ behave like independent and identically distributed (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 parametrization in (1.1) and, hence, replaces the regressor $\widetilde{X}_j$ by

$$X_j = -2\log\lambda_j \tag{2.5}$$

(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, or mean squared error or the asymptotic 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) and used by Robinson (1995b) 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 mean-squared error (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 regressor $\widetilde{X}_j$ replaced by $X_j$ ( and correspondingly with the $-0.5$ term dropped in the definition of $\hat{d}_{GPH}$).

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

**Assumption HDB1.** $m = m(n) \to \infty$ and $\frac{m\log m}{n} \to 0$ 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

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

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

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

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

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 $\lambda_j^2$ to the pseudo-regression model (2.4). Furthermore, additional bias terms can be eliminated by adding the regressors $\lambda_j^4, ..., \lambda_j^{2r}$ 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 $\widehat{d}_r$, 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 $\widehat{d}_r$ is the LS estimator of the coefficient on $X_j$ from the regression of $\log I_j$ on 1, $X_j$, $\lambda_j^2, \lambda_j^4, ..., \lambda_j^{2r}$ for $j = 1, ..., m$ for some non-negative integer $r$. It is defined explicitly in (3.8) below. Note that only even powers of $\lambda_j$ are employed in the regression. Odd powers of $\lambda_j$ 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 $\widehat{d}_r$, 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 $\lambda$ 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 $\int_{-\pi}^{\pi} |\lambda|^{-2d} g(\lambda) d\lambda < \infty$.

Assumption 2 allows one to develop an $[s]$ term Taylor expansion of $\log g(\lambda_j)$

about $\lambda = 0$:

$$\begin{aligned}\log(g_j/g_0) &= \sum_{k=1}^{[s]} \frac{b_k}{k!}\lambda_j^k + O(\lambda_j^s), \text{ where} \\ g_j &= g(\lambda_j),\ g_0 = g(0), \\ b_k &= \frac{d^k}{d\lambda^k}\log g(\lambda)\Big|_{\lambda=0}, \end{aligned} \tag{3.1}$$

and $O(\cdot)$ holds uniformly over $j = 1, ..., m$ and all $n \geq 1$.

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 $k \leq [s]$. In consequence,

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

and $O(\cdot)$ holds uniformly over $j = 1, ..., m$ and all $n \geq 1$. 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 $\lambda_j^{2k}$ for $k = 1, ..., r$ and the remainder:

$$\begin{aligned}\log(g_j/g_0) &= \sum_{k=1}^{\min\{[s/2],r\}} \frac{b_{2k}}{(2k)!}\lambda_j^{2k} + R_j, \text{ where} \\ R_j &= \sum_{k=\min\{[s/2],r\}+1}^{[s/2]} \frac{b_{2k}}{(2k)!}\lambda_j^{2k} + O(\lambda_j^s) \\ &= 1(s \geq 2+2r)\frac{b_{2+2r}}{(2+2r)!}\lambda_j^{2+2r} + O(\lambda_j^q), \\ q &= \min\{s, 4+2r\}, \end{aligned} \tag{3.3}$$

and $O(\cdot)$ is uniform over $j = 1, ..., m$ and all $n$. If $s$ is an integer, then (3.1)–(3.3) hold with $O(\cdot)$ replaced by $o(\cdot)$ throughout by the continuity of the $s$-th order derivative of $g$.

Let

$$Q_{k,j} = \lambda_j^k \text{ for } j = 1, ..., m \text{ and } k = 1, 2, ... . \tag{3.4}$$

Let $\log I$, $X$, $Q_k$, $R$, and $\varepsilon$ denote column $m$-vectors whose $j$-th elements are $\log I_j$, $X_j$, $Q_{k,j}$, $R_j$, and $\varepsilon_j$, respectively, where $\varepsilon_j = \log(I_j/f_j) + C$ and $f_j = f(\lambda_j)$. Let $Q$ denote the $m \times r$ matrix whose $k$-th column is $Q_{2k}$ for $k = 1, ..., r$. Let $1_m$ denote a column $m$-vector of ones. Let $b$ denote the column $r$-vector whose $k$-th element is $b_{2k}/(2k)!$ for $k = 1, ..., \min\{[s/2], r\}$ and 0 for $k = \min\{[s/2], r\} + 1, ..., r$. 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:

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

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

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

The pseudo-regression model in deviation from mean form is

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

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, we have

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

(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 $Q^* b$ in (3.7), which includes the $\lambda_j^{2k}$ terms for $k = 1, ..., \min\{[s/2], r\}$ 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 $\mu_r$ be a column $r$-vector with $k$-th element $\mu_{r,k}$ and $\Gamma_r$ be an $r \times r$ matrix with $(i, k)$ element given by $[\Gamma_r]_{i,k}$, where

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

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

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

For example, $c_0 = 1$, $c_1 = 9/4$, $c_2 = 3.52$, $c_3 = 4.79$, $c_4 = 6.06$, and $c_5 = 7.33$.

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

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

Let

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

For example, $\tau_0 = -2.19$, $\tau_1 = 2.23$, $\tau_2 = -0.793$, $\tau_3 = .146$, $\tau_4 = -0.0164$, and $\tau_5 = .00125$.

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

*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 $n^\gamma$ 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 $n^\gamma$ 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.** Theorem 1 also holds when the regressor $X_j$ is replaced by $\widetilde{X}_j$ and $\hat{d}_r$ is defined to be $-0.5$ times the LS coefficient on the regressor $\widetilde{X}_j$ from the regression of $\log I_j$ on 1, $\widetilde{X}_j$, $\lambda_j^2, \lambda_j^4, ..., \lambda_j^{2r}$ for $j = 1, ..., m$. 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).

**4.** If the condition in Assumption 2 that $g$ is smooth of order $s$ at zero is replaced by the assumption that $g$ satisfies the expansion of (3.2) for some constants $\{b_{2k} : 1 \leq k \leq [s/2]\}$ and the Lipschitz condition: $|g(\lambda_1) - g(\lambda_2)| \leq K_3|\lambda_1 - \lambda_2|$ for all $0 < \lambda_1 < \lambda_2 \leq \delta_3$ for some constants $K_3 < \infty$ and $\delta_3 > 0$, then the results of the Theorem still hold.

**5.** The proof of Theorem 1 relies on the Lemmas of HDB that are used to prove their Theorem 1. The proof is given in the Appendix. The proof of Comment 4 is given in the second to last paragraph of the proof of Theorem 3. Note that in Comment 4 the Lipschitz condition is used in place of Assumption 2 of Robinson (1995b).

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

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

and $m \sim n^{2\phi/(2\phi+1)}$ means $m/n^{2\phi/(2\phi+1)}$ converges to a finite positive constant as $n \to \infty$.

For this choice,

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

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

If $s$ and $r$ are arbitrarily large, then $MSE(\hat{d}_r)$ converges to zero at a rate arbitrarily close to the rate $n^{-1}$ of parametric estimators.

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 $m \propto n^\gamma$ 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

$$\begin{aligned} MSE(\hat{d}_r) &= \tau_r^2 b_{2+2r}^2 \frac{m^{4+4r}}{n^{4+4r}}(1+o(1)) + O(\frac{m^{1+2r}\log^3 m}{n^{2+2r}}) \\ &\quad + \frac{\pi^2}{24}\frac{c_r}{m}(1+o(1)). \end{aligned} \tag{3.16}$$

(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:

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

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).

## 3.2 Asymptotic Normality

We now show that the bias-reduced estimator $\widehat{d_r}$ 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:

**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,*

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

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

The restriction on $m$ for the GPH estimator is $m = o(n^{4/5})$ in the asymptotic normality result of HDB (which corresponds to $r = 0$ and $s \geq 2$ in Assumption 3.) For example, for $n$ =100, 300, and 1000, the term $n^{4/5}$ equals 40, 96, and 251 respectively. In contrast, Assumption 3 requires $m = o(n^{(4+4r)/(5+4r)})$ when $r \geq 1$ and $s \geq 2 + 2r$. For the same values of $n$, the term $n^{(4+4r)/(5+4r)}$ equals 60, 159, and 464 when $r = 1$ and 70, 193, and 588 when $r = 2$.

**2.** When $s = 3$, the assumptions on $g$ in Theorem 2 are the same as those used in HDB to obtain the asymptotic normality of $\widehat{d}_{GPH}$. For this choice of $s$, the restriction on $m$ in Assumption 3 is $m = o(n^{6/7})$ for any $r \geq 1$, whereas the restriction on $m$ in HDB for asymptotic normality of $\widehat{d}_{GPH}$ is $m = o(n^{4/5})$. Hence, under the same conditions as in HDB, $\widehat{d}_r$ is asymptotically normal with a faster rate of convergence than $\widehat{d}_{GPH}$ for an appropriate choice of $m$.

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

**4.** For the processes considered in Section 5, we find by simulation that the variance of $m^{1/2}(\widehat{d}_r - d)$ can be estimated more accurately in finite samples by $\pi^2/(6X^{*\prime}M_{Q^*}X^*/m)$, which is the usual regression variance expression with regression error variance $\pi^2/6$, than by $\pi^2 c_r/24$, which equals $\lim_{n\rightarrow\infty} \pi^2/(6X^{*\prime}M_{Q^*}X^*/m)$ (see Lemma 2(j) of the Appendix).

# 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) 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 $\widehat{d}_r$ 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 $\widehat{d}_r$. 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 $k = 1, ..., [s]$. For this class of spectral densities, 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 equal to zero at $\lambda = 0$. For example, a fractionally differenced autoregressive moving average process has non-zero derivatives 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.

When we expand the class of functions to include functions $g(\lambda)$ that are smooth of order $s$ and have non-zero derivatives 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 $a = (a_0, a_{00}, a_1, ..., a_{[s/2]})'$, $\delta = (\delta_1, \delta_2, \delta_3)'$, and $K = (K_1, K_2, K_3)'$ be positive finite constants with $a_0 < a_{00}$ and $\delta_1 < 1/2$. We consider the following class of spectral densities:

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

If $g$ is an even function on $[-\pi, \pi]$ that is smooth of order $s \geq 1$ 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 $a$, $\delta$, 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 $k = 1, ..., [s/2]$ 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

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

The optimal rate results are given in the following Theorem:

**Theorem 3** *Let $s$ and the elements of $a = (a_0, a_{00}, a_1, ..., a_{[s/2]})'$, $\delta = (\delta_1, \delta_2, \delta_3)'$, 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*

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

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

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

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

*where $\widehat{d}_{r,m}$ denotes the bias-reduced estimator $\widehat{d}_r$ 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 $p$-th power absolute error loss function $\ell(x) = |x|^p$ for any $p > 0$. The upper bound on the risk of the bias-reduced estimator $\widehat{d}_r$ given in part (b) is for the quadratic loss function. This result implies a similar result for any loss function $\ell(\cdot)$ for which $E_f\ell(n^{s/(2s+1)}(\widehat{d}_{r,m} - d_f)) \leq h(n^{s/(2s+1)}(E_f(\widehat{d}_{r,m} - d_f)^2)^{1/2})$ 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 $p$-th 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, $\infty$, 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 Monte Carlo Experiment

## 5.1 Experimental Design

In this section, we compare the finite sample behavior of the estimators $\widehat{d}_0$, $\widehat{d}_1$, and $\widehat{d}_2$. The estimator $\widehat{d}_0$ is the standard log-periodogram estimator, whereas $\widehat{d}_1$ and $\widehat{d}_2$ are bias-reduced log-periodogram estimators. We consider stationary Gaussian

ARFIMA$(1, d, 1)$ processes with autoregressive parameter (AR) $\phi$ and moving average (MA) parameter $\theta$. When $d = 0$, the model is

$$Y_t = \phi Y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1} \text{ for } t = 1, ..., n, \tag{5.1}$$

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

We consider the processes that correspond to all possible combinations of

$$\begin{aligned} d &= 0,\ .4,\ -.4, \\ \phi &= .9,\ .6,\ .3,\ 0,\ -.3,\ -.6,\ -.9, \text{ and} \\ \theta &= .9,\ .6,\ .3,\ 0,\ -.3,\ -.6,\ -.9. \end{aligned} \tag{5.2}$$

We consider sample sizes $n = 512$ and $2,048$ (which facilitate use of the fast Fourier transform to compute the periodogram.) Because the results are quite similar for a wide variety of different parameter combinations, we only explicitly report results for a small subset of the parameter combinations.

We use $1,000$ simulation repetitions for each parameter combination. For each parameter combination that is actually reported in the Figures below, however, we use $20,000$ simulation repetitions. The differences between using $1,000$ and $20,000$ repetitions are quite small.

We calculate the biases, standard deviations, and root mean squared errors (rmse's) of $\widehat{d}_0$, $\widehat{d}_1$, and $\widehat{d}_2$ as functions of $m$ for $m = 4, 5, ..., n/2$. 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% confidence intervals (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 $m^{1/2}(\widehat{d}_r - d)$ 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

$$\left[ \widehat{d}_r - z_{.95} \left( \frac{\pi^2}{6 X^{*\prime} M_{Q^*} X^*} \right)^{1/2},\ \widehat{d}_r + z_{.95} \left( \frac{\pi^2}{6 X^{*\prime} M_{Q^*} X^*} \right)^{1/2} \right] \text{ for } r = 0, 1, 2, \tag{5.3}$$

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.

## 5.2 Simulation Results

We discuss the results for $d = 0$ and $n = 512$ first. We find that for any given positive AR parameter $\phi$ 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 non-negative values of the AR parameter and these results can be well summarized by considering the parameter combinations in which the MA parameter $\theta$ 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 report results for $\theta = -.9$, which yields a MA(1) process with positive autocorrelation. In sum, with little loss of generality, we report in Figures 1–4 results for the following parameter combinations: $(\phi, \theta) = (.9, 0)$, $(.6, 0)$, $(.3, 0)$, and $(0, -.9)$, respectively, where in each case $d = 0$ and $n = 512$.

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 $\widehat{d}_0$ grows very rapidly as $m$ increases, whereas the biases of $\widehat{d}_1$ and especially $\widehat{d}_2$ grow much more slowly. It is apparent that the bias-reducing features of $\widehat{d}_1$ and $\widehat{d}_2$ 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 $\widehat{d}_0$ is less than that of $\widehat{d}_1$ and $\widehat{d}_2$ 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 $(\phi, \theta)$ 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 smaller for $\widehat{d}_1$ and $\widehat{d}_2$ than for $\widehat{d}_0$, which is in accord with the asymptotic results. In addition, one sees that the rmse graph for $\widehat{d}_0$ rises very steeply from its minimal value, whereas the rmse graphs for $\widehat{d}_1$ and $\widehat{d}_2$ rise more slowly. In consequence, $\widehat{d}_1$ and $\widehat{d}_2$ have low rmse's over wider ranges of $m$ values and, hence, are not as sensitive to the choice of $m$ as $\widehat{d}_0$.

The CI coverage probability graphs in Figure 1(d) show that $\widehat{d}_0$ has true coverage probability close to .9 only for very small values of $m$. This is due to the bias of $\widehat{d}_0$ for larger values of $m$. In contrast, the coverage probabilities of $\widehat{d}_1$ and $\widehat{d}_2$ are close to .9 for a wider range of values of $m$, due to their smaller biases. Thus, $\widehat{d}_1$ and $\widehat{d}_2$ yield CI's that are more robust to the choice of $m$ than does $\widehat{d}_0$. On the other hand, the larger standard deviations of $\widehat{d}_1$ and $\widehat{d}_2$ lead to larger average lengths of their CI's than those of $\widehat{d}_0$, 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 $\widehat{d}_0$ increases more slowly as $m$ increases than in Figure 1(a). The biases of $\widehat{d}_1$ and $\widehat{d}_2$ 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 $\widehat{d}_1$ and $\widehat{d}_2$ in Figure 2(c) are slightly lower than those of $\widehat{d}_0$ due to their lower biases. In addition, the rmse graphs for $\widehat{d}_1$ and $\widehat{d}_2$ are quite flat in Figure 2(c), which implies that a wide range of values of $m$ yield low rmse.

Figure 3 provides results for the AR(1) model with AR parameter .3. As expected, the biases of all three estimators are reduced further from those reported in Figures 1 and 2. In fact, the biases of $\widehat{d}_1$ and $\widehat{d}_2$ are quite small across the entire range of $m$ values. As a result, the coverage probabilities of the CI's based on $\widehat{d}_1$ and $\widehat{d}_2$ are quite robust to the choice of $m$—much more so than $\widehat{d}_0$. The rmse's of all three estimators are fairly flat, which indicates that all three are relatively robust to the choice of $m$.

Figure 4 reports results for the MA(1) model with MA parameter $-.9$. The bias of $\widehat{d}_1$ is negative in this case, so we graph the bias of $-\widehat{d}_1$, which equals the negative of the bias of $\widehat{d}_1$ because $d = 0$. The results of Figure 4 are similar to those of Figure 3 except that the bias of $\widehat{d}_1$ is negative and the bias and rmse of $\widehat{d}_0$ rise more sharply for large values of $m$. Again the biases of $\widehat{d}_1$ and $\widehat{d}_2$ 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 $m$.

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 $m$.

Finally, we discuss the results for the larger sample size $n = 2,048$. The results are relatively easy to describe. For every $(\phi, \theta)$ parameter combination, the bias and coverage probability graphs are very similar to their $n = 512$ counterparts and the standard deviation, rmse, and average CI length graphs are very similar in shape to those for $n = 512$ but are shifted down toward the horizontal axis. In consequence, for brevity, we only report results for the parameter combination $(\phi, \theta) = (.9, 0)$, $d = 0$, and $n = 2,048$. See Figure 5. The similarity of the bias graphs for $n = 512$ and $n = 2,048$ 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 $n = 2,048$ 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 $2,048$ 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 $2,048$.

In sum, the Monte Carlo simulation results show that $\widehat{d}_1$ and $\widehat{d}_2$ have lower biases, higher standard deviations, and lower rmse's compared to $\widehat{d}_0$ for a wide range of stationary Gaussian ARFIMA(1, d, 1) processes, as the asymptotic results suggest. The lower biases lead to good CI coverage probabilities for $\widehat{d}_1$ and $\widehat{d}_2$ over a wider range of $m$ values than for $\widehat{d}_0$. On the other hand, the lower standard deviation of $\widehat{d}_0$ leads to shorter CI intervals than CI's based on $\widehat{d}_1$ and $\widehat{d}_2$. The rmse graphs for $\widehat{d}_1$ and $\widehat{d}_2$ are flatter than those for $\widehat{d}_0$, which implies that $\widehat{d}_1$ and $\widehat{d}_2$ are less sensitive to the choice of $m$ than is $\widehat{d}_0$. The results are essentially the same for the three values of $d$ considered: $-.4$, 0, and .4. The basic pattern of results are the same for sample sizes $n = 512$ and $n = 2,048$.

## 6 Appendix of Proofs

The estimator $\widehat{d}_r$ 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

$$Z_{k,j} = (j/m)^k. \tag{6.1}$$

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

$$Z^* = Z - 1_m \overline{Z}', \text{ where } \overline{Z} = \frac{1}{m} Z' 1_m. \tag{6.2}$$

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* $i, k = 1, 2, ...,$ *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* $O(m^{q+1}/n^{q+1})$ *replaced by* $o(m^{q+1}/n^{q+1})$.

**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* $i, k = 1, 2, ...,$ *provided* $\Gamma_r$ *is nonsingular in parts* (j) *and* (k). *When* $s$ *is an integer, the results in parts* (d) *and* (e) *hold with* $O(m^{q+1}/n^{q+1})$ *replaced by* $o(m^{q+1}/n^{q+1})$.

**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 $\xi_r$,

$$Z^{*\prime}R = \frac{(2\pi)^{2+2r}(2+2r)b_{2+2r}}{(3+2r)!(3+2r)}\xi_r\frac{m^{3+2r}}{n^{2+2r}}(1+o(1)) + O(\frac{m^{q+1}}{n^q}). \tag{6.3}$$

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

$$X^{*'}M_{Z^*}R = -\frac{2(2\pi)^{2+2r}(2+2r)}{(3+2r)!(3+2r)}(1-\mu_r'\Gamma_r^{-1}\xi_r)b_{2+2r}\frac{m^{3+2r}}{n^{2+2r}}(1+o(1)) + O(\frac{m^{q+1}}{n^q})). \tag{6.4}$$

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

$$X^{*'}M_{Z^*}R = X^{*\prime}R - X^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}R = O(\frac{m^{s+1}}{n^s}). \tag{6.5}$$

Finally, by Lemma 2(f)-(i),

$$X^{*'}M_{Z^*}E\varepsilon = O(\log^3 m) + 2\mu_r'\Gamma_r^{-1}1_r(1+o(1))O(\log^2 m) = O(\log^3 m). \tag{6.6}$$

Equations (3.9), (6.4), and (6.6), Lemma 2(j), and the definition of $\tau_r$ combine to establish part (a) when $s \geq 2+2r$. Equations (3.9), (6.5), and (6.6), 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 (6.3)–(6.5) 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 $S_{xx}$ by $X^{*'}M_{Z^*}X^* = (4m/c_r)(1+o(1))$ 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 sequence $\{a_j : j = 1, ..., m\}$ for which

$$\max_{1\leq j\leq m}|a_j| = O(\log m) \text{ and } \sum_{j=1}^{m} a_j^2 = K_0 m(1+o(1)) \text{ for some } K_0 > 0. \tag{6.7}$$

In our case, $a_j = [M_{Z^*}X^*]_j$ and $\sum_{j=1}^{m} a_j^2 = X^{*'}M_{Z^*}X^* = (4m/c_r)(1+o(1))$ by Lemma 2(j). The first condition of (6.7) holds by Lemma 2(k). The second holds with $K_0 = 4m/c_r$. From the proof of Theorem 1 of HDB, we have

$$Var(\hat{d}_r) = \frac{1}{(X^{*\prime}M_{Z^*}X^*)^2}\frac{4m}{c_r}\frac{\pi^2}{6}(1+o(1)) = \frac{\pi^2}{24}\frac{c_r}{m}(1+o(1)). \tag{6.8}$$

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. □

**Proof of Lemma 1.** To prove part (a), by the definition of $\lambda_j$, we have

$$\sum_{j=1}^{m} \log \lambda_j = -m \log n + m \log(2\pi) + \sum_{j=1}^{m} \log j. \tag{6.9}$$

We estimate the sum by integrals:

$$\begin{aligned} \sum_{j=1}^{m} \log j &\geq \int_{1}^{m} \log x dx = m \log m - m + 1 \text{ and} \\ \sum_{j=1}^{m} \log j &\leq \int_{2}^{m+1} \log x dx = (m+1)(\log(m+1) - 1) - 2\log 2 + 2. \end{aligned} \tag{6.10}$$

Thus,

$$\sum_{j=1}^{m} \log j = m \log m - m + O(\log m). \tag{6.11}$$

Combining $X_j = -2 \log \lambda_j$, (6.9), and (6.11) establishes part (a).

The proofs of parts (b) and (c) are straightforward.

To prove part (d), we write

$$\sum_{j=1}^{m} (X_j + 2\log(2\pi/n)) Z_{k,j} = -2 \sum_{j=1}^{m} \frac{j^k}{m^k} \log j. \tag{6.12}$$

Again we estimate a sum by integrals:

$$\int_{1}^{m} x^k \log x \; dx \leq \sum_{j=1}^{m} j^k \log j \leq \int_{2}^{m+1} x^k \log x \; dx. \tag{6.13}$$

Because $\int_a^b x^k \log x \; dx = [\frac{1}{k+1} x^{k+1} \log x - \frac{1}{(k+1)^2} x^{k+1}]_a^b$, we obtain

$$\sum_{j=1}^{m} j^k \log j = \frac{1}{k+1} m^{k+1} \log m - \frac{1}{(k+1)^2} m^{k+1} + o(m^{k+1}). \tag{6.14}$$

Equations (6.12) and (6.14) combine to establish part (d).

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,

$$\max_{1\leq j\leq m} |E\varepsilon_j| = \max_{1\leq j\leq m} |E \log(I_j/f_j) + C| = O(1). \tag{6.15}$$

Furthermore, by Lemma 6 of HDB,

$$|E\varepsilon_j| = O(\frac{\log j}{j}) \text{ uniformly for } \log^2 m \leq j \leq m. \tag{6.16}$$

We have

$$\sum_{j=\log^2 m+1}^{m} \frac{\log j}{j} \leq \int_{\log^2 m}^{m} \frac{\log x}{x} dx = \frac{1}{2} [\log^2 x]_{\log^2 m}^{m} \leq \log^2 m. \tag{6.17}$$

Hence,

$$\sum_{j=1}^{m} |E\varepsilon_j| = \sum_{j=1}^{\log^2 m} |E\varepsilon_j| + \sum_{j=\log^2 m+1}^{m} |E\varepsilon_j|$$
$$\leq O(\log^2 m) + O(1) \sum_{j=\log^2 m+1}^{m} \frac{\log j}{j} = O(\log^2 m). \tag{6.18}$$

Part (h) is proved using (6.14)–(6.16):

$$|Z_k' E\varepsilon| \leq |\sum_{j=1}^{\log^2 m} (j/m)^k E\varepsilon_j| + |\sum_{j=\log^2 m+1}^{m} (j/m)^k E\varepsilon_j|$$
$$= O(1) \sum_{j=1}^{\log^2 m} (j/m)^k + O(\frac{1}{m^k}) \sum_{j=\log^2 m+1}^{m} j^{k-1} \log j$$
$$= O(1) \frac{\log^{2k+2} m}{m^k} + O(\log m) = O(\log m). \ \Box \tag{6.19}$$

**Proof of Lemma 2.** Part (a) is established as follows:

$$X^{*\prime} X^* = (X + 1_m 2\log(2\pi/n))'(X + 1_m 2\log(2\pi/n)) - (1_m'(X + 1_m 2\log(2\pi/n)))^2/m$$
$$= 4\sum_{j=1}^{m} \log^2 j - (-2m\log m + 2m + O(\log m))^2/m, \tag{6.20}$$

where the second equality uses Lemma 1(a). Estimating $\sum_{j=1}^{m} \log^2 j$ by the corresponding integral, as in (6.10), yields

$$\sum_{j=1}^{m} \log^2 j = m\log^2 m - 2m\log m + 2m + O(\log^2 m), \tag{6.21}$$

because $\int_a^b \log^2 x \, dx = [x\log^2 x - 2x\log x + 2x]_a^b$. Combining (6.20) and (6.21) gives the desired result.

Part (b) is established using Lemma 1(b) and (c):

$$Z_i^{*\prime} Z_k^* = Z_i' Z_k - \frac{1}{m} 1_m' Z_i 1_m' Z_k$$
$$= \frac{1}{i+k+1} m(1+o(1)) - \frac{1}{m}\frac{1}{i+1} m \frac{1}{k+1} m(1+o(1))$$
$$= (\frac{1}{i+k+1} - \frac{1}{(i+1)(k+1)}) m(1+o(1)). \tag{6.22}$$

Part (c) is established using Lemma 1(a), (b), and (d):

$$X^{*\prime} Z_k^* = X^{*\prime} Z_k = (X + 1_m 2\log(2\pi/n))' Z_k - (\frac{1}{m} 1_m' X + 2\log(2\pi/n)) 1_m' Z_k$$

$$
\begin{aligned}
&= -\frac{2}{k+1} m \log m + \frac{2}{(k+1)^2} m(1+o(1)) \\
&\quad - \left(-2\log m + 2 + O(\frac{\log m}{m})\right)\left(\frac{1}{k+1} m + O(1)\right) \\
&= -\frac{2k}{(k+1)^2} m(1+o(1)). \qquad (6.23)
\end{aligned}
$$

Part (d) is established as follows. By the definition of $R_j$ in (3.3) and Lemma 2(c), we have

$$
\begin{aligned}
X^{*\prime}R &= 1(s \geq 2+2r)\frac{b_{2+2r}}{(2+2r)!} X^{*\prime} Z^*_{2+2r} (2\pi m/n)^{2+2r} + \sum_{j=1}^{m} (X_j - \overline{X}) O(\lambda_j^q) \\
&= -1(s \geq 2+2r) \frac{2(2\pi)^{2+2r}(2+2r)b_{2+2r}}{(3+2r)!(3+2r)} \frac{m^{3+2r}}{n^{2+2r}}(1+o(1)) + \sum_{j=1}^{m}(X_j - \overline{X})O(\lambda_j^q).
\end{aligned} \qquad (6.24)
$$

Next, we have

$$
\begin{aligned}
&|\frac{1}{2}\sum_{j=1}^{m}(X_j - \overline{X})O(\lambda_j^q)| \\
&= |\sum_{j=1}^{m}(\log j - \log m + 1 + O(\frac{\log m}{m}))O(\lambda_j^q)| \\
&\leq |O(1)| \sum_{j=1}^{m} \lambda_j^q |\log j - \log m| + |\sum_{j=1}^{m}(1 + O(\frac{\log m}{m}))O(\lambda_j^q)| \\
&= O(n^{-q}) \sum_{j=1}^{m} j^q (\log m - \log j) + O(\frac{m^{q+1}}{n^q}) \\
&= O(n^{-q})(\frac{1}{(q+1)^2} m^{q+1} + O(m^q \log m) + o(m^{q+1})) + O(\frac{m^{q+1}}{n^q}) \\
&= O(\frac{m^{q+1}}{n^q}), \qquad (6.25)
\end{aligned}
$$

using Lemma 1(a), the triangle inequality, (6.14), the result that $\sum_{j=1}^{m} j^q = m^{q+1}/(q+1) + O(m^q)$, and the uniformity of the $O(\cdot)$ terms over $j = 1, ..., m$. Equations (6.24) and (6.25) combine to establish part (d).

Part (e) is established using Lemma 1(b), (e), and (f):

$$
\begin{aligned}
Z_k^{*\prime} R &= Z_k' R - \frac{1}{m} 1_m' Z_k 1_m' R \\
&= 1(s \geq 2+2r)(\frac{(2\pi)^{2+2r} b_{2+2r}}{(2+2r)!(3+2r+k)} \frac{m^{3+2r}}{n^{2+2r}}(1+o(1)) + O(\frac{m^{q+1}}{n^q}) \\
&\quad - \frac{1}{m}\left(\frac{1}{k+1} m(1+o(1))\right)
\end{aligned}
$$

$$\times\left(1(s\geq 2+2r)(\frac{(2\pi)^{2+2r}b_{2+2r}}{(3+2r)!}\frac{m^{3+2r}}{n^{2+2r}}(1+o(1))+O(\frac{m^{q+1}}{n^q})\right)$$
$$= 1(s\geq 2+2r)\frac{(2\pi)^{2+2r}b_{2+2r}}{(2+2r)!}(\frac{1}{3+2r+k}-\frac{1}{(k+1)(3+2r)})\frac{m^{3+2r}}{n^{2+2r}}(1+o(1))$$
$$+O(\frac{m^{q+1}}{n^q}). \tag{6.26}$$

When $s$ is an integer, the $O(\cdot)$ terms are $o(\cdot)$ in the proofs of parts (d) and (e) because the same holds in the definition of $R_j$ in (3.3).

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

Part (g) is established using Lemma 1(b), (g), and (h):

$$\begin{aligned}Z_k^{*\prime}E\varepsilon &= Z_k'E\varepsilon-\frac{1}{m}1_m'Z_k1_m'E\varepsilon\\ &= O(\log m)-\frac{1}{m}\left(\frac{1}{k+1}m+O(1)\right)O(\log^2 m))\\ &= O(\log^2 m).\end{aligned} \tag{6.27}$$

Parts (h) and (i) hold by Lemma 2(b) and (c) and the definitions of $\Gamma_r$ and $\mu_r$. Part (j) holds by Lemma 2(a), (h), and (i) using the definition of $c_r$.

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

$$M_{Z^*}X^*=X^*-Z^*(Z^{*'}Z^*)^{-1}Z^{*'}X^*=O(\log m)1_m+2Z^*\Gamma_r^{-1}\mu_r(1+o(1)). \tag{6.28}$$

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

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

$$\omega(\lambda)=\frac{1}{(2\pi n)^{1/2}}\sum_{t=1}^{n}Y_t\exp(i\lambda_j t)\text{ and }u(\lambda)=\omega(\lambda)/f^{1/2}(\lambda). \tag{6.29}$$

**Lemma 3** (a) *Suppose Assumption 2 holds. Then, uniformly over* $j,k=1,...,m$ *and all* $n\geq 1$,

(i) $Eu(\lambda_j)\overline{u(\lambda_j)}=1+O(\frac{\log j}{j})$,

(ii) $Eu(\lambda_j)u(\lambda_j)=O(\frac{\log j}{j})$,

(iii) $Eu(\lambda_j)\overline{u(\lambda_k)}=O(\frac{\log j}{k})$, *and*

(iv) $Eu(\lambda_j)u(\lambda_k)=O(\frac{\log j}{k})$.

(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.

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 $0 < \lambda_1 < \lambda_2 \leq \widetilde{\delta} = \min\{\delta_2, \delta_3\}$,

$$|f(\lambda_1) - f(\lambda_2)| \leq C|\lambda_1|^{-1-2d_f}|\lambda_1 - \lambda_2| \tag{6.30}$$

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

$$\begin{aligned} |f(\lambda_1) - f(\lambda_2)| &= |\lambda_1^{-2d_f} g(\lambda_1) - \lambda_2^{-2d_f} g(\lambda_2)| \\ &\leq \lambda_1^{-2d_f}|g(\lambda_1) - g(\lambda_2)| + |g(\lambda_2)| \cdot |\lambda_1^{-2d_f} - \lambda_2^{-2d_f}| \\ &\leq K_3\lambda_1^{-2d_f}|\lambda_1 - \lambda_2| + 2|d_f|C_5\lambda_1^{-1-2d_f}|\lambda_1 - \lambda_2| \end{aligned} \tag{6.31}$$

for all $0 < \lambda_1 < \lambda_2 \leq \widetilde{\delta}$, 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 $\lambda_2$, 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 $\varepsilon$ by $\widetilde{\delta}$.

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 integrals $\int_{-\lambda_j/2}^{\lambda_j/2}$ in the first and last lines on p. 1062 and the second lines on pp. 1064 and 1065 hold uniformly over $|d_f| < 1/2$. The difficulty is that $\sup_{|d_f|<1/2} \int_0^{\lambda_j/2} \lambda^{-2d_f} d\lambda = \sup_{|d_f|<1/2} \lambda_j^{1-2d_f}/(1 - 2d_f) = \infty$. It may be possible to circumvent this problem by using a different argument to determine the order of the $\int_{-\lambda_j/2}^{\lambda_j/2}$ integrals than that used in Robinson (1995b). This is probably what GRS had in mind.

We impose the condition $\int_{-\pi}^{\pi} f(\lambda)d\lambda \leq K_1$ 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 $f \in \mathcal{F}(s, a, \delta, K)$. $\square$

**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 $\varepsilon_j$ 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

$$\begin{aligned} m^{1/2}(\widehat{d}_r - d) &= m^{1/2}(X^{*\prime} M_{Z^*} X^*)^{-1} X^{*\prime} M_{Z^*}(R + \varepsilon) \\ &= O(m^{-1/2}) X^{*\prime} M_{Z^*} R + (1 + o(1)) \frac{c_r}{4m^{1/2}} \sum_{j=1}^{m} [X^{*\prime} M_{Z^*}]_j \varepsilon_j. \end{aligned} \tag{6.32}$$

By (6.4), when $s \geq 2 + 2r$,

$$O(m^{-1/2}) X^{*\prime} M_{Z^*} R = O(\frac{m^{2.5+2r}}{n^{2+2r}}) + O(\frac{m^{q+0.5}}{n^q}) = o(1), \tag{6.33}$$

using Assumption 3 to obtain the second equality. By (6.5), when $s < 2 + 2r$,

$$O(m^{-1/2}) X^{*\prime} M_{Z^*} R = O(\frac{m^{s+0.5}}{n^s}) = o(1), \tag{6.34}$$

using Assumption 3 to obtain the second equality.

Hence, it suffices to show that

$$m^{-1/2} \sum_{j=1}^{m} A_j \varepsilon_j \to_d N(0, \frac{4}{c_r} \frac{\pi^2}{6}). \tag{6.35}$$

We write

$$\begin{aligned} &m^{-1/2} \sum_{j=1}^{m} A_j \varepsilon_j = T_1 + T_2 + T_3, \text{ where } T_1 := m^{-1/2} \sum_{j=1}^{\log^8 m} A_j \varepsilon_j, \\ &T_2 := m^{-1/2} \sum_{j=1+\log^8 m}^{m^{0.5+\delta}} A_j \varepsilon_j, \text{ and } T_3 := m^{-1/2} \sum_{j=1+m^{0.5+\delta}}^{m} A_j \varepsilon_j \end{aligned} \tag{6.36}$$

for some $0 < \delta < 0.5$. The proofs in HDB that $T_1 = o_p(1)$ and $T_2 = o_p(1)$ also are valid in our case, because $\max_{1 \leq j \leq m} |A_j| = O(\log m)$ 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) $\alpha$ equals $\min\{s, 2\}$ in our notation, so that a large value of $s$ does not increase $\alpha$ above 2.

Instead of using the method of HDB, we show that $T_3 c_r^{1/2}/2 \to_d N(0, \pi^2/6)$ by altering the asymptotic normality result given in (5.14) of Robinson (1995b). Robinson's (5.14) states that the normalized sum of random variables

$m^{-1/2}\sum_{j=1+m^{0.5+\delta}}^{m} a_j v U_j$ is asymptotically normal with mean zero and variance $\Omega$, where $\{a_j : j = 1, ..., m\}$ are constants that satisfy the conditions in his equation (5.15), $v$ is a non-zero constant, and $\{U_j : j = 1, ..., m\}$ are the random variables defined by

$$U_j := \varepsilon_j + \log\{g(\lambda_j)/g(0)\}, \tag{6.37}$$

using our notation. (This is actually a special case of Robinson's result when his $J = 1$ and $l = m^{0.5+\delta}$.) We alter this result by replacing the $U_j$ by $\varepsilon_j$ and prove the altered result using an alteration of the proof given by Robinson. Specifically, the altered (5.14) result is that $m^{-1/2}\sum_{j=1+m^{0.5+\delta}}^{m} a_j v \varepsilon_j \to_d N(0, \Omega)$ as $m \to \infty$, where $\Omega$ is as in Robinson, provided our Assumptions 2 and 3 hold and the constants $\{a_j : j = 1, ..., m\}$ satisfy the conditions in Robinson's (5.15).

Robinson's proof of (5.14) 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 $\varepsilon_j$ 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 not used in his proof of (5.14) and, hence, is not needed for our altered (5.14) result.) This completes the proof of our altered (5.14) result.

Now, we apply the altered (5.14) result with $a_j := A_j c_r^{1/2}/2$ (and with $J = 1$, $l = m^{0.5+\delta}$, and $v = 2/c_r^{1/2}$ in Robinson's notation). The first condition of Robinson's (5.15) holds by Lemma 2(k). The second condition of (5.15) holds because

$$\sum_{j=1+m^{0.5+\delta}}^{m} A_j^2 = \sum_{j=1}^{m} A_j^2 - \sum_{j=1}^{m^{0.5+\delta}} A_j^2 = 4m/c_r + o(m), \tag{6.38}$$

using Lemma 2(j) and (k). The third condition of (5.15) holds because

$$\begin{aligned}\sum_{j=1+m^{0.5+\delta}}^{m} |A_j|^p &\leq 2^{p-1} \sum_{j=1+m^{0.5+\delta}}^{m} |X_j^*|^p + 2^{p-1} \sum_{j=1+m^{0.5+\delta}}^{m} |\widetilde{Z}_j^{*\prime}(Z^{*\prime}Z^*)^{-1}Z^{*\prime}X^*|^p \\ &= O(m)\end{aligned} \tag{6.39}$$

for all $p \geq 1$, where $\widetilde{Z}_j^*$ denotes the $j$-th row of $Z^*$ written as a column vector of dimension $r$. The equality in (6.39) 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, (5.15) of Robinson (1995b) holds. Using Lemma 7 of HDB, we find that the asymptotic covariance matrix $\Omega$ in (5.14) is $\pi^2/6$.

The above results combine to establish (6.35), which completes the proof. $\square$

**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

$$\int_{-\pi}^{\pi} (f_n^*(\lambda) - f_0(\lambda))^2 d\lambda \leq C_0 n^{-1} \text{ for some constant } C_0 < \infty, \tag{6.40}$$

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 $\widetilde{f}_n$ that satisfies the results of their Lemma 2 and is of the form

$$\widetilde{f}_n(\lambda) = \widetilde{c}_n |\lambda|^{-h_n} (1 + \widetilde{\Delta}_n(\lambda)), \tag{6.41}$$

where $\{\widetilde{c}_n : n \geq 1\}$ 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 $\lambda$ 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 $\mathcal{F}(s, a, \delta, K)$, 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 $\mathcal{F}(s, a, \delta, K)$. 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 $\mathcal{F}(s, a, \delta, K)$. 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 $\mathcal{F}(s, a, \delta, K)$, 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:

$$f_n(\lambda) = \begin{cases} c_n \lambda^{-h_n} & \text{for } \lambda \in (0, \delta_n] \\ 1 & \text{for } \lambda \in (\delta_n, \pi] \end{cases} = c_n \lambda^{-h_n} (1 + \Delta_n(\lambda)), \text{ where}$$

$$h_n = \kappa\delta_n^s \text{ for some } \kappa > 0,\ \delta_n = n^{-1/(2s+1)},\ c_n = 1 + \log\delta_n^{h_n}, \text{ and}$$

$$\Delta_n(\lambda) = \begin{cases} 0 & \text{for } \lambda \in (0, \delta_n] \\ c_n^{-1}\lambda^{h_n} - 1 & \text{for } \lambda \in (\delta_n, \pi] \end{cases}. \tag{6.42}$$

(Note that in the notation of GRS $s$ is $\beta$.)

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:

$$\begin{aligned} f_n^*(\lambda) &= \begin{cases} c_n\lambda^{-h_n} & \text{for } \lambda \in (0, \delta_n] \\ c_n\delta_n^{-h_n} & \text{for } \lambda \in (\delta_n, \pi] \end{cases} \\ &= c_n\lambda^{-h_n}(1 + \Delta_n^*(\lambda)) \\ &= \lambda^{-h_n}g_n^*(\lambda), \text{ where} \\ \Delta_n^*(\lambda) &= \begin{cases} 0 & \text{for } \lambda \in [0, \delta_n] \\ (\lambda/\delta_n)^{h_n} - 1 & \text{for } \lambda \in (\delta_n, \pi] \end{cases} \text{and} \\ g_n^*(\lambda) &= c_n(1 + \Delta_n^*(\lambda)). \end{aligned} \tag{6.43}$$

Now we show that $f_n^* \in \mathcal{F}(s, a, \delta, K)$ for $n$ large and $\kappa$ small (where $\kappa$ 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 $\kappa$ sufficiently small. The first inequality holds because $c_n\delta_n^{-h_n} \leq 1$, as shown below in (6.50). The second inequality holds by Lemma 1(i) of GRS. Thus, $f_n^*$ is of the form (6.41) and $g_n^*$ satisfies condition (ii) of $\mathcal{F}(s, a, \delta, K)$ with $g_k = 0$ for $k = 1, ..., [s/2]$ 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 $\mathcal{F}(s, a, \delta, K)$ 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 $\mathcal{F}(s, a, \delta, K)$ 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

$$\begin{aligned} |g_n^*(\lambda_1) - g_n^*(\lambda_2)| &= |\lambda_1^{h_n} f_n^*(\lambda_1) - \lambda_2^{h_n} f_n^*(\lambda_2)| \\ &\leq |\lambda_1^{h_n}| \cdot |f_n^*(\lambda_1) - f_n^*(\lambda_2)| + |f_n^*(\lambda_2)| \cdot |\lambda_1^{h_n} - \lambda_2^{h_n}| \\ &\leq \zeta h_n \lambda_*^{h_n - 1}|\lambda_1 - \lambda_2|, \end{aligned} \tag{6.44}$$

where $\lambda_*$ lies between $\lambda_1$ and $\lambda_2$, the first inequality holds by the triangle inequality, and the second inequality holds using $f_n^*(\lambda_1) = f_n^*(\lambda_2)$, $|f_n^*(\lambda_2)| \leq \zeta$ for all $n$ for some constant $\zeta \geq 1$, 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 $\lambda_2$. We have

$$h_n\lambda_*^{h_n-1} \leq h_n\delta_n^{h_n-1} = (h_n/\delta_n)(1+o(1)) = \kappa n^{-(s-1)/(2s+1)}(1+o(1)) = \kappa O(1) \leq K_3/\zeta \tag{6.45}$$

for $n$ large and $\kappa$ small, where the first equality holds because $\delta_n^{h_n} \to 1$ as $n \to \infty$, as shown below. Equations (6.44) and (6.45) combine to yield $|g_n^*(\lambda_1) - g_n^*(\lambda_2)| \leq$

$K_3|\lambda_1 - \lambda_2|$ for $n$ sufficiently large. If $\lambda_1 \in (0, \delta_n]$ and $\lambda_2 \in (\delta_n, \pi]$, then

$$\begin{aligned}|g_n^*(\lambda_1) - g_n^*(\lambda_2)| &\leq |g_n^*(\lambda_1) - g_n^*(\delta_n)| + |g_n^*(\delta_n) - g_n^*(\lambda_2)| \\ &\leq K_3|\lambda_1 - \lambda_2|, \end{aligned} \tag{6.46}$$

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 $\mathcal{F}(s, a, \delta, K)$ for $n$ large.

Fifth, we have

$$\int_{-\pi}^{\pi} f_n^*(\lambda)d\lambda/2 = \int_0^{\delta_n} c_n\lambda^{-h_n}d\lambda + \int_{\delta_n}^{\pi} c_n\delta_n^{-h_n}d\lambda = \frac{c_n\delta_n^{1-h_n}}{1-h_n} + c_n\delta_n^{-h_n}(\pi - \delta_n) = \pi + o(1), \tag{6.47}$$

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 (6.47) is $\pi(a_0 + a_{00})/2 + o(1)$, which is less than or equal to $K_1/2$ for $n$ large because $2\pi a_{00} \leq K_1$ by assumption.

Next, we show that (6.40) 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

$$\begin{aligned}\int_{-\pi}^{\pi}(f_n^*(\lambda) - f_0(\lambda))^2 d\lambda &= 2\int_0^{\delta_n}(f_n(\lambda) - f_0(\lambda))^2 d\lambda + 2\int_{\delta_n}^{\pi}\xi_n^2 d\lambda \\ &\leq C_0 n^{-1} + O(n^{-4s/(2s+1)}\log^4 n) \end{aligned} \tag{6.48}$$

for some constant $C_0 > 0$, where the inequality uses Lemma 1(ii) of GRS for the bound on the first term. Thus, (6.40) 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 $\xi_n$ 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 $\kappa_1$ and $\gamma$ as above, this gives

$$\delta_n^{h_n} = n^{\kappa_1 n^{-\gamma}} = 1 + (n^{x_{1n}}\log n)\kappa_1 n^{-\gamma} = 1 + O(n^{-s/(2s+1)}\log n), \tag{6.49}$$

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 (6.49) gives

$$0 < \xi_n = (\delta_n^{h_n} - 1 - \log\delta_n^{h_n})/\delta_n^{h_n} = O((\delta_n^{h_n} - 1)^2) = O(n^{-2s/(2s+1)}\log^2 n). \tag{6.50}$$

Next, we prove part (b) of the Theorem. The function $\log(1+x)$ satisfies $\log(1+x) = x + \gamma(x)$, where $|\gamma(x)| \leq C_1 x^2$ for all $|x| \leq x_0$ for some $x_0 > 0$ and $C_1 < \infty$. Let

$g$ be a function that satisfies conditions (i) and (ii) of $\mathcal{F}(s,a,\delta,K)$. Then,

$$\begin{aligned}
\log(g(\lambda)/g(0)) &= \log\left(1+\sum_{k=1}^{[s/2]}(g_k/g(0))\lambda^{2k}+\Delta(\lambda)/g(0)\right)\\
&= \sum_{k=1}^{[s/2]}(g_k/g(0))\lambda^{2k}+\Delta(\lambda)/g(0)+\gamma\left(\sum_{k=1}^{[s/2]}(g_k/g(0))\lambda^{2k}+\Delta(\lambda)/g(0)\right)\\
&= \sum_{k=1}^{[s/2]}(g_k/g(0))\lambda^{2k}+\zeta(\lambda), \qquad (6.51)
\end{aligned}$$

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

Now, the result of part (b) holds for the estimator $\widehat{d}_{r,m_n}$ for a single density $f\in\mathcal{F}(s,a,\delta,K)$ and a single sequence $\{m_n : m_n\in J_n(s,D_0),\ n\geq 1\}$ by Theorem 1, because (i) Assumption 1 holds, (ii) Assumption 2 holds except that $g$ is not necessarily smooth of order $s\geq 1$, (iii) the expansion (6.51) 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 (6.51), (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 $r\geq(s-2)/2$ implies that $E_f\widehat{d}_{r,m_n}-d_f = O(m^s/n^s)=O(n^{-s/(2s+1)})$ and $Var_f(\widehat{d}_{r,m_n})=O(m^{-1})=O(n^{-2s/(2s+1)})$. Hence, it suffices to show that the results of Theorem 1 hold uniformly over $f\in\mathcal{F}(s,a,\delta,K)$ 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 (6.51) over functions $g$ that satisfy conditions (i)–(iii) of $\mathcal{F}(s,a,\delta,K)$.

Note that we impose the condition $g(0)\geq a_0$ in $\mathcal{F}(s,a,\delta,K)$ so that $g_k/g(0)$ in (6.51) is uniformly bounded and $\zeta(\lambda)$ in (6.51) satisfies $|\zeta(\lambda)|\leq C_2\lambda^s$. $\square$

# Footnotes

[1]The first author gratefully acknowledges the research support of the National Science Foundation via grant number SBR-9730277. The authors thank Marc Henry and Peter Robinson for helpful comments.

[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.

# 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 (1999): "Local Polynomial Whittle Estimation of Long-range Dependence," manuscript in preparation, Cowles Foundation, Yale University.

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.

Deo, R. and C. M. Hurvich (1999): "On the Log Periodogram Regression Estimator of the Memory Parameter in Long Memory Stochastic Volatility Models," working paper, Stern School of Business, New York University.

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., 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,* forthcoming.

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. 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, 220–232. New York: Springer-Verlag.

Henry, M. (1999): "Robust Automatic Bandwidth for Long Memory," working paper, Department of Economics, Columbia University.

Hurvich, C. M. (2000): "Model Selection for Broadband Semiparametric Estimation of Long Memory in Time Series," Working Paper No. SOR-2000-4, Stern School of Business, New York University.

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 (1997): "Broadband Semiparametric Estimation of the Memory Parameter of a Long-memory Time Series Using Fractional Exponential Models," Working Paper No. SOR-97-2, Stern School of Business, New York University.

Hurvich, C. M. and R. Deo (1998): "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. 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.

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.

Lobato, I. and P. M. Robinson (1996): "Average Periodogram Estimation of Long Memory," *Journal of Econometrics,* 73, 303–324.

Moulines, E. and P. Soulier (1999): "Broad Band Log-periodogram Regression of Time Series with Long Range Dependence," *Annals of Statistics,* 27, 1415–1439.

———(2000): "Data-driven Order Selection for Projection Estimator of the Spectral Density of Time Series with Long Range Dependence," *Journal of Time Series Analysis,* 21, forthcoming.

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.

Shimotsu, K. and P. C. B. Phillips (1997): "Pooled Log Periodogram Regression," working paper, Cowles Foundation, Yale University.

Velasco, C. (1998a): "Non-Gaussian Log-periodogram Regression," working paper, Departmento de Estadística y Econometría, Universidad Carlos III de Madrid.

———(1998b): "Non-stationary Log-periodogram Regression," working paper, Departmento de Estadística y Econometría, Universidad Carlos III de Madrid.

**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.

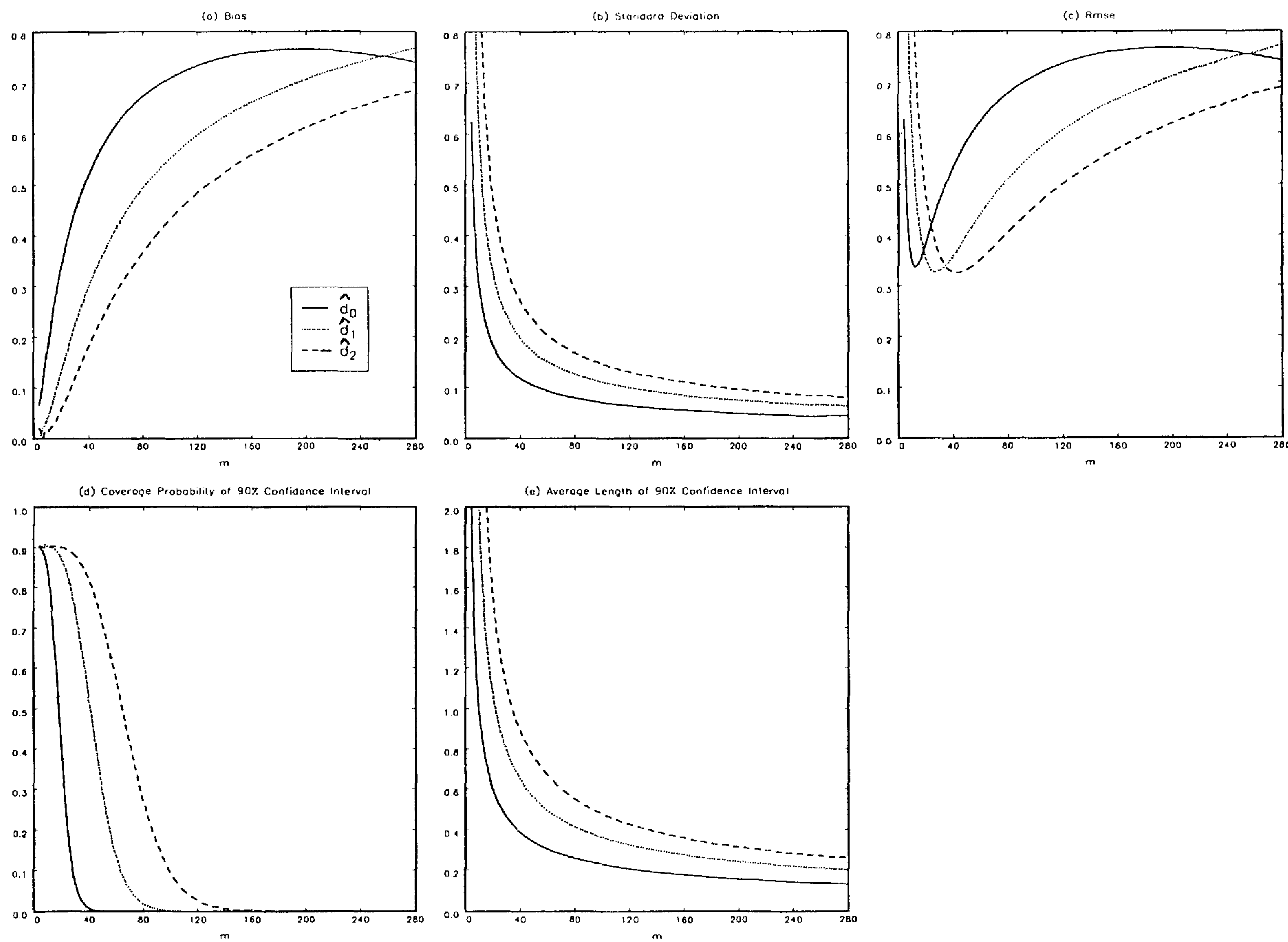

**Figure 2.** Performance of the log-periodogram regression estimators $\hat{d}_0$, $\hat{d}_1$, and $\hat{d}_2$ for an AR(1) process with AR parameter $\phi = .6$, for sample size $n = 512$, computed using 20,000 simulation repetitions.

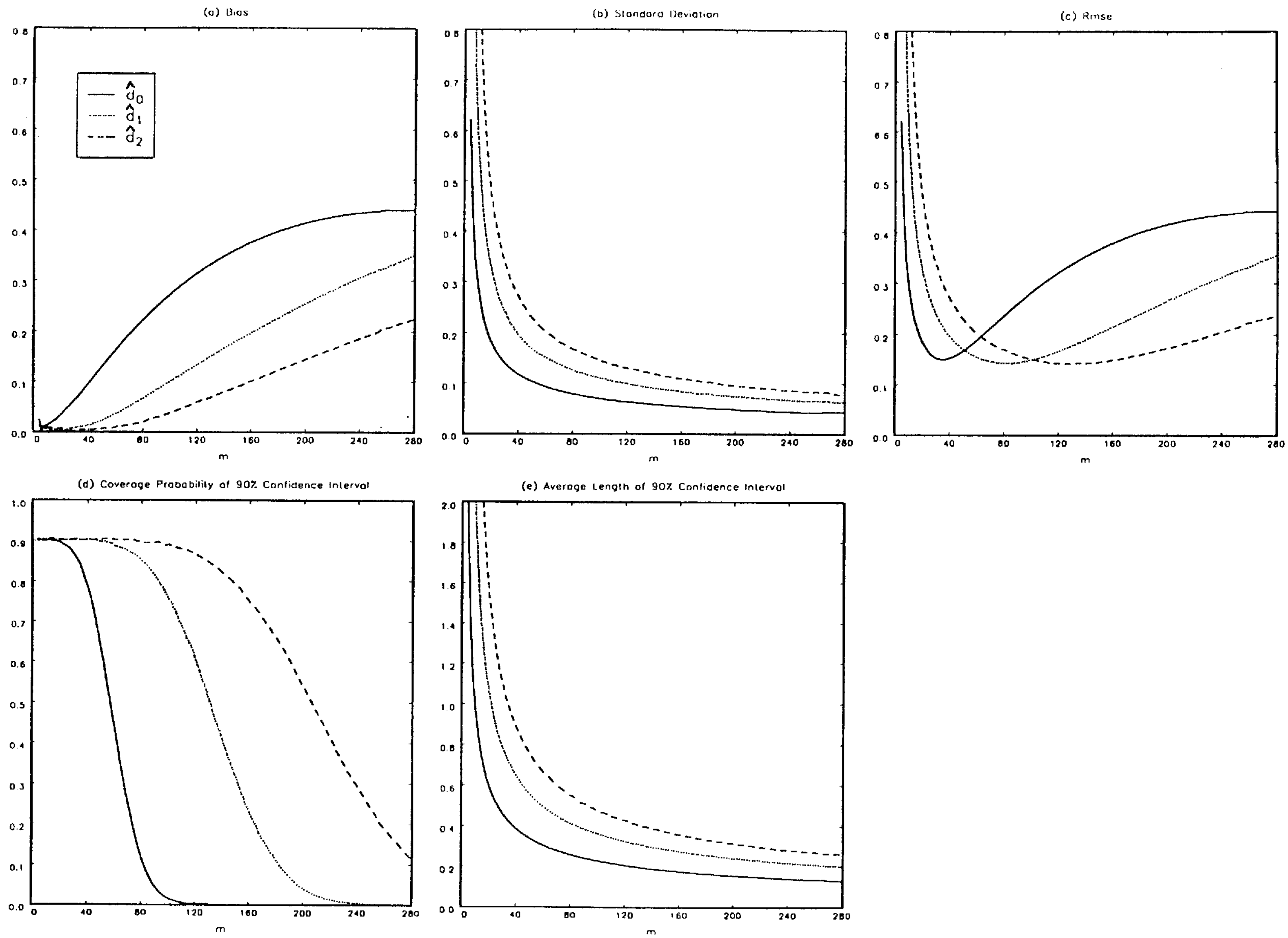

**Figure 3.** Performance of the log-periodogram regression estimators $\hat{d}_0$, $\hat{d}_1$, and $\hat{d}_2$ for an AR(1) process with AR parameter $\phi = .3$, for sample size $n = 512$, computed using 20,000 simulation repetitions.

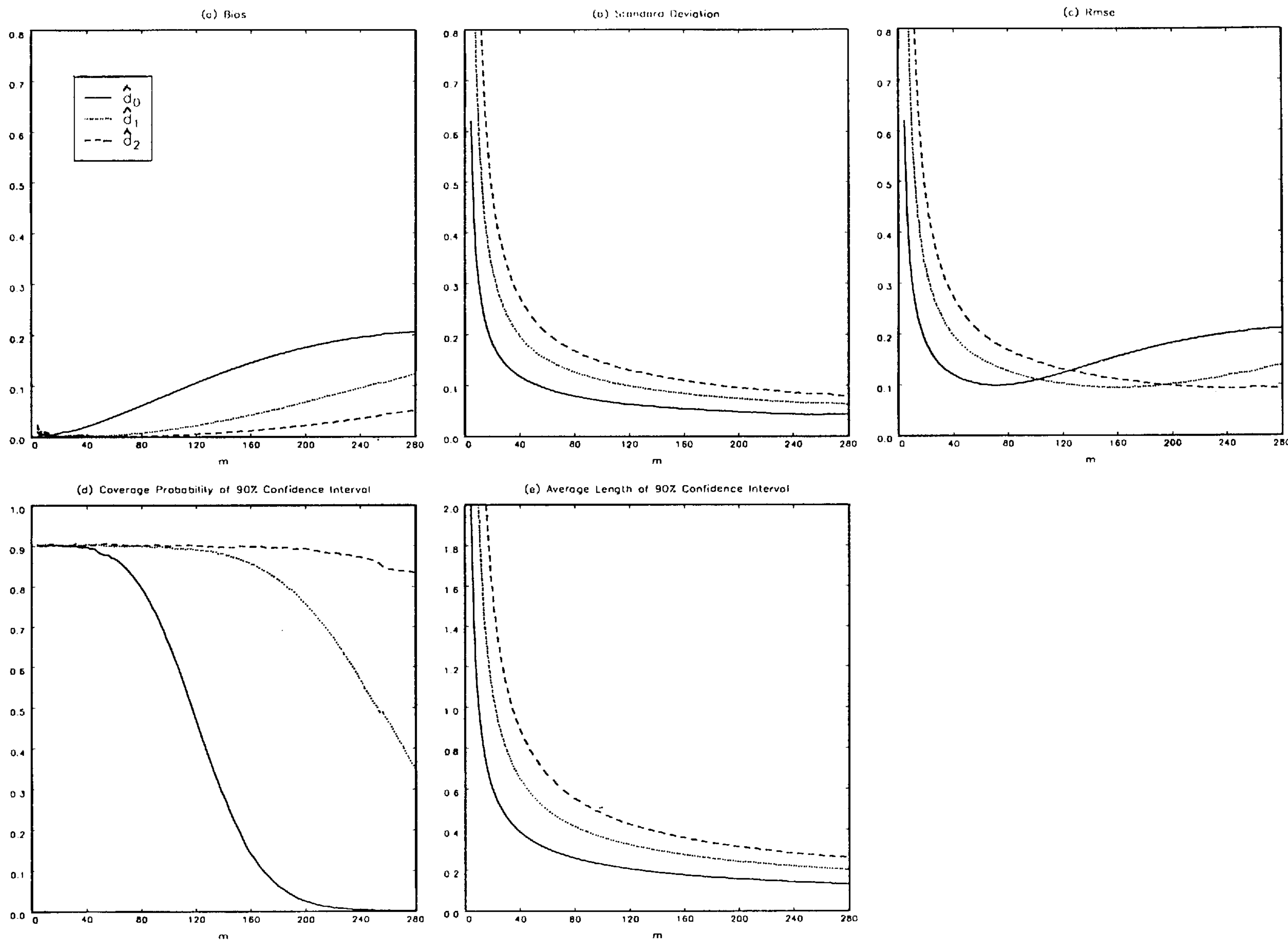

**Figure 4.** Performance of the log-periodogram regression estimators $\hat{d}_0$, $\hat{d}_1$, and $\hat{d}_2$ for an MA(1) process with MA parameter $\theta = -.9$, for sample size $n = 512$, computed using 20,000 simulation repetitions.

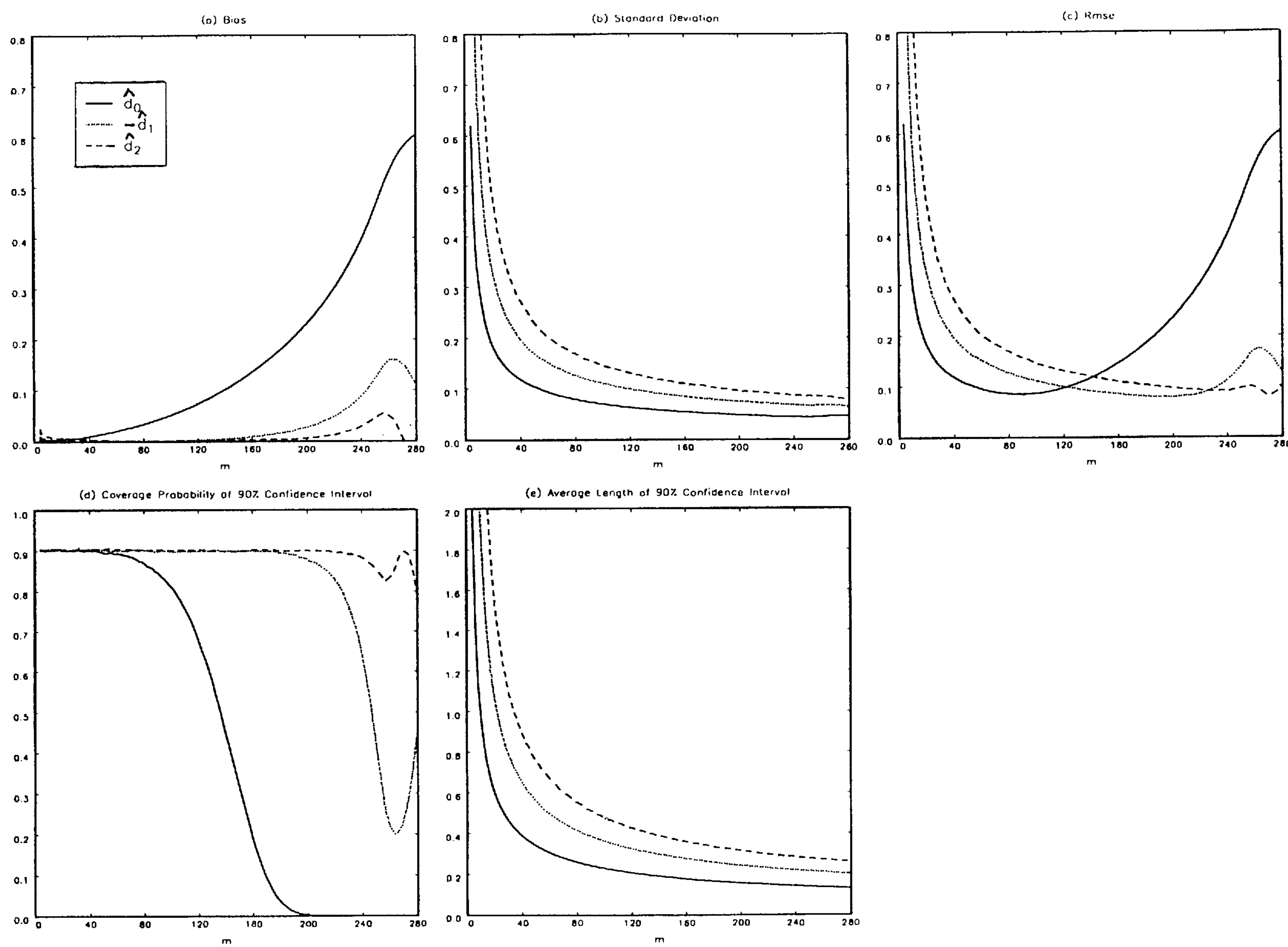

**Figure 5.** Performance of the log-periodogram regression estimators $\hat{d}_0$, $\hat{d}_1$, and $\hat{d}_2$ for an AR(1) process with AR parameter $\phi = .9$, for sample size $n = 2,048$, computed using 20,000 simulation repetitions.