THE MEAN SQUARED ERROR OF GEWEKE AND PORTER-HUDAK'S ESTIMATOR OF THE MEMORY PARAMETER OF A LONG-MEMORY TIME SERIES

By Clifford M. Hurvich, Rohit Deo and Julia Brodsky

Department of Statistics and Operations Research, New York University

First version received

Abstract. We establish some asymptotic properties of a log-periodogram regression estimator for the memory parameter of a long-memory time series. We consider the estimator originally proposed by Geweke and Porter-Hudak (The estimation and application of long memory time series models. *J. Time Ser. Anal.* 4 (1983), 221–37). In particular, we do not omit any of the low frequency periodogram ordinates from the regression. We derive expressions for the estimator's asymptotic bias, variance and mean squared error as functions of the number of periodogram ordinates, *m*, used in the regression. Consistency of the estimator is obtained as long as $m \to \infty$ and $n \to \infty$ with $(m \log m)/n \to 0$, where n is the sample size. Under these and the additional conditions assumed in this paper, the optimal *m*, minimizing the mean squared error, is of order $O(n^{4/5})$. We also establish the asymptotic normality of the estimator. In a simulation study, we assess the accuracy of our asymptotic theory on mean squared error for finite sample sizes. One finding is that the choice $m = n^{1/2}$, originally suggested by Geweke and Porter-Hudak (1983), can lead to performance which is markedly inferior to that of the optimal choice, even in reasonably small samples.

Keywords. Periodogram; semiparametric estimation.

1. INTRODUCTION

We consider a semiparametric model for a stationary Gaussian long-memory time series. According to this model, the spectral density of the time series $\{Y_t\}$ is given by

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

where $d \in (-0.5, 0.5)$ is the memory parameter, and $f^*(\cdot)$ is an even, positive, continuous function on $[-\pi, \pi]$ bounded above and bounded away from zero with first derivative $f^{*\prime}(0) = 0$ and second and third derivatives bounded in a neighborhood of zero. The function f^* endows the model (1) with a short-term correlation structure which is free of any parametrically imposed constraints. For this reason the semiparametric model (1) may be preferable to the assumption that $\{Y_t\}$ obeys a fractional ARIMA(p, d, q) model with p and q finite, either known or unknown. Although the fractional ARIMA model is in fact a special case of (1) obtained by assuming f^* to be the spectral density of a stationary invertible ARMA(p, q) process, we do not impose this or any other parametric restriction on f^* here.

The parameter d controls the long-memory aspects of the series, and will typically need to be estimated from a data set $y_0, \ldots, y_{n-1}$. Geweke and Porter-Hudak (1983) proposed a semiparametric estimator of d based on the first m periodogram ordinates

$$I_j = \frac{1}{2\pi n}\left|\sum_{t=0}^{n-1} y_t \exp(i\omega_j t)\right|^2 \qquad j = 1, \ldots, m$$

where $\omega_j = 2\pi j/n$ and m is a positive integer. Their estimator, hereafter called the GPH estimator, is given by $-1/2$ times the least squares estimate of the slope parameter in an ordinary linear regression of $\{\log I_j\}_{j=1}^m$ on the explanatory variable $X_j = \log|1 - \exp(-i\omega_j)| = \log|2\sin(\omega_j/2)|$, together with a constant term. Therefore, the GPH estimator can be written as

$$\hat{d} = \frac{-0.5\sum_{j=1}^m (X_j - \overline{X})\log I_j}{\sum_{k=1}^m (X_k - \overline{X})^2} \tag{2}$$

where

$$\overline{X} = \frac{1}{m}\sum_{k=1}^m X_k.$$

The GPH estimator can be motivated heuristically by noting that

$$\log I_j = (\log f_0^* - C) - 2dX_j + \log(f_j^*/f_0^*) + \varepsilon_j \tag{3}$$

where $\varepsilon_j = \log(I_j/f_j) + C$, $f_j = f(\omega_j)$, $f_j^* = f^*(\omega_j)$, and $C = 0.577\,216\ldots$ is Euler's constant. It seems reasonable in discussions of the asymptotic properties of the GPH estimator to assume that $m \to \infty$ so that the variance of $\hat{d}$ will decrease to zero as $n \to \infty$, and also to assume that $m/n \to 0$, so that bias due to the non-constancy of $\log(f_j^*/f_0^*)$ will tend to zero. Unfortunately, formula (3) does not lead to any easy proof of the properties of the GPH estimator, since the 'errors' $\{\varepsilon_j\}$ are not independent and identically distributed (see Kunsch, 1986; Hurvich and Beltrao, 1993; Robinson, 1995). Among other difficulties, $\lim_{n\to\infty} E(\varepsilon_j)$ depends on j, most strongly when j is close to zero, so the $\{\varepsilon_j\}$ may contribute to the bias of $\hat{d}$.

Although the GPH estimator is widely used in practice, its consistency for all $d \in (-0.5, 0.5)$ has not heretofore been established, so far as we are aware. Robinson (1995) did prove consistency and asymptotic normality for a modified regression estimator which regresses $\{\log I_j\}_{j=l+1}^m$ on $\{X_j\}_{j=l+1}^m$ where l is a lower trunction point which tends to infinity more slowly than m. However, simulations (e.g. Hurvich and Beltrao, 1994) indicate that the modified estimator is typically outperformed in finite samples by the GPH estimator itself. The reason is that any bias reduction resulting from the omission of the first l periodogram ordinates from the regression is more than offset by inflation of the variance.

An important practical problem in the implementation of the GPH estimator is the choice of m. Clearly this choice entails a bias–variance tradeoff. Hurvich and Beltrao (1994) discussed a data-driven method of choosing m by minimizing the estimated mean squared error of $\hat{d}$. However, the even more basic theoretical problem of establishing the asymptotic mean squared error of $\hat{d}$ as a function of m, including both bias and variance terms, has remained unsolved. The primary aim of this paper is to obtain this mean squared error, assuming (as in Robinson, 1995) that the time series is Gaussian. This leads to an expression for the asymptotically optimal (though not data-driven) choice of m in terms of n and f^*. The optimal m is of order $O(n^{4/5})$. Also, our formula for the mean squared error implies that the GPH estimator is consistent under the additional condition that $(m \log m)/n \to 0$. Finally, we establish the asymptotic normality of the GPH estimator, employing the same lemmas used to derive the mean squared error.

We present simulation results to assess the accuracy of our asymptotic theory on mean squared error for finite sample sizes. One finding is that the choice $m = n^{1/2}$, originally suggested by Geweke and Porter-Hudak (1983), can lead to performance which is markedly inferior to that of the asymptotically optimal choice in reasonably small samples.

2. THEORETICAL RESULTS

Combining Equations (2) and (3) and defining $a_j = X_j - \overline{X}$ yields

$$\hat{d} - d = -\frac{1}{2S_{xx}} \sum_{j=1}^{m} a_j \log f_j^* - \frac{1}{2S_{xx}} \sum_{j=1}^{m} a_j \varepsilon_j$$

where

$$S_{xx} = \sum_{k=1}^{m} (X_k - \overline{X})^2 = \sum_{k=1}^{m} a_k^2.$$

Consequently, the bias of $\hat{d}$ is given by

$$E(\hat{d} - d) = -\frac{1}{2S_{xx}} \sum_{j=1}^{m} a_j \log f_j^* - \frac{1}{2S_{xx}} \sum_{j=1}^{m} a_j E(\varepsilon_j) \tag{4}$$

while the variance of $\hat{d}$ is given by

$$\begin{aligned}\operatorname{var}(\hat{d}) &= \frac{1}{4S_{xx}^2}\operatorname{var}\left(\sum_{j=1}^{m} a_j\varepsilon_j\right)\\ &= \frac{1}{4S_{xx}^2}\sum_{j=1}^{m} a_j^2 \operatorname{var}(\varepsilon_j) + \frac{1}{2S_{xx}^2}\sum_{k=1}^{m}\sum_{j=k+1}^{m} a_j a_k \operatorname{cov}(\varepsilon_j,\, \varepsilon_k).\end{aligned} \tag{5}$$

We assume the following conditions.

CONDITION 1. $m \to \infty$, $n \to \infty$, with $m/n \to 0$ and $(m \log m)/n \to 0$.

CONDITION 2. $f^{*\prime}(0) = 0$, $|f^{*\prime\prime}(\omega)| < \tilde{B}_2 < \infty$, and $|f^{*\prime\prime\prime}(\omega)| < \tilde{B}_3 < \infty$, for all ω in a neighborhood of zero.

Under these conditions, Hurvich and Beltrao (1994) showed that $S_{xx} = m + \mathrm{o}(m)$ and that $a_j = \mathrm{O}(\log m)$. The first term on the right-hand side of Equation (4) is the contribution to the bias of $\hat{d}$ due to the non-constancy of f^*, and is expressed asymptotically by the following lemma. (All proofs of lemmas are given in the Appendix.)

LEMMA 1

$$-\frac{1}{2S_{xx}}\sum_{j=1}^{m} a_j \log f_j^* = \frac{-2\pi^2}{9}\frac{f^{*\prime\prime}(0)}{f^*(0)}\frac{m^2}{n^2} + \mathrm{o}\left(\frac{m^2}{n^2}\right).$$

To obtain asymptotic expressions for the other terms in Equations (4) and (5), we will need approximations to the means, variances and covariances of the $\{\varepsilon_j\}$. We start by considering $\operatorname{cov}(\varepsilon_j,\, \varepsilon_k)$ with $j > k$.

Note that the normalized periodogram can be expressed as

$$\frac{I_j}{f_j} = \left(\frac{A_j}{f_j^{1/2}}\right)^2 + \left(\frac{B_j}{f_j^{1/2}}\right)^2$$

where A_j and B_j are the cosine and sine coefficients:

$$A_j = \frac{1}{(2\pi n)^{1/2}}\sum_{t=0}^{n-1} y_t \cos(\omega_j t) \qquad B_j = \frac{1}{(2\pi n)^{1/2}}\sum_{t=0}^{n-1} y_t \sin(\omega_j t).$$

Since $0 < j < n/2$, it follows that $E(A_j) = E(B_j) = 0$, even if $E(y_t) \neq 0$. Since the series is Gaussian and since $\varepsilon_j = \log(I_j/f_j) + C$, the joint distribution of $(\varepsilon_j,\, \varepsilon_k)$ is determined by the covariance matrix Σ of

$$v = \left(\frac{A_j}{f_j^{1/2}}, \frac{B_j}{f_j^{1/2}}, \frac{A_k}{f_k^{1/2}}, \frac{B_k}{f_k^{1/2}}\right)' = (v_1,\, v_2,\, v_3,\, v_4)'.$$

The vector v has a four-dimensional multivariate Gaussian distribution with mean zero. Let $\alpha_{jk} = \max\{|\mathrm{cov}(A_j/f_j^{1/2}, A_k/f_k^{1/2})|, |\mathrm{cov}(A_j/f_j^{1/2}, B_k/f_k^{1/2})|, |\mathrm{cov}(B_j/f_j^{1/2}, A_k/f_k^{1/2})|, |\mathrm{cov}(B_j/f_j^{1/2}, B_k/f_k^{1/2})|\}$.

LEMMA 2. $\mathrm{cov}(\varepsilon_j, \varepsilon_k) = \mathrm{O}(\alpha_{jk}^2)$, *uniformly for* $\log^2 m \leqslant k < j \leqslant m$.

The following lemma is an easy consequence of Robinson (1995, Theorem 2, parts (c) and (d)).

LEMMA 3. $\alpha_{jk} = \mathrm{O}(\log j/k)$, *uniformly for* $1 \leqslant k < j \leqslant m$.

To approximate the quantities $E(\varepsilon_j)$ and $\mathrm{var}(\varepsilon_j)$ which appear in Equations (4) and (5), we adapt a result from Robinson (1995) to conclude that

$$E\left(\frac{A_j^2}{f_j}\right) = \tfrac{1}{2} + \mathrm{O}\left(\frac{\log j}{j}\right) \qquad E\left(\frac{B_j^2}{f_j}\right) = \tfrac{1}{2} + \mathrm{O}\left(\frac{\log j}{j}\right)$$

and

$$E\left(\frac{A_j B_j}{f_j}\right) = \mathrm{O}\left(\frac{\log j}{j}\right)$$

uniformly for $1 \leqslant j \leqslant m$. It should be noted that the corresponding results in Robinson (1995) have an additional $\mathrm{O}(j^2/n^2)$ term arising from his use of a different normalization, e.g.

$$E\left(\frac{A_j^2}{|\omega_j|^{-2d}}\right) = \tfrac{1}{2} + \mathrm{O}\left(\frac{\log j}{j}\right) + \mathrm{O}\left(\frac{j^2}{n^2}\right).$$

This additional term does not arise in our case, since our normalized sine and cosine coefficients are $A_j/f_j^{1/2}$ and $B_j/f_j^{1/2}$ rather than the $A_j/|\omega_j|^{-d}$ and $B_j/|\omega_j|^{-d}$ used by Robinson (1995).

LEMMA 4. $\lim_n \inf_{1 \leqslant j \leqslant m} E(I_j/f_j) > 0$.

LEMMA 5. $\lim_n \sup_{1 \leqslant j \leqslant m} E[\{\log(I_j/f_j)\}^2] < \infty$.

By an argument similar to the proof of Lemma 2, we obtain the following.

LEMMA 6. $E(\varepsilon_j) = \mathrm{O}(\log j/j)$, *uniformly for* $\log^2 m \leqslant j \leqslant m$.

LEMMA 7. $\mathrm{var}(\varepsilon_j) = \pi^2/6 + \mathrm{O}(\log j/j)$, *uniformly for* $\log^2 m \leqslant j \leqslant m$.

The second term of the right-hand side of Equation (4) is the contribution to the bias of $\hat{d}$ due to the long-memory nature of the process. It is given asymptotically by Lemma 8.

LEMMA 8

$$-\frac{1}{2S_{xx}}\sum_{j=1}^{m} a_j E(\varepsilon_j) = \mathrm{O}\left(\frac{\log^3 m}{m}\right).$$

This term will be asymptotically negligible compared with the bias due to f^* provided that $m = Kn^{\beta}$ where $K \neq 0$ and $\beta > 2/3$. It will not be negligible, however, if m grows sufficiently slowly, for example if $\beta \leqslant 2/3$.

We now state our main theorem, which is proved below.

THEOREM 1. *Under Conditions 1 and 2,*

$$E(\hat{d} - d) = \frac{-2\pi^2}{9}\frac{f^{*\prime\prime}(0)}{f^*(0)}\frac{m^2}{n^2} + \mathrm{o}\left(\frac{m^2}{n^2}\right) + \mathrm{O}\left(\frac{\log^3 m}{m}\right) \tag{6}$$

$$\mathrm{var}\,(\hat{d}) = \frac{\pi^2}{24m} + \mathrm{o}\left(\frac{1}{m}\right) \tag{7}$$

$$\begin{aligned}\mathrm{MSE}(\hat{d}) = E(\hat{d} - d)^2 = \frac{4\pi^4}{81}\left\{\frac{f^{*\prime\prime}(0)}{f^*(0)}\right\}^2\frac{m^4}{n^4} + \frac{\pi^2}{24m} + \mathrm{O}\left\{\frac{m(\log^3 m)}{n^2}\right\} \\ + \mathrm{o}\left(\frac{m^4}{n^4}\right) + \mathrm{o}\left(\frac{1}{m}\right).\end{aligned} \tag{8}$$

Since the mean squared error tends to zero, we obtain the corollary.

COROLLARY. Under Conditions 1 and 2, $\hat{d}$ is a consistent estimator of d.

It is noteworthy that consistency of $\hat{d}$ holds no matter how slowly m tends to infinity. For example, the choice $m = \log(\log n)$ would yield a consistent estimator.

Neglecting the remainder terms in the mean squared error (Equation (8)), assuming that $f^{*\prime\prime}(0) \neq 0$ and minimizing with respect to m yields the asymptotically optimal choice for m:

$$m^{\mathrm{THR}} = \left(\frac{27}{128\pi^2}\right)^{1/5}\left\{\frac{f^*(0)}{f^{*\prime\prime}(0)}\right\}^{2/5} n^{4/5}. \tag{9}$$

Note that the remainder term $\mathrm{O}\{m(\log^3 m)/n^2\}$ in Equation (8) is negligible compared with both the first and the second terms on the right-hand side of Equation (8) as long as $m = Kn^{\beta}$ for $2/3 < \beta < 1$. Consequently this remainder term does not affect the mean squared error for β in a neighborhood of the minimizing value, $\beta = 4/5$.

PROOF OF THEOREM 1. Equation (6) follows directly from Lemmas 1 and 8. To prove Equation (7), we note from the first part of Equation (5) that

$$\operatorname{var}(\hat{d}) = \frac{1}{4S_{xx}^2} \operatorname{var}(T_1 + T_2)$$

where

$$T_1 = \sum_{j=1}^{\log^6 m} a_j \varepsilon_j \qquad T_2 = \sum_{j=1+\log^6 m}^{m} a_j \varepsilon_j.$$

Now,

$$\begin{aligned}\operatorname{var}(T_1) &= \sum_{j=1}^{\log^6 m} a_j^2 \operatorname{var}(\varepsilon_j) + 2 \sum_{k=1}^{\log^6 m} \sum_{j=k+1}^{\log^6 m} a_j a_k \operatorname{cov}(\varepsilon_j, \varepsilon_k) \\ &= \mathrm{O}[\log^8 m + \log^{14} m \sup_j \{\operatorname{var}(\varepsilon_j)\}^{1/2} \sup_k \{\operatorname{var}(\varepsilon_k)\}^{1/2}] = \mathrm{O}(\log^{14} m)\end{aligned}$$

since $\operatorname{var}(\varepsilon_j)$ is uniformly bounded in j by Lemma 5. Now using Lemmas 7, 2 and 3 we obtain

$$\begin{aligned}\operatorname{var}(T_2) &= \sum_{j=1+\log^6 m}^{m} a_j^2 \operatorname{var}(\varepsilon_j) + 2 \sum_{k=1+\log^6 m}^{m} \sum_{j=k+1}^{m} a_j a_k \operatorname{cov}(\varepsilon_j, \varepsilon_k) \\ &= \sum_{j=1+\log^6 m}^{m} a_j^2 \left\{ \frac{\pi^2}{6} + \mathrm{O}\left(\frac{\log j}{j} \right) \right\} + \mathrm{O}\left(\log^2 m \sum_{k=1+\log^6 m}^{m} \sum_{j=k+1}^{m} \alpha_{jk}^2 \right) \\ &= \frac{\pi^2 m}{6} + \mathrm{o}(m) + \mathrm{O}\left(\log^2 m \sum_{j=1+\log^6 m}^{m} \frac{\log j}{j} \right) \\ &\quad + \mathrm{O}\left(\log^2 m \sum_{k=1+\log^6 m}^{m} \sum_{j=k+1}^{m} \frac{\log^2 j}{k^2} \right) \\ &= \frac{\pi^2 m}{6} + \mathrm{o}(m) + \mathrm{O}(\log^4 m) + \mathrm{O}\left(\log^4 m \sum_{k=1+\log^6 m}^{m} \frac{m}{k^2} \right) \\ &= \frac{\pi^2 m}{6} + \mathrm{o}(m) + \mathrm{O}\left(\frac{m}{\log^2 m} \right) \\ &= \frac{\pi^2 m}{6} + \mathrm{o}(m).\end{aligned}$$

Thus

$$\begin{aligned}\operatorname{var}(\hat{d}) &= \frac{1}{4S_{xx}^2}\{\operatorname{var}(T_1) + \operatorname{var}(T_2) + 2\operatorname{cov}(T_1, T_2)\} \\ &= \mathrm{O}\left(\frac{\log^{14} m}{m^2}\right) + \frac{\pi^2}{24m} + \mathrm{o}\left(\frac{1}{m}\right) + \mathrm{O}\left(\log^7 m \frac{m^{1/2}}{m^2}\right) \\ &= \frac{\pi^2}{24m} + \mathrm{o}\left(\frac{1}{m}\right)\end{aligned}$$ ■

THEOREM 2. *If* $m = \mathrm{o}(n^{4/5})$ *and* $\log^2 n = \mathrm{o}(m)$, *then*

$$m^{1/2}(\hat{d} - d) \xrightarrow{\mathrm{D}} \mathrm{N}(0, \pi^2/24).$$

The proof of Theorem 2 is given in the Appendix.

3. MONTE CARLO RESULTS

We performed a simulation study to assess the adequacy of our symptotic expression for MSE, and to compare the GPH estimator ($m = n^{1/2}$) with estimators based on an optimal choice of m. The specific choice $m = n^{1/2}$ was originally suggested by Geweke and Porter-Hudak (1983) because it gave good results in their simulations. We considered five processes, all ARIMA(1, d, 0) of form $(1 - B)^d(1 - \phi B)x_t = \varepsilon_t$, where the $\{\varepsilon_t\}$ are independent identically distributed standard normal and B is the backshift operator. The values for (d, ϕ) were (0.1, 0.3), (0.3, 0.1), (0.3, 0.3), (0.3, 0.9) and (0.49, 0.3). For each of these five processes, and for each of the four sample sizes $n = 200$, 700, 1500 and 2000, we generated 1000 realizations using the algorithm of Haslett and Raftery (1989), as implemented in the command arima.fracdiff.sim of S-PLUS. For each realization, we evaluated $\hat{d}_m$ for $m = 2, \ldots, n/2 - 1$, where $\hat{d}_m$ is the regression estimator of Equation (2) which uses the first m ordinates of the log periodogram. Define the Monte Carlo MSE as the average over the 1000 realizations of $(\hat{d}_m - d)^2$ for $m = 2, \ldots, n/2 - 1$. Define the theoretical MSE to be the expression given in Equation (8), omitting the remainder terms. Define m^{THR} and m^{MC} to be the minimizers of the theoretical and Monte Carlo MSE, respectively. Note that m^{THR} is given by Equation (9). Here, we present some selected results which are representative of the entire study.

Figures 1–5 present plots of the theoretical and Monte Carlo MSEs for $(d, \phi) = (0.3, 0.3)$ and $(d, \phi) = (0.3, 0.9)$, at various sample sizes. Overall, Figures 1–3 (for $\phi = 0.3$) indicate that the Monte Carlo MSE is higher than the theoretical MSE when m is small, while the situation is reversed when m is large. The discrepancy for small m is presumably caused by the omission of the $\mathrm{O}\{(\log^3 m)/m\}$ term for the bias due to long memory (Lemma 8) and the $\mathrm{o}(1/m)$ term in the expression for $\operatorname{var}(\hat{d})$ (Equation (7)). The discrepancy for

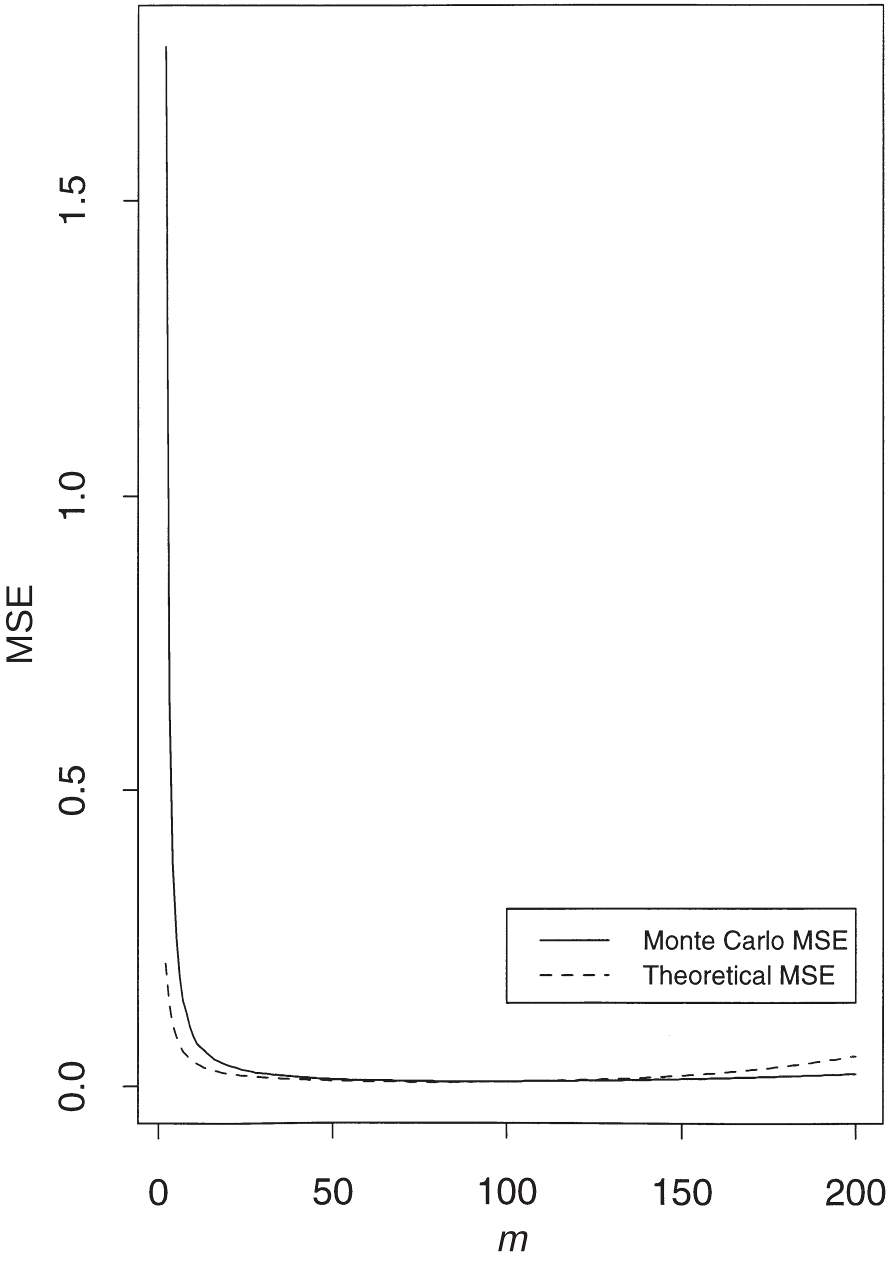

FIGURE 1. MSEs for the regression estimator, $d = 0.3$, $\phi = 0.3$, $n = 700$.

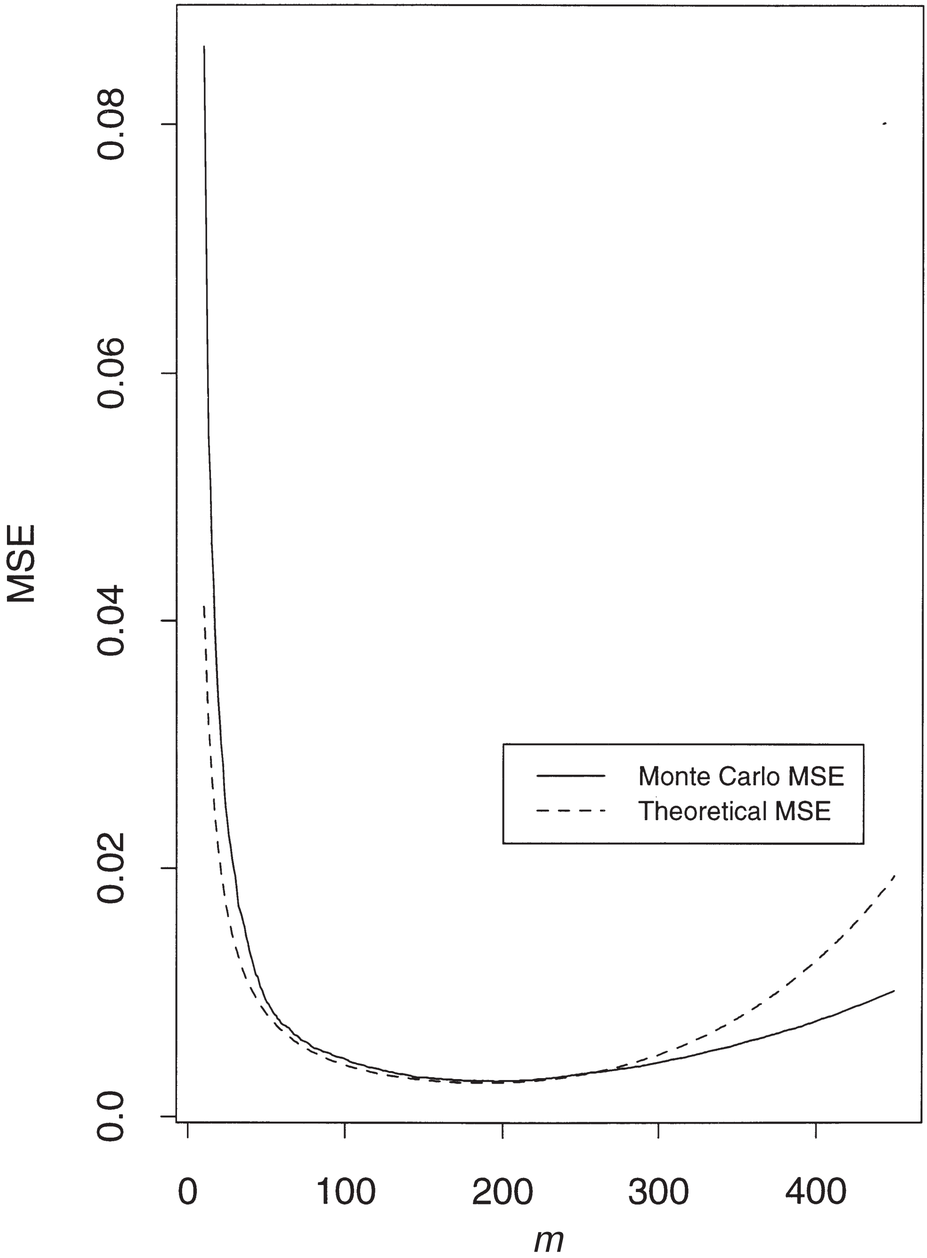

FIGURE 2. MSEs for the regression estimator, $d = 0.3$, $\phi = 0.3$, $n = 2000$.

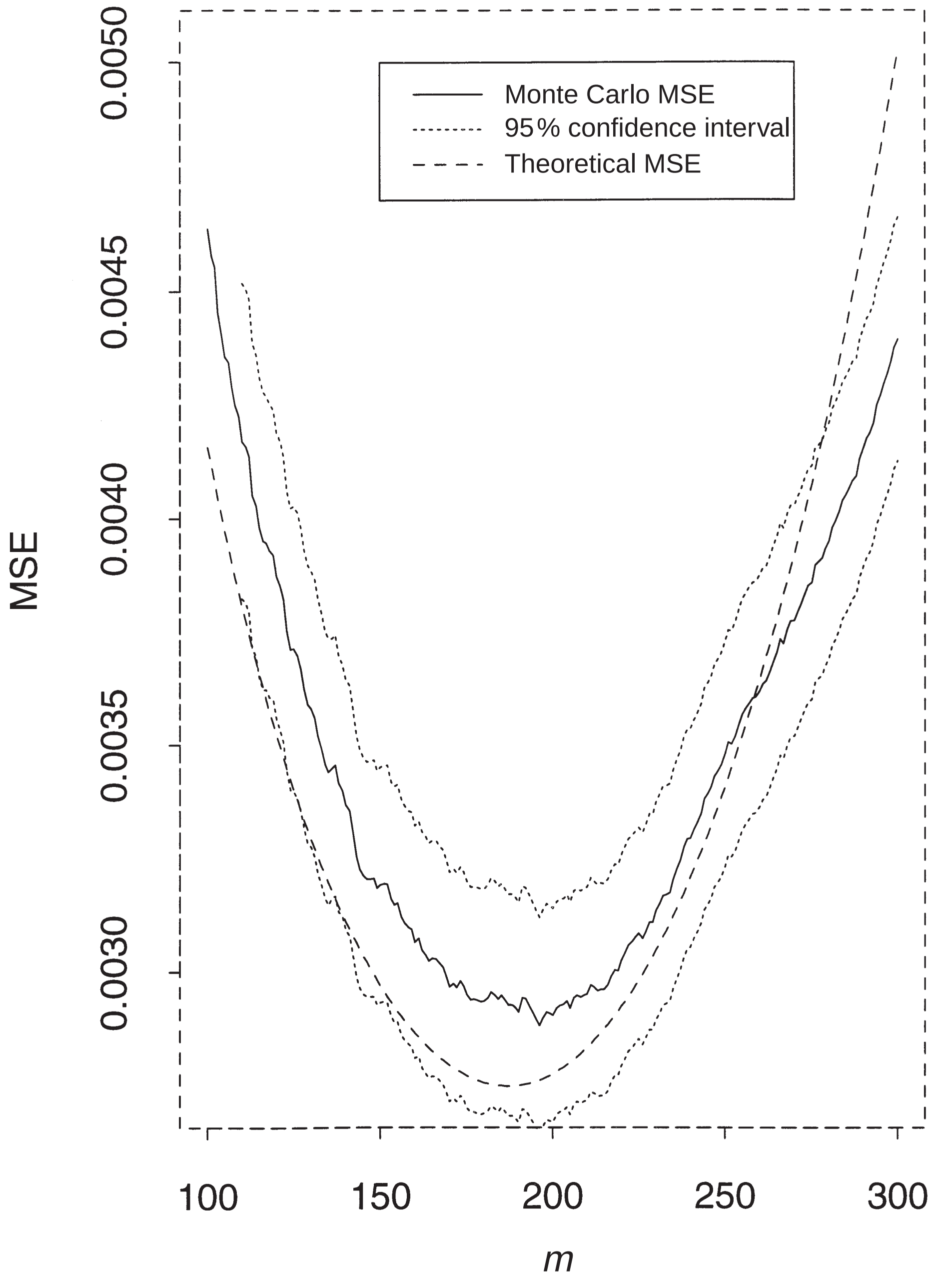

FIGURE 3. MSEs for the regression estimator, $d = 0.3$, $\phi = 0.3$, $n = 2000$.

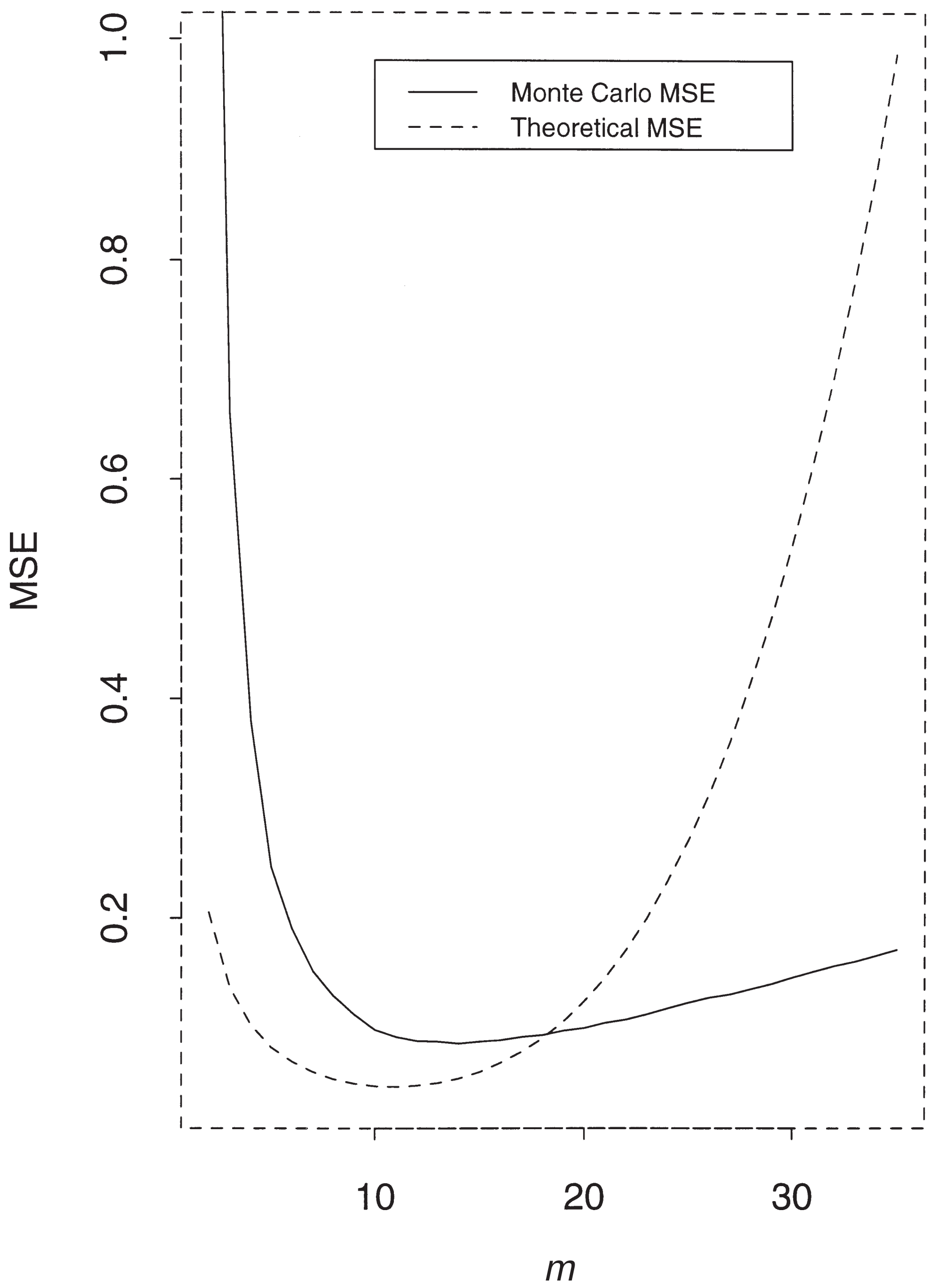

FIGURE 4. MSEs for the regression estimator, $d = 0.3$, $\phi = 0.9$, $n = 700$.

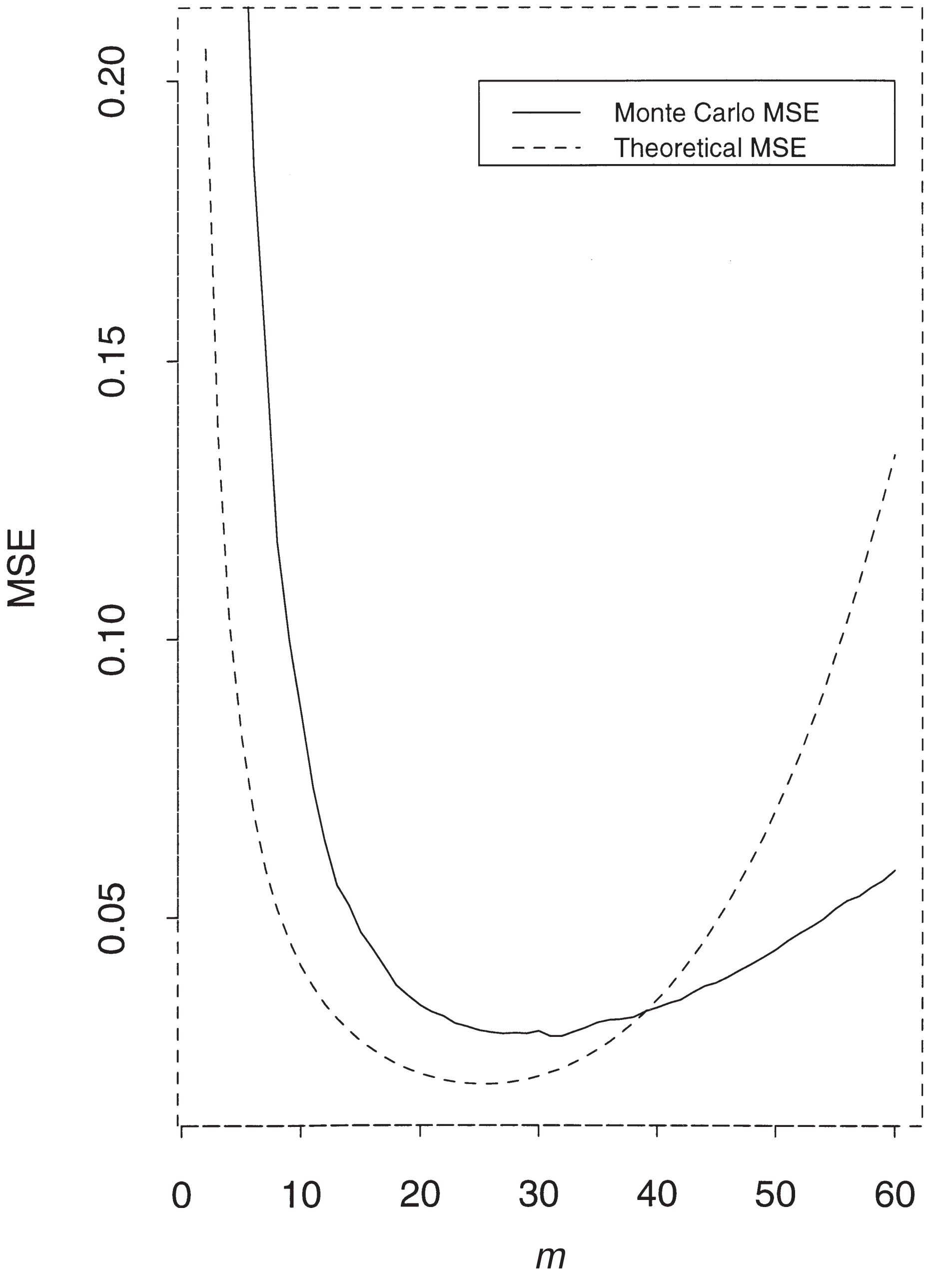

FIGURE 5. MSEs for the regression estimator, $d = 0.3$, $\phi = 0.9$, $n = 2000$.

large m presumably results from the eventual breakdown of the Taylor series used to derive the bias due to f^*. As the sample size increases from $n = 700$ (Figure 1) to $n = 2000$ (Figure 2), the agreement between the theoretical and the Monte Carlo MSE improves. Even at $n = 700$, the Monte Carlo MSE at m^{THR} is less than 1.05 times the Monte Carlo MSE at its minimizer, m^{MC} (see Table I). The close-up for $d = 0.3$, $\phi = 0.3$ and $n = 2000$ (Figure 3) reveals that the Monte Carlo MSE is not statistically significantly different from the theoretical MSE in a region of values of m near m^{MC}.

The results were similar to the above for the model with $(d, \phi) = (0.3, 0.1)$. However, when ϕ was increased to 0.9, with d remaining at 0.3 (see Figures 4 and 5), agreement between the Monte Carlo and theoretical MSEs was not good even at $n = 2000$, although it improved as the sample size was increased from 700 to 2000. This discrepancy is presumably due to the inaccuracy of the Taylor series approximation for the bias term. In spite of this, for $n \geqslant 700$, the Monte Carlo MSE at m^{THR} never exceeded 1.14 times the Monte Carlo MSE at its minimizer, m^{MC} (see Table II).

To examine the impact of the memory parameter on the performance of the estimate of d, we present boxplots $\hat{d}^{\mathrm{MC}} - d$ (Figure 6) for the models $(d, \phi) = (0.3, 0.3)$ and $(d, \phi) = (0.49, 0.3)$, where $\hat{d}^{\mathrm{MC}} = \hat{d}_{m^{\mathrm{MC}}}$ is the (Monte Carlo) optimal estimator of d. For each sample size considered ($n = 200, 700, 2000$), the distributions of $\hat{d}^{\mathrm{MC}} - d$ seem virtually indistinguishable as d is changed from 0.3 to 0.49. This finding is consistent with the fact that the theoretical MSE (Equation (8)) does not depend on d.

In Tables I and II, we present detailed results on the estimation of d for the models $(d, \phi) = (0.3, 0.3)$ and $(d, \phi) = (0.3, 0.9)$, respectively. We consider the

TABLE I

PERFORMANCE OF GPH AND OPTIMAL ESTIMATORS OF d FOR ARIMA(1, d, 0) MODEL WITH $d = 0.3$, $\phi = 0.3$

		$\hat{d}^{\mathrm{THR}}$	$\hat{d}^{\mathrm{MC}}$	$\hat{d}^{\mathrm{GPH}}$	min (THR MSE)/ min (MC MSE)
$n = 200$					0.71
	m	30	39	14	
	Mean ($\hat{d}$)	0.368	0.393	0.327	
	MSE ($\hat{d}$)	0.0259	0.0245	0.0543	
$n = 700$					0.88
	m	81	85	26	
	Mean ($\hat{d}$)	0.343	0.345	0.319	
	MSE($\hat{d}$)	0.00759	0.00725	0.0245	
$n = 1500$					0.81
	m	148	146	39	
	Mean ($\hat{d}$)	0.331	0.329	0.310	
	MSE($\hat{d}$)	0.00429	0.00426	0.0147	
$n = 2000$					0.95
	m	187	196	45	
	Mean ($\hat{d}$)	0.326	0.329	0.310	
	MSE($\hat{d}$)	0.00293	0.00288	0.0108	

TABLE II
PERFORMANCE OF GPH AND OPTIMAL ESTIMATORS OF d FOR ARIMA(1, d, 0) MODEL WITH $d = 0.3$, $\phi = 0.9$

		$\hat{d}^{\text{THR}}$	$\hat{d}^{\text{MC}}$	$\hat{d}^{\text{GPH}}$	min (THR MSE)/ min (MC MSE)
$n = 200$					0.43
	m	4	8	14	
	Mean ($\hat{d}$)	0.569	0.723	0.865	
	MSE ($\hat{d}$)	0.505	0.206	0.372	
$n = 700$					0.55
	m	11	14	26	
	Mean ($\hat{d}$)	0.417	0.475	0.622	
	MSE($\hat{d}$)	0.0918	0.0858	0.1275	
$n = 1500$					0.66
	m	20	24	39	
	Mean ($\hat{d}$)	0.386	0.409	0.501	
	MSE($\hat{d}$)	0.0440	0.0387	0.0554	
$n = 2000$					0.70
	m	25	31	45	
	Mean ($\hat{d}$)	0.380	0.405	0.466	
	MSE($\hat{d}$)	0.0298	0.0288	0.0382	

three regression estimators $\hat{d}^{\text{THR}}$, $\hat{d}^{\text{MC}}$ and $\hat{d}^{\text{GPH}}$ corresponding to the choices $m = m^{\text{THR}}$, $m = m^{\text{MC}}$ and $m = n^{1/2}$, respectively. Note that neither $\hat{d}^{\text{THR}}$ nor $\hat{d}^{\text{MC}}$ can be computed in practice, since both would require knowledge of the true spectral density. For each sample size and estimator, we present the value of m which was used, the average value of the estimator, and the Monte Carlo MSE of the estimator. In addition, for each sample size we present the ratio of the minimum theoretical MSE to the minimum of the Monte Carlo MSE. For both $\phi = 0.3$ and $\phi = 0.9$, the bias and Monte Carlo MSE of all estimators decrease as the sample size increases, although bias and MSE are much larger for $\phi = 0.9$ than for $\phi = 0.3$. The Monte Carlo ratio $\text{MSE}(\hat{d}^{\text{THR}})/\text{MSE}(\hat{d}^{\text{MC}})$ approaches unity as the sample size increases, as does the ratio of the minimum theoretical MSE to the minimum of the Monte Carlo MSE. These findings are consistent with a good asymptotic agreement between the theoretical and Monte Carlo MSEs. On the other hand, the Monte Carlo ratio $\text{MSE}(\hat{d}^{\text{GPH}})/\text{MSE}(\hat{d}^{\text{MC}})$ is typically substantially greater than unity. This is indicative of the sub-optimality of the choice $m = n^{1/2}$. For example, in the case $\phi = 0.3$, the Monte Carlo ratio $\text{MSE}(\hat{d}^{\text{GPH}})/\text{MSE}(\hat{d}^{\text{MC}})$ increases from 2.22 at $n = 200$ to 3.75 at $n = 2000$. In the case $\phi = 0.9$, the ratio actually decreases from 1.81 at $n = 200$ to 1.33 at $n = 2000$. The decrease can be explained by noting that, for the sample sizes considered here, $n^{1/2}$ is larger than the optimal value of m, but the gap narrows as the sample size increases. Eventually, if larger sample sizes are considered, the ratio $n^{1/2}/m^{\text{MC}}$ should approach zero and the ratio $\text{MSE}(\hat{d}^{\text{GPH}})/\text{MSE}(\hat{d}^{\text{MC}})$ should approach infinity, assuming that $m^{\text{MC}} = \text{O}(n^{4/5})$.

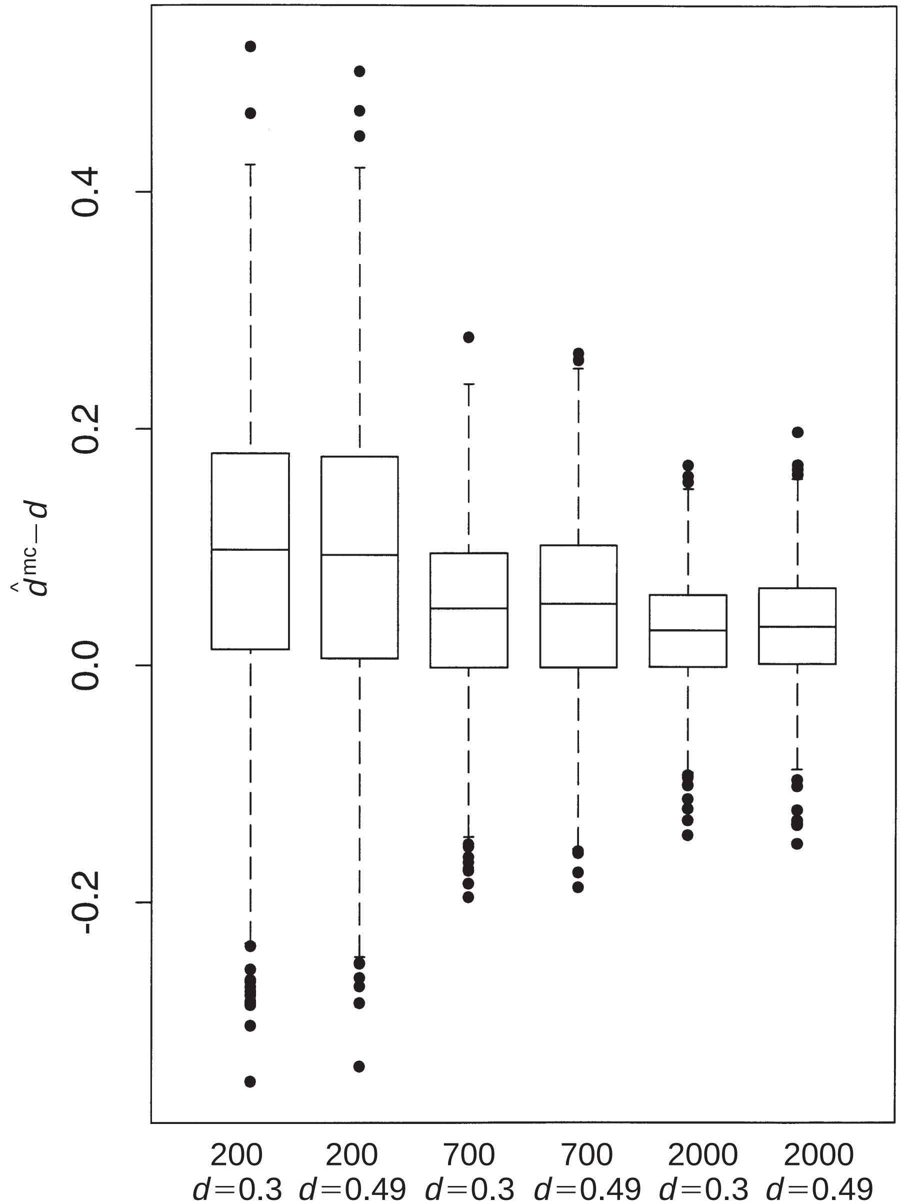

FIGURE 6. Boxplots of $\hat{d}^{\mathrm{MC}} - d$, $\phi = 0.3$, $d = 0.3$ and $d = 0.49$, $n = 200$, 700, 2000.

Figure 7 (for $d = 0.3$, $\phi = 0.3$) and Figure 8 (for $d = 0.3$, $\phi = 0.9$) present boxplots of $\hat{d}^{\mathrm{THR}}$, $\hat{d}^{\mathrm{GPH}}$ and $\hat{d}^{\mathrm{MC}}$ for $n = 200$ and $n = 2000$. These boxplots once again underscore the sub-optimality of $\hat{d}^{\mathrm{GPH}}$. Nevertheless, it must be stressed that the theoretically optimal estimator $\hat{d}^{\mathrm{THR}}$ cannot be computed in

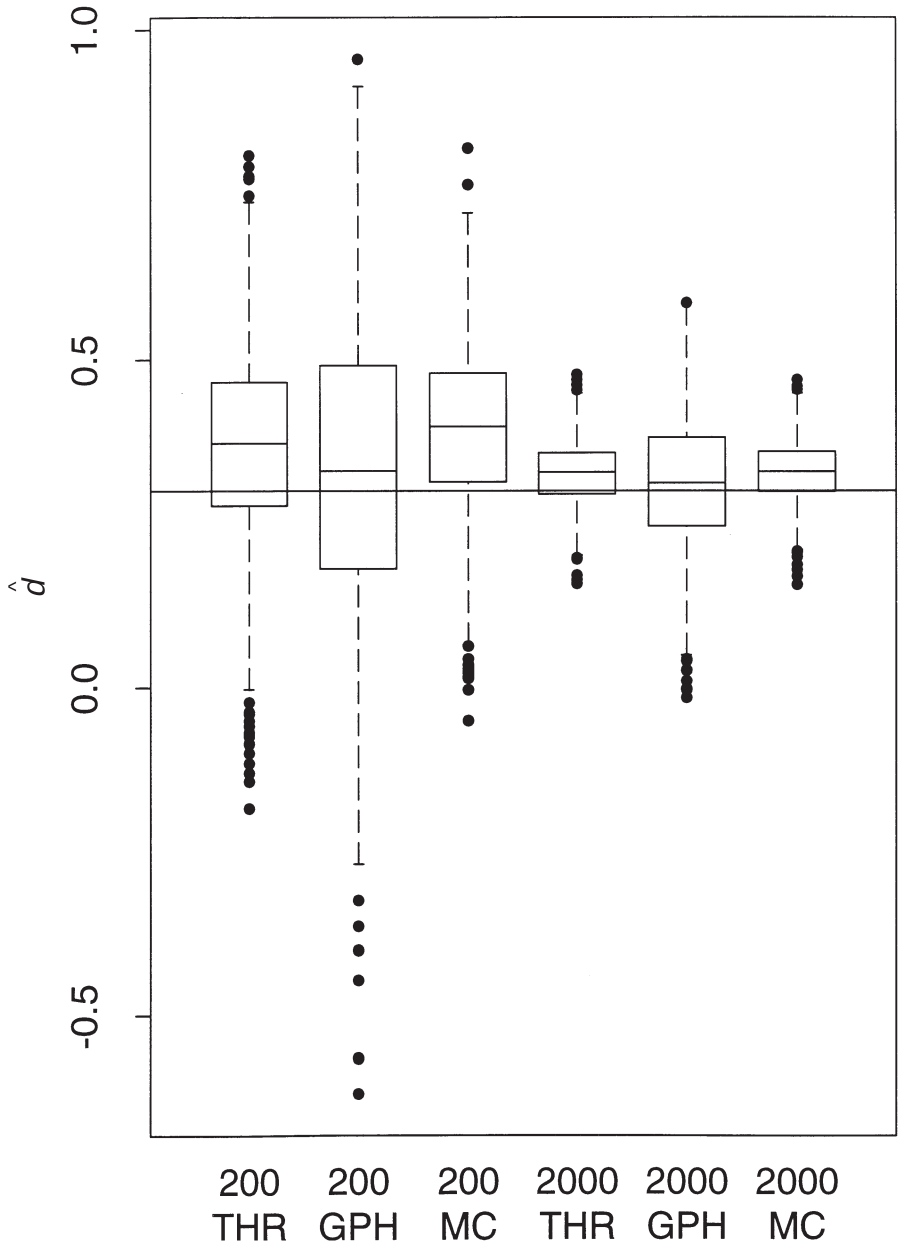

FIGURE 7. Boxplots of $\hat{d}^{\mathrm{THR}}$, $\hat{d}^{\mathrm{GPH}}$, $\hat{d}^{\mathrm{MC}}$, $d = 0.3$, $\phi = 0.3$, $n = 200,\ 2000$.

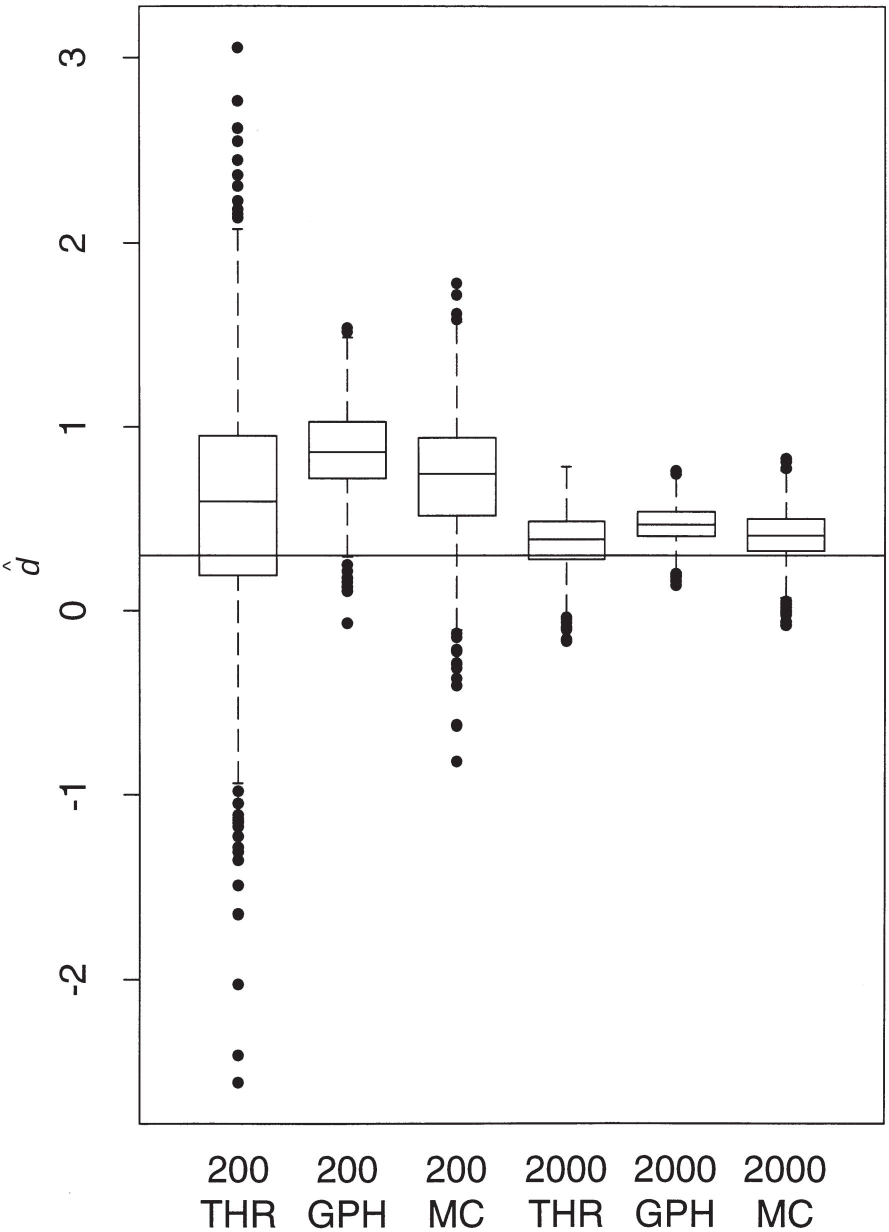

FIGURE 8. Boxplots of $\hat{d}^{\mathrm{THR}}$, $\hat{d}^{\mathrm{GPH}}$, $\hat{d}^{\mathrm{MC}}$, $d = 0.3$, $\phi = 0.9$, $n = 200$, 2000.

practice since the theoretical MSE cannot be evaluated without knowledge of the true spectrum.

4. DISCUSSION

In Hurvich and Beltrao (1994), which concerns the data-driven selection of m, a theorem was presented on the asymptotic equivalence between the mean squared error of $\hat{d}$ and the mean integrated squared error of the log of the corresponding ARIMA(0, $\hat{d}$, 0) spectral density estimate. One of the assumptions of the theorem was that $m = \mathrm{o}(n^{4/5})$. This permitted the use of theorems from Robinson (1995) which also made this assumption, but the assumption is now seen to be much too restrictive for our purposes since we have shown here (Equation (9)) that the optimal value of m is $\mathrm{O}(n^{4/5})$ if $f^{*\prime\prime}(0) \neq 0$. Fortunately, it is easily seen from examining its proof that the theorem from Hurvich and Beltrao (1994) continues to hold under the far less restrictive condition

$$\frac{n^{2\alpha-2} E(\hat{d} - d)}{\mathrm{MSE}(\hat{d})} \to 0 \tag{10}$$

where the number of Fourier frequencies used in computing the mean integrated squared error is $\mathrm{O}(n^{\alpha})$. Indeed, using Equations (6) and (8) for $E(\hat{d} - d)$ and $\mathrm{MSE}(\hat{d})$, it follows that for any $\alpha < 4/5$ (such as the value $\alpha = 0.6$, which was used in the simulations of Hurvich and Beltrao (1994)) Equation (10) holds for $m = \mathrm{O}(n^{\beta})$ with any β, $0 < \beta < 1$.

APPENDIX

PROOF OF LEMMA 1. By Condition 2, for $1 \leqslant j \leqslant m$ there exist ξ_j with $0 \leqslant \xi_j \leqslant \omega_j$ such that

$$\log f_j^* = \log f_0^* + \frac{\omega_j^2}{2} \frac{f^{*\prime\prime}(0)}{f_0^*} + \frac{\omega_j^3}{6} g(\xi_j)$$

where

$$g(\omega) = \frac{f^{*\prime\prime\prime}(\omega)}{f^*(\omega)} - \frac{3 f^{*\prime}(\omega) f^{*\prime\prime}(\omega)}{\{f^*(\omega)\}^2} + \frac{2\{f^{*\prime}(\omega)\}^3}{\{f^*(\omega)\}^3}.$$

Thus,

$$\sum_{j=1}^{m} a_j \log f_j^* = \frac{f^{*\prime\prime}(0)}{2 f^*(0)} \sum_{j=1}^{m} a_j \omega_j^2 + R_1$$

where

$$R_1 = \sum_{j=1}^{m} a_j \frac{\omega_j^3}{6} g(\xi_j).$$

Since $|g(\xi_j)|$ is bounded uniformly in j for sufficiently large n, and since $a_j = \mathrm{O}(\log m)$, we find that

$$R_1 = \mathrm{O}(n^{-3} \log m \sum_{j=1}^{m} j^3) = \mathrm{O}(n^{-3} m^4 \log m).$$

From Hurvich and Beltrao (1994),

$$a_j = \log j - \frac{1}{m} \sum_{k=1}^{m} \log k + \mathrm{O}\left(\frac{m^2}{n^2}\right)$$

where the remainder term is uniform in j. Using this together with the formulas

$$\sum_{j=1}^{m} j^2 \log j = \tfrac{1}{6} m(m+1)(2m+1) \log m - \frac{m^3}{9} + \mathrm{o}(m^3)$$

$$\frac{1}{m} \sum_{k=1}^{m} \log k = \log m - 1 + \mathrm{o}(1) \qquad \sum_{j=1}^{m} j^2 = \tfrac{1}{6} m(m+1)(2m+1)$$

it follows that

$$\begin{aligned}\sum_{j=1}^{m} a_j \log f_j^* &= \frac{2\pi^2}{n^2} \frac{f^{*\prime\prime}(0)}{f^*(0)} \left\{ \sum_{j=1}^{m} j^2 \log j - \frac{1}{m} \sum_{k=1}^{m} \log k \sum_{j=1}^{m} j^2 + \mathrm{O}\left(\frac{m^2}{n^2}\right) \sum_{j=1}^{m} j^2 \right\} \\ &\quad + \mathrm{O}(n^{-3} m^4 \log m) \\ &= \frac{2\pi^2}{n^2} \frac{f^{*\prime\prime}(0)}{f^*(0)} \left\{ -\frac{m^3}{9} + \frac{m(m+1)(2m+1)}{6} + \mathrm{o}(m^3) \right\}.\end{aligned}$$

Since $S_{xx} = m\{1 + \mathrm{o}(1)\}$ and since $n^{-3} m^3 \log m = \mathrm{o}(m^2/n^2)$ by Condition 1, we conclude that

$$\begin{aligned}-\frac{1}{2S_{xx}} \sum_{j=1}^{m} a_j \log f_j^* &= -\frac{\pi^2}{mn^2} \frac{f^{*\prime\prime}(0)}{f^*(0)} \left\{ \frac{4m^3}{18} + \mathrm{o}(m^3) \right\} + \mathrm{O}(n^{-3} m^3 \log m) \\ &= \frac{-2\pi^2}{9} \frac{f^{*\prime\prime}(0)}{f^*(0)} \frac{m^2}{n^2} + \mathrm{o}(m^2/n^2).\end{aligned}$$

PROOF OF LEMMA 2. The following argument is adapted from Robinson (1995). Define $\chi_j = \log(I_j/f_j) - E\{\log(I_j/f_j)\}$. It is desired to show that $E[\chi_j \chi_k] = \mathrm{O}(\alpha_{jk}^2)$, uniformly for $\log^2 m \leqslant k < j \leqslant m$. In the remainder of this proof, any mention of uniformity is taken to mean that the property holds uniformly for $\log^2 m \leqslant k < j \leqslant m$. Define $\psi = \Sigma^{-1}$, where $\Sigma = \mathrm{cov}(\nu)$. Partition Σ and ψ as

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \qquad \psi = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix}$$

where the Σ_{ij} and ψ_{ij} $(i, j = 1, 2)$ are 2×2 matrices. By the formulas for the inverse of a partitioned matrix,

$$\psi_{11} = \Sigma_{11}^{-1} + \Sigma_{11}^{-1} \Sigma_{12} (\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})^{-1} \Sigma_{21} \Sigma_{11}^{-1}$$

$$\psi_{12} = -\Sigma_{11}^{-1} \Sigma_{12} (\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})^{-1}$$

$$\psi_{21} = \psi_{12}' \qquad \psi_{22} = (\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})^{-1}.$$

From Robinson (1995, Theorem 2, parts (a)–(d)), we have $\Sigma = \frac{1}{2}I_4^* = o(1)$ uniformly, where I_4^* is a 4×4 identity matrix. We also have that $\Sigma_{11}^{-1} = O(1)$ and $(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} = O(1)$ uniformly. Using these results together with the formula for ψ_{12} given above, we conclude that $\psi_{12} = O(\alpha_{jk})$ uniformly.

Define

$$\tilde{\psi} = \begin{bmatrix} \psi_{11} & 0 \\ 0 & \psi_{22} \end{bmatrix}$$

and define $\overline{\psi} = \psi - \tilde{\psi}$. We have

$$\begin{aligned} E[\chi_j\chi_k] &= (2\pi)^{-2}|\psi|^{1/2}\int\!\!\int\!\!\int\!\!\int \chi_j\chi_k \exp\left(-\frac{\nu'\psi\nu}{2}\right) d\nu \\ &= (2\pi)^{-2}|\psi|^{1/2}\int\!\!\int\!\!\int\!\!\int \chi_j\chi_k \exp\left(-\frac{\nu'\tilde{\psi}\nu}{2}\right) d\nu \\ &\quad + (2\pi)^{-2}|\psi|^{1/2}\int\!\!\int\!\!\int\!\!\int \chi_j\chi_k \exp\left(-\frac{\nu'\tilde{\psi}\nu}{2}\right)\left\{\exp\left(-\frac{\nu'\overline{\psi}\nu}{2}\right) - 1\right\} d\nu. \end{aligned} \tag{A1}$$

We will show that the two terms on the right-hand side of (A1) are $O(\alpha_{jk}^2)$ uniformly, so that $E[\chi_j\chi_k] = O(\alpha_{jk}^2)$, uniformly.

We now consider the first term on the right-hand side of (A1). If we define $\nu_{(j)} = (\nu_1, \nu_2)'$ and $\nu_{(k)} = (\nu_3, \nu_4)'$ then this term can be written as

$$(2\pi)^{-2}|\psi|^{1/2}\int\!\!\int \chi_j \exp\left(-\frac{\nu_{(j)}'\psi_{11}\nu_{(j)}}{2}\right) d\nu_{(j)} \int\!\!\int \chi_k \exp\left(-\frac{\nu_{(k)}'\psi_{22}\nu_{(k)}}{2}\right) d\nu_{(k)}. \tag{A2}$$

We will show that the first double integral in (A2) is $O(\alpha_{jk}^2)$ uniformly, and that the second is $O(1)$, uniformly. The first integral is

$$\int\!\!\int \chi_j \exp\left(-\frac{\nu_{(j)}'\Sigma_{11}^{-1}\nu_{(j)}}{2}\right) \exp\left(-\frac{\nu_{(j)}'M_{11}\nu_{(j)}}{2}\right) d\nu_{(j)} \tag{A3}$$

where $M_{11} = \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{21}\Sigma_{11}^{-1}$. The entries of M_{11} are all $O(\alpha_{jk}^2)$, uniformly.

Note that $|\nu_{(j)}'M_{11}\nu_{(j)}| \leqslant 3K_{11}\|\nu_{(j)}\|^2$, where K_{11} is the largest absolute entry of M_{11} and $\|\cdot\|$ is the Euclidean norm. Since by the mean value theorem $|e^u - 1| \leqslant |u|e^{|u|}$ for all u, we obtain

$$\exp\left(-\frac{\nu_{(j)}'M_{11}\nu_{(j)}}{2}\right) = 1 + O\{K_{11}\|\nu_{(j)}\|^2 \exp(\tfrac{3}{2}K_{11}\|\nu_{(j)}\|^2)\}. \tag{A4}$$

Thus, (A3) is equal to

$$\int\!\!\int \chi_j \exp\left(-\frac{\nu_{(j)}\Sigma_{11}^{-1}\nu_{(j)}}{2}\right) + O\left\{\int\!\!\int |\chi_j| K_{11}\|\nu_{(j)}\|^2 \exp\left(-\frac{\nu_{(j)}'(\Sigma_{11}^{-1} - 3K_{11}I_2^*)\nu_{(j)}}{2}\right)\right\} \tag{A5}$$

where I_2^* is a 2×2 identity matrix. The first term of (A5) is zero since it is the expectation of χ_j. Since $K_{11} = O(\alpha_{jk}^2) = O\{(\log^2 j)/k^2\}$ uniformly by Robinson (1995, Theorem 2, parts (a)–(d)), we see that $\Sigma_{11}^{-1} - 3K_{11}I_2^*$ is positive definite for sufficiently large n, uniformly. Since $(\Sigma_{11} - 3K_{11}I_2^*)^{-1} = 2I_2^* + o(1)$ uniformly, we conclude that the second term in (A5) is $O(\alpha_{jk}^2)$. Since $\psi_{22} = (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} = 2I_2^* + o(1)$ uniformly, we conclude that the second double integral in (A2) is $O(1)$ uniformly. Since

$\Sigma = (1/2)I_4^* + o(1)$ uniformly, it follows that $|\psi| = O(1)$ uniformly, and hence that the first term in (A1) is $O(\alpha_{jk}^2)$ uniformly.

Next, we consider the second term on the right-hand side of (A1). By the mean value theorem, we have $|e^u - 1 - u| \leqslant (1/2)u^2 e^{|u|}$ for all u. Thus,

$$\exp\left(-\frac{\nu'\overline{\psi}\nu}{2}\right) - 1 = -\frac{\nu'\overline{\psi}\nu}{2} + O\{K_1^2\|\nu\|^4 \exp(2K_1\|\nu\|^2)\}$$

where K_1 is the largest absolute entry of $\overline{\psi}$. Note that $K_1 = O(\alpha_{jk})$ uniformly. Thus, the second term on the right-hand side of (A1) is

$$(2\pi)^{-2}|\psi|^{1/2}\int\!\!\int\!\!\int\!\!\int \chi_j\chi_k \exp\left(-\frac{\nu'\tilde{\psi}\nu}{2}\right)\left(-\frac{\nu'\overline{\psi}\nu}{2}\right) d\nu$$
$$+ O\left\{\alpha_{jk}^2 \int\!\!\int\!\!\int\!\!\int \|\nu\|^4 |\chi_j\chi_k| \exp\left(-\frac{\nu'(\tilde{\psi} - 4K_1 I_4^*)\nu}{2}\right) d\nu\right\}. \qquad \text{(A6)}$$

The first term of (A6) is zero. To see this, note first that it is a linear combination of $E_{\tilde{\psi}}[\chi_j\chi_k A_j A_k]$, $E_{\tilde{\psi}}[\chi_j\chi_k A_j B_k]$ $E_{\tilde{\psi}}[\chi_j\chi_k B_j A_k]$ and $E_{\tilde{\psi}}[\chi_j\chi_k B_j B_k]$ where $E_{\tilde{\psi}}$ denotes the expectation assuming that ν is multivariate normal with mean zero and covariance matrix $\tilde{\psi}$. But $\operatorname{cov}(\nu) = \tilde{\psi}$ imples that the vectors (A_j, B_j) and (A_k, B_k) are independent. Thus, for example, $E_{\tilde{\psi}}[\chi_j\chi_k A_j A_k] = E_{\tilde{\psi}}[\chi_j A_j]E_{\tilde{\psi}}[\chi_k A_k]$, and both of these expectations are zero because χ_j and χ_k are even functions of (A_j, B_j) and (A_k, B_k) respectively, and because the densities for (A_j, B_j) and (A_k, B_k) are also even functions.

Since $K_1 = O(\alpha_{jk}) = o(1)$ uniformly, we conclude that $\tilde{\psi} - 4K_1 I_4^* = 2I_4^* + o(1)$ uniformly. Thus, the second term of (A6) is $O(\alpha_{jk}^2)$ uniformly and the lemma is proved.

Proof of Lemma 3. The discrete Fourier transform of $\{y_t\}$ is given by

$$J_j = \frac{1}{(2\pi n)^{1/2}} \sum_{t=0}^{n-1} y_t \exp(i\omega_j t) = A_j + iB_j$$

so that $I_j = |J_j|^2$. Note that

$$\alpha_{jk}^2 \leqslant \frac{1}{f_j f_k}\{\operatorname{cov}^2(A_j, A_k) + \operatorname{cov}^2(A_j, B_k) + \operatorname{cov}^2(B_j, A_k) + \operatorname{cov}^2(B_j, B_k)\}$$
$$= \frac{1}{2f_j f_k}\{|E(J_j\overline{J}_k)|^2 + |E(J_jJ_k)|^2\}.$$

Let

$$W(\lambda) = \frac{1}{(2\pi n)^{1/2}} \sum_{t=0}^{n-1} y_t \exp\{i\lambda(t+1)\}$$

as in Robinson (1995). Since $J_j = e^{-i\omega_j} W(\omega_j)$, we have

$$\alpha_{jk}^2 \leqslant \frac{1}{2f_j f_k}\{|E[W(\omega_j)\overline{W(\omega_k)}]|^2 + |E[W(\omega_j)W(\omega_k)]|^2\} = O\left(\frac{\log^2 j}{k^2}\right)$$

uniformly for $1 \leqslant k < j \leqslant m$ by Robinson (1955, Theorem 2, parts (c) and (d)).

Proof of Lemma 4. For $j = 1$, we have from Hurvich and Beltrao (1993, Theorem 1) that $\lim_{n\to\infty} E(I_1/f_1) > 0$. Define

$$F_n(\lambda) = \frac{1}{2\pi n}\frac{\sin^2(n\lambda/2)}{\sin^2(\lambda/2)}$$

the Fejer kernel. Since $F_n(\lambda) \geqslant 0$, and since f^* is continuous at zero, there exist finite positive constants C_1 and C_2 such that for all j with $2 \leqslant j \leqslant m$, and for all sufficiently large n,

$$\begin{aligned}
E\left(\frac{I_j}{f_j}\right) &= \frac{1}{f_j}\int_{-\pi}^{\pi} F_n(\omega_j - \mu) f(\mu)\, d\mu \\
&\geqslant \int_{\omega_j - 2\pi/n}^{\omega_j + 2\pi/n} \frac{F_n(\omega_j - \mu) f(\mu)}{f_j}\, d\mu \\
&\geqslant C_1 \int_{\omega_j - 2\pi/n}^{\omega_j + 2\pi/n} F_n(\omega_j - \mu)\, d\mu \\
&= C_1 \int_{-2\pi}^{2\pi} \frac{1}{n}\frac{1}{2\pi n}\frac{\sin^2(\lambda/2)}{\sin^2\{\lambda/(2n)\}}\, d\lambda \\
&= C_2 \int_{-2\pi}^{2\pi} \frac{\{\lambda/(2n)\}^2}{\sin^2\{\lambda/(2n)\}}\frac{\sin^2(\lambda/2)}{(\lambda/2)^2}\, d\lambda \\
&\geqslant C_2 \int_{-2\pi}^{2\pi} \frac{\sin^2(\lambda/2)}{(\lambda/2)^2}\, d\lambda > 0.
\end{aligned}$$

PROOF OF LEMMA 5. For any $\alpha > 0$ and $\beta > 0$ there exist finite constants C_1 and C_2 such that, for all x,

$$|\log|x|| \leqslant C_1|x|^{-\beta} + C_2|x|^{\alpha}.$$

Thus,

$$\begin{aligned}
E\left[\left\{\log\left(\frac{I_j}{f_j}\right)\right\}^2\right] &\leqslant C_1^2 E\left(\left|\frac{I_j}{f_j}\right|^{-2\beta}\right) + C_2^2 E\left(\left|\frac{I_j}{f_j}\right|^{2\alpha}\right) \\
&\quad + 2C_1C_2\left\{E\left(\left|\frac{I_j}{f_j}\right|^{-2\beta}\right)\right\}^{1/2}\left\{E\left(\left|\frac{I_j}{f_j}\right|^{2\alpha}\right)\right\}^{1/2}.
\end{aligned}$$

We set $\alpha = 1/2$ and $\beta = 1/8$. From Robinson (1995, Theorem 2, part (a)) it follows that

$$\lim_n \sup_{1\leqslant j\leqslant m} E\left(\left|\frac{I_j}{f_j}\right|^{2\alpha}\right) < \infty.$$

We now show that

$$\lim_n \sup_{1\leqslant j\leqslant m} E\left(\left|\frac{I_j}{f_j}\right|^{-2\beta}\right) < \infty.$$

Let Σ_{jn} be the 2×2 covariance matrix of $g_j = (A_j/f_j^{1/2}, B_j/f_j^{1/2})'$. Then there exists an orthonormal matrix P_{jn} such that $P_{jn}\Sigma_{jn}P'_{jn} = \operatorname{diag}(\lambda_{jn1}, \lambda_{jn2})$, where $0 < \lambda_{jn1} \leqslant \lambda_{jn2} < \infty$ are the eigenvalues of Σ_{jn}. Furthermore

$$I_j/f_j = g_j'g_j = g_j'P'_{jn}P_{jn}g_j = h_j'h_j$$

where $h_j = P_{jn} g_j$ is a bivariate Gaussian vector with mean zero and covariance matrix diag $(\lambda_{jn1}, \lambda_{jn2})$. Also,

$$\lambda_{jn2} \geqslant \max\left\{E\left(\frac{A_j^2}{f_j}\right), E\left(\frac{B_j^2}{f_j}\right)\right\} \geqslant \tfrac{1}{2}E\left(\frac{I_j}{f_j}\right)$$

and thus by Lemma 4

$$\lim_n \inf_{1\leqslant j\leqslant m} \lambda_{jn2} > 0. \tag{A7}$$

Thus,

$$E(I_j/f_j)^{-1/4} = E(h_j' h_j)^{-1/4} \leqslant E(h_{j2}^2)^{-1/4} = E|h_{j2}|^{-1/2}. \tag{A8}$$

Since h_{j2} is normally distributed with mean zero and variance λ_{jn2}, we conclude from Equations (A7) and (A8) that

$$\lim_n \sup_{1\leqslant j\leqslant m} E(I_j/f_j)^{-1/4} \leqslant \lim_n \sup_{1\leqslant j\leqslant m} E|h_{j2}|^{-1/2} < \infty.$$

PROOF OF LEMMA 6 AND 7. The proofs of Lemmas 6 and 7 are very similar to the proof of Lemma 2, so we omit these proofs to save space.

PROOF OF LEMMA 8

$$\begin{aligned}\left|\frac{1}{2S_{xx}}\sum_{j=1}^{m} a_j E(\varepsilon_j)\right| &\leqslant \left|\frac{1}{2S_{xx}}\sum_{j=1}^{\log^2 m} a_j E(\varepsilon_j)\right| + \left|\frac{1}{2S_{xx}}\sum_{j=1+\log^2 m}^{m} a_j E(\varepsilon_j)\right| \\ &= \mathrm{O}\left(\frac{\log^3 m}{m}\right) + \mathrm{O}\left(m^{-1}\log m \sum_{1+\log^2 m}^{m} \frac{\log j}{j}\right) = \mathrm{O}\left(\frac{\log^3 m}{m}\right)\end{aligned}$$

by Lemmas 5 and 6.

PROOF OF THEOREM 2. First, we note that

$$m^{1/2}(\hat{d} - d) = -\frac{m^{1/2}}{2S_{xx}}\sum_{j=1}^{m} a_j \log f_j^* - \frac{m}{2S_{xx}}\frac{1}{m^{1/2}}\sum_{j=1}^{m} a_j \varepsilon_j.$$

Using Lemma 1, we obtain

$$-\frac{m^{1/2}}{2S_{xx}}\sum_{j=1}^{m} a_j \log f_j^* = \mathrm{o}(1).$$

Now, consider

$$\frac{1}{m^{1/2}}\sum_{j=1}^{m} a_j \varepsilon_j \equiv T_1 + T_2 + T_3 \tag{A9}$$

where

$$T_1 = \frac{1}{m^{1/2}}\sum_{j=1}^{\log^8 m} a_j \varepsilon_j$$

$$T_2 = \frac{1}{m^{1/2}} \sum_{j=1+\log^8 m}^{m^{0.5+\delta}} a_j \varepsilon_j$$

and

$$T_3 = \frac{1}{m^{1/2}} \sum_{j=1+m^{0.5+\delta}}^{m} a_j \varepsilon_j$$

where $0 < \delta < 0.5$. We have

$$\mathrm{pr}\,(|T_1| > \varepsilon) \leqslant \frac{E(|T_1|)}{\varepsilon} \leqslant \frac{1}{\varepsilon} \frac{\mathrm{O}(\log^9 m)}{m^{1/2}} = \mathrm{o}(1).$$

Hence,

$$T_1 = \mathrm{o_P}(1). \tag{A10}$$

From Lemma 5, we have $E(\varepsilon_k^2) = \mathrm{O}(1)$ uniformly in $1 \leqslant k \leqslant m$. Also,

$$|E(\varepsilon_j \varepsilon_k)| \leqslant |\,\mathrm{cov}\,(\varepsilon_j,\, \varepsilon_k)| + |E(\varepsilon_j)E(\varepsilon_k)|$$

$$= \mathrm{O}(\alpha_{jk}^2) + \mathrm{O}\left(\frac{\log j}{j} \frac{\log k}{k}\right) = \mathrm{O}\left(\frac{\log^2 m}{k^2}\right)$$

uniformly in $\log^2 m \leqslant k < j \leqslant m$ by Lemmas 2, 6 and 3. Thus,

$$\begin{aligned}
\mathrm{pr}\,(|T_2| > \varepsilon) &\leqslant E(T_2^2)/\varepsilon^2 \\
&= \frac{1}{\varepsilon^2} \frac{1}{m} \sum_{k=1+\log^8 m}^{m^{0.5+\delta}} a_k^2 E(\varepsilon_k^2) + \frac{2}{\varepsilon^2 m} \sum_{k=1+\log^8 m}^{m^{0.5+\delta}} \sum_{j=k+1}^{m^{0.5+\delta}} a_j a_k E(\varepsilon_j \varepsilon_k) \\
&= \mathrm{O}\left(\frac{\log^2 m\, m^{0.5+\delta}}{m}\right) + \mathrm{O}\left(\frac{\log^2 m}{m} \sum_{k=1+\log^8 m}^{m^{0.5+\delta}} \sum_{j=k+1}^{m^{0.5+\delta}} \frac{\log^2 m}{k^2}\right) \\
&= \mathrm{o}(1) + \mathrm{O}\left(\frac{\log^4 m}{m} \frac{m^{0.5+\delta}}{\log^8 m}\right) = \mathrm{o}(1).
\end{aligned}$$

Thus,

$$T_2 = \mathrm{o_P}(1). \tag{A11}$$

We now prove that T_3 is asymptotically normal by approximating it with another sequence which is known to be asymptotically normal from Robinson (1995). Let

$$U_j = \log I_j - \log f^*(0) + C + 2d \log \omega_j$$

as defined in Equation (2.4) of Robinson (1995). Then we have

$$U_j = \varepsilon_j + \log\left\{\frac{f^*(\omega_j)}{f^*(0)}\right\} - 2d \log\left\{\frac{|1 - \exp(-i\omega_j)|}{\omega_j}\right\}.$$

Hence,

$$T_3 \equiv T_{31} + T_{32} + T_{33} \tag{A12}$$

where

$$T_{31} = \frac{1}{m^{1/2}} \sum_{j=1+m^{0.5+\delta}}^{m} a_j U_j$$

$$T_{32} = -\frac{1}{m^{1/2}} \sum_{j=1+m^{0.5+\delta}}^{m} a_j \log\left\{\frac{f^*(\omega_j)}{f^*(0)}\right\}$$

and

$$T_{33} = \frac{2d}{m^{1/2}} \sum_{j=1+m^{0.5+\delta}}^{m} a_j \log\left\{\frac{|1-\exp(-i\omega_j)|}{\omega_j}\right\}.$$

Now

$$\begin{aligned} T_{32} &= -\frac{1}{m^{1/2}}\sum_{j=1}^{m} a_j \log\left\{\frac{f^*(\omega_j)}{f^*(0)}\right\} + \frac{1}{m^{1/2}}\sum_{j=1}^{m^{0.5+\delta}} a_j \log\left\{\frac{f^*(\omega_j)}{f^*(0)}\right\} \\ &= -\frac{1}{m^{1/2}}\sum_{j=1}^{m} a_j \log f^*(\omega_j) + \frac{\log f^*(0)}{m^{1/2}}\sum_{j=1}^{m} a_j + \frac{1}{m^{1/2}}\sum_{j=1}^{m^{0.5+\delta}} a_j \log\left\{\frac{f^*(\omega_j)}{f^*(0)}\right\}. \end{aligned} \tag{A13}$$

The first term of the right-hand side of (A13) is o(1) by Lemma 1. The second term is identically zero since $\sum_{j=1}^{m} a_j = 0$. Since

$$\log f^*(\omega_j) = \log f^*(0) + \frac{\omega_j^2}{2}\left[\frac{f^{*\prime\prime}(\xi_j)}{f^*(\xi_j)} - \frac{f^{*\prime}(\xi_j)}{\{f^*(\xi_j)\}^2}\right]$$

for some $0 \leqslant \xi_j \leqslant \omega_j$, we conclude from Condition 2 that $\log\{f^*(\omega_j)/f^*(0)\} = \mathrm{O}(j^2/n^2)$ uniformly for $1 \leqslant j \leqslant m$. Thus, the third term of (A13) is

$$\begin{aligned} \frac{1}{m^{1/2}}\sum_{j=1}^{m^{0.5+\delta}} a_j \log\left\{\frac{f^*(\omega_j)}{f^*(0)}\right\} &= \mathrm{O}\left(\frac{\log m}{m^{1/2}}\sum_{j=1}^{m^{0.5+\delta}} \frac{j^2}{n^2}\right) \\ &= \mathrm{O}\left(\frac{\log m}{m^{1/2}}\frac{1}{n^2}m^{3(0.5+\delta)}\right) = \mathrm{o}(1) \end{aligned}$$

for δ sufficiently small. Hence

$$T_{32} = \mathrm{o}(1). \tag{A14}$$

We now show that $T_{33} = \mathrm{o}(1)$. From page 300 of Hurvich and Beltrao (1994), we have for $1 \leqslant j \leqslant m$ that

$$\log|1-\exp(-i\omega_j)| = \log\omega_j + \tfrac{1}{2}\log\cos\xi_j$$

where $0 \leqslant \xi_j \leqslant \omega_j$ and

$$|\log\cos\xi_j| \leqslant \frac{\omega_m^2}{2}\frac{1}{\cos^2\omega_m} = \mathrm{O}\left(\frac{m^2}{n^2}\right)$$

uniformly in $1 \leqslant j \leqslant m$. Hence

$$\log\left\{\frac{|1-\exp(-i\omega_j)|}{\omega_j}\right\} = \mathrm{O}\left(\frac{m^2}{n^2}\right)$$

uniformly in $1 \leqslant j \leqslant m$. Thus,

$$T_{33}^2 \leqslant \frac{4d^2}{m} \sum_{j=1}^{m} a_j^2 \sum_{j=1}^{m} \left[\log \left\{ \frac{|1 - \exp(-i\omega_j)|}{\omega_j} \right\} \right]^2$$
$$= \mathrm{O}(1/m)\, \mathrm{O}(m)\, \mathrm{O}(m^5/n^4) = \mathrm{o}(1).$$

Thus,

$$T_{33} = \mathrm{o}(1). \tag{A15}$$

We now prove that $T_{31} \xrightarrow{\mathrm{D}} \mathrm{N}(0, \pi^2/6)$. Note that

$$a_j = \mathrm{O}(\log m) \tag{A16}$$

uniformly in $1 \leqslant j \leqslant m$. Also,

$$\sum_{j=1+m^{0.5+\delta}}^{m} a_j^2 = \sum_{j=1}^{m} a_j^2 - \sum_{j=1}^{m^{0.5+\delta}} a_j^2 = m + \mathrm{o}(m) + \mathrm{O}(m^{0.5+\delta} \log^2 m) = m + \mathrm{o}(m). \tag{A17}$$

Finally, using the fact (see Hurvich and Beltrao, 1994, pp. 300–1) that

$$a_j = \log j - \frac{1}{m} \sum_{k=1}^{m} \log k + \mathrm{O}\left(\frac{m^2}{n^2} \right)$$

uniformly in $1 \leqslant j \leqslant m$, we have for any $p \geqslant 2$ that

$$\begin{aligned}
\sum_{j=1+m^{0.5+\delta}}^{m} |a_j|^p &= \sum_{j=1+m^{0.5+\delta}}^{m} \left| \log j - \frac{1}{m} \sum_{k=1}^{m} \log k + \mathrm{O}\left(\frac{m^2}{n^2} \right) \right|^p \\
&\leqslant 2^{p-1} \sum_{j=1+m^{0.5+\delta}}^{m} \left| \log j - \frac{1}{m} \sum_{k=1}^{m} \log k \right|^p + 2^{p-1} \mathrm{O}\left(\frac{m^{2p}}{n^{2p}} \right) \mathrm{O}(m) \\
&= 2^{p-1} \sum_{j=1+m^{0.5+\delta}}^{m} \left| \log j - \frac{1}{m} \sum_{k=1+m^{0.5+\delta}}^{m} \log k - \frac{1}{m} \sum_{k=1}^{m^{0.5+\delta}} \log k \right|^p + \mathrm{O}(m) \\
&\leqslant (2^{p-1})(2^{p-1}) \sum_{j=1+m^{0.5+\delta}}^{m} \left| \log j - \frac{1}{m} \sum_{k=1+m^{0.5+\delta}}^{m} \log k \right|^p \\
&\quad + 2^{p-1} \left(\frac{m^{0.5+\delta} \log m^{0.5+\delta}}{m} \right)^p \mathrm{O}(m) + \mathrm{O}(m) \\
&= \mathrm{O}(m)
\end{aligned} \tag{A18}$$

since

$$\sum_{j=m^{0.5+\delta}}^{m} \left| \log j - \frac{1}{m} \sum_{k=m^{0.5+\delta}}^{m} \log k \right|^p = \mathrm{O}(m)$$

as shown in Robinson (1995, p. 1067). From (A16), (A17) and (A18), we conclude that the sequence $\{a_j\}$ satisfies (5.15) of Robinson (1995), and hence

$$T_{31} \xrightarrow{\mathrm{D}} \mathrm{N}(0, \pi^2/6) \tag{A19}$$

by (5.14) of Robinson (1995). Using (A14), (A15) and (A19) in (A12) yields

$$T_3 \xrightarrow{\mathrm{D}} \mathrm{N}(0, \pi^2/6). \tag{A20}$$

Using (A10), (A11) and (A20) in (A9) proves the theorem.

REFERENCES

GEWEKE, J. and PORTER-HUDAK, S. (1983) The estimation and application of long memory time series models. *J. Time Ser. Anal.* 4, 221–37.

HASLETT, J. and RAFTERY, A. E. (1989) Space–time modelling with long-memory dependence: assessing Ireland's wind power resources (with discussion). *J. R. Stat. Soc., Ser. C* 38, 1–21.

HURVICH, C. M. and BELTRAO, K. I. (1993) Asymptotics for the low-frequency ordinates of the periodogram of a long memory time series. *J. Time Ser. Anal.* 14, 455–72.

—— and —— (1994) Automatic semiparametric estimation of the memory parameter of a long memory time series. *J. Time. Ser. Anal.* 15, 285–302.

KUNSCH, H. R. (1986) Discrimination between monotonic trends and long-range dependence. *J. Appl. Probab.* 23, 1025–30.

ROBINSON, P. M. (1995) Log-periodogram regression of time series with long range dependence. *Ann. Stat.* 23, 1048–72.