Working Paper 93-34
Statistics and Econometrics Series 24
November 1993

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

OPTIMAL SPECTRAL BANDWIDTH FOR LONG MEMORY

Miguel A. Delgado and Peter M. Robinson*

Abstract

For long range dependent time series with a spectral singularity at frequency zero, a theory for optimal bandwidth choice in non-parametric analysis of the singularity was developed by Robinson (1991b). The optimal bandwidths are described and compared with those in case of analysis of a smooth spectrum. They are also analysed in case of fractional ARIMA models and calculated as a function of the self similarity parameter in some special cases. Feasible data dependent approximations to the optimal bandwidth are discussed.

Key Words

Long range dependence; Spectrum; Self similarity parameter; Semiparametric estimation; Optimal bandwidth choice; ARFIMA models.

*Delgado, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid; Robinson, London School of Economics. This article is based on research funded by the Economic Social Research Council (ESRC) reference number: R000233609, and by the Spanish Dirección General de Investigación Científica y Técnica (DGICYT), reference number: PB92-0247.

1. INTRODUCTION

A theory of optimal bandwidth choice in nonparametric spectral estimation was developed many years ago (see eg. Parzen 1957). This theory in large part precedes the corresponding optimal bandwidth literature for nonparametric probability density and regression estimation, though it has not been developed to the same extent. There are considerable similarities between the two types of theory. In both cases, a nonparametric estimate of an unknown function at a given point of the domain borrows information from neighbouring points. The extent of such information is largely determined by a "bandwidth" number, and the choice of this considerably affects the estimate. Too large a bandwidth tends to be associated with a large bias, too small a bandwidth with a large variance. One usually seeks a bandwidth which balances bias and imprecision. A mathematically simple way of doing this consists of minimizing a form of mean squared error of the nonparametric estimate, either at a particular point of interest, or else averaged across an interval or possibly the whole domain. Typically, a closed form formula for an 'optimal' bandwidth results, depending on the precise way the nonparametric estimate has been implemented and on features of the nonparametric function, in particular, smoothness properties.

In the spectral estimation situation, and the probability density and regression situations, it is typically assumed that the unknown function is at least finite at all points at which it is estimated. This assumption may be controversial in case of spectral estimation. Some plots of spectral estimates exhibit sharp peaks (so that it has long been common practice to use a logarithmic scale), and this could be consistent with a singularity in the spectral density. Likewise, plots of sample autocorrelations are sometimes indicative of a slow rate of decay. Consequently, there has been considerable study of 'long range dependent' parametric and nonparametric models which imply a singularity in the spectral density, typically at zero frequency.

Recently, Robinson (1991 b) has developed some optimality theory for nonparametric frequency domain estimation in case of long range dependence.

The present paper elaborates on and extends his work. The following section briefly reviews related results for "short memory" time series, by which we mean here ones with finite spectral density. Section 3 discusses Robinson's (1991a) optimal bandwidth results for long range dependence. In Section 4 these formulae are further analyzed and numerically illustrated in case of fractional ARIMA (ARFIMA) models. Feasible approximations of the optimal bandwidth are derived in Section 5.

2. OPTIMAL BANDWIDTH FOR THE SMOOTH SPECTRUM ESTIMATE

First we introduce some notation. Denote by X_t, t= 0, ± 1, ±2, ... a discrete parameter covariance stationary time series; for the sake of simplicity we suppose X_t is also Gaussian, though our conclusions have more general relevance. Denote the lag-j autocovariance of X_t by

$$\gamma_j = E\Big((X_j - E(X_0))(X_0 - E(X_0))\Big), \quad j= 0, \pm 1, \pm 2, \ldots,$$

and the spectral density of X_t by

$$f(\lambda) = (2\pi)^{-1} \sum_{j=-\infty}^{\infty} \gamma_j \cos j\lambda, \quad -\pi \le \lambda \le \pi.$$

For a realization of size n, introduce the periodogram

$$I(\lambda) = (2\pi n)^{-1} \Big| \sum_{t=1}^{n} X_t\, e^{it\lambda} \Big|^2. \tag{2.1}$$

All estimates in the paper will depend on $I(\lambda)$ computed at frequencies $\lambda_j = 2\pi j/n$ for integer j. Notice that, $E(X_0)$ is not assumed to be zero (or known) and for $j \neq 0(\mathrm{mod}(n))$, $I(\lambda_j)$ is invariant to any location change in the X_t.

Because we focus on estimation around zero frequency when we deal with long range dependence, we shall consequently give formulae for estimates of a smooth spectral density only at zero frequency; in the long range dependent case, our results go through in case of estimation around another non-zero frequency at which there is known to be a spectral singularity, and of course there is no loss of generality in looking at $\lambda = 0$ in the smooth case.

We suppose now that $0 < f(0) < \infty$ and

$$f(\lambda)/f(0) = 1 + E_\alpha \lambda^\alpha + o(|\lambda|^\alpha), \text{ as } \lambda \longrightarrow 0+, \tag{2.2}$$

for some $\alpha \in (0, 2)$, where $0 < |E_\alpha| < \infty$. This condition essentially says that, in a neighbourhodd of $\lambda = 0$, $f(\lambda)$ satisfies a Lipschitz condition of degree α for $0 < \alpha \le 1$, or $f(\lambda)$ is differentiable and its derivative satisfy a Lipschitz condition of degree $\alpha - 1$, and is zero at $\lambda = 0$, for $1 < \alpha \le 2$. In case $\alpha = 1$ we have $E_1 = \partial \log f(0)/\partial\lambda$ and in case $\alpha = 2$ we have $E_2 = (\partial^2 f(0)/\partial\lambda^2)/(2 f(0))$.

The averaged periodogram form of estimate of $f(0)$ which we consider (see e.g. Brillinger 1975, Robinson 1983) is

$$\hat{f}(0) = 2\pi n^{-1} \sum_{j=1}^{n} K_n(\lambda_j) I(\lambda_j), \tag{2.3}$$

where

$$K_n(\lambda) = \frac{n}{2m} \sum_{j=-\infty}^{\infty} K\left(\frac{n}{2m} (\lambda + 2\pi j)\right). \tag{2.4}$$

Here m is an integer between 1 and n, depending on n in asymptotic theory, while $K(\lambda)$ is a real, even function satisfying

$$\int_{-\infty}^{\infty} |K(\lambda)| \, d\lambda < \infty, \qquad \int_{-\infty}^{\infty} K(\lambda) \, d\lambda = 1. \tag{2.5}$$

m is the bandwidth number, which is regarded as tending to infinity as n does but more slowly, $K(\lambda)$ is the spectral window, a simple leading case of which is

$$K(\lambda) = \begin{cases} (2\pi)^{-1}, & |\lambda| \le \pi \\ 0, & |\lambda| > \pi. \end{cases} \tag{2.6}$$

For future use, introduce also

$$c_\theta = \int_{-\infty}^{\infty} |\lambda|^\theta K(\lambda)\, d\lambda, \tag{2.7}$$

$$d_\phi = \int_{-\infty}^{\infty} |\lambda|^\phi K(\lambda)^2\, d\lambda, \tag{2.8}$$

where it will be taken for granted that the integrals exist for the θ and ϕ values used.

Under the above conditions and additional regularity conditions one has for the scaled mean squared error of $\hat{f}(0)$,

$$E\{(\hat{f}(0) - f(0))/f(0)\}^2 \sim 2\pi d_0 m^{-1} + E_\alpha^2\, c_\alpha^2\, (2\, m\, n^{-1})^{2\alpha}, \tag{2.9}$$

when

$$m^{-1} + m\, n^{-1} \longrightarrow 0 \text{ as } n \longrightarrow \infty. \tag{2.10}$$

This is minimized with respect to m by

$$m_{opt} = \left((\pi\, d_0)/(\alpha\, 2^{2\alpha}\, E_\alpha\, c_\alpha^2) \right)^{1/(2\alpha+1)} \quad n^{2\alpha/(2\alpha+1)}. \tag{2.11}$$

In case K is given by (2.6), we have

$$c_\alpha = \pi^\alpha/(\alpha + 1),\; d_0 = (2\pi)^{-1}, \tag{2.12}$$

and thus,

$$E\left\{[\hat{f}(0) - f(0)]/f(0)\right\}^2 \sim m^{-1} + E_\alpha^2\, \lambda_m^{2\alpha}/(\alpha+1)^2, \tag{2.13}$$

and the optimal m is

$$m_{opt} = \left(\frac{(\alpha+1)^2}{2\alpha\,(2\pi)^{2\alpha}\, E_\alpha^2} \right)^{\frac{1}{2\alpha+1}} n^{\frac{2\alpha}{2\alpha+2}}. \tag{2.14}$$

3. OPTIMAL BANDWIDTH UNDER LONG RANGE DEPENDENCE

We now consider processes with spectrum satisfying

$$f(\lambda) \sim g_H(\lambda) = L(\lambda^{-1})\,\lambda^{1-2H} \text{ as } \lambda \longrightarrow 0+,$$

where $L(\lambda)$ is slowly varying at infinity; that is,

$$\frac{L(t\lambda)}{L(\lambda)} \longrightarrow 1 \text{ as } \lambda \longrightarrow \infty, \quad \text{for all } t>0, \tag{3.1}$$

and $1/2 < H < 1$. Clearly $f(0)$ is now infinite and it is no longer meaningful to estimate it. However, it is of some interest to investigate the impact on optimal bandwidth in case one attempts to estimate $f(0)$ in the incorrect belief that it is finite. In addition Robinson (1991 b) has shown that an optimal type of spectral bandwidth is relevant to the choice of bandwidth in the semiparametric estimate of H proposed by Robinson (1991 a).

The criterion (2.9) is no longer relevant. However, Robinson (1991 b) suggested the extended criterion

$$E\left\{(\hat{F}(\lambda_m) - G_H(\lambda_m))/G_H(\lambda_m)\right\}^2, \tag{3.2}$$

where

$$\hat{F}(\lambda_m) = \frac{2\pi}{n}\sum_{j=1}^{m} I(\lambda_j), \tag{3.3}$$

and

$$G_H(\lambda) = L(\lambda^{-1})\,\lambda^{2(1-H)}/(2(1-H)) \sim \int_0^\lambda g_H(\lambda)\,d\lambda \sim \int_0^\lambda f(\lambda)\,d\lambda, \text{ as } \lambda \longrightarrow 0+. \tag{3.4}$$

Notice that in case K is given by (2.6), $\hat{f}(0) = \hat{F}(\lambda_m)/\lambda_m$ reduces to the left hand side of (2.9). To extend condition (2.2) it is assumed for $\alpha \in (0, 2]$,

$$f(\lambda)/g_H(\lambda) = 1 + E_\alpha(H)\,\lambda^\alpha + o(\lambda^\alpha) \text{ as } \lambda \longrightarrow 0+, \tag{3.5}$$

where $0< |E_\alpha(H)| < \infty$, $1/2< H< 1$. Notice that $E_\alpha(H)$ in general depends on H as well as α, as will be illustrated subsequently.

Consider the case $1/2 < H < 3/4$. Introduce also the regularity condition

$$\sup_{-\Delta\lambda\le \mu \le \delta\lambda} \frac{|f(\lambda) - f(\lambda-\mu)|}{|\mu|\; g_H(|\mu|)} = O(\lambda^{-1}), \text{ as } \lambda \longrightarrow \infty, \tag{3.6}$$

for any $\delta \in (0, 1)$ and $\Delta \in (1, \infty)$; the condition is discussed by Robinson (1991 b). Under these conditions and (2.10), Robinson (1991 b) established that

$$(3.2) \sim 4(1-H)^2 \left[\frac{1}{(3- 4H)\, m} + \left\{\frac{E_\alpha(H)}{2- 2H + \alpha}\right\}^2 \lambda_m^{2\alpha}\right], \text{ as } n \longrightarrow \infty, \tag{3.7}$$

and an optimal m is

$$m_{opt}(H) = \left\{ \frac{(2 - 2H + \alpha)^2}{2\alpha\ (2\pi)^{2\alpha}\ E_\alpha^2(H)(3- 4H)}\right\}^{1/(2\alpha+ 1)} n^{2\alpha/(2\alpha+1)}. \tag{3.8}$$

Notice that both formulae are independent of the slowing varying function L, so that the results have the advantage of being valid when the functional form of L is unknown. Note also that the rate of convergence in (3.8) is identical to that in (2.11), so that long range dependence, in the case $1/2 < H < 3/4$, affects only the multiplying factor in the optimal m. Notice finally that the formulae (3.7) and (3.8) reduce to (2.9) and (2.11) on taking $H=1/2$ and $g_{1/2}(0)= f(0)$.

When $3/4 < H< 1$, $f(\lambda)$ is no longer square-integrable on a neighbourhood of the origin. In this case, instead of (3.6), it is assumed that

$$\gamma_j \sim D_H\ L(j)\ j^{2(H-1)} \text{ as } j \longrightarrow \infty, \tag{3.9}$$

where $D_H= 2\,\Gamma(2(1-H))\cos((1-H)\pi)$. Robinson (1991b) showed that (3.1) and (3.9) implies that

$$\lambda|f(\lambda) - f(\lambda-\mu)| = O\left(|\mu|\ g_H(|\mu|)\right) \text{ as } |\mu| \longrightarrow 0, \tag{3.10}$$

uniformly in $\lambda \in (0, \pi)$. Assumption (3.9) is stronger than (3.6). Furthermore, (3.10) implies that $f(\lambda)$ satisfies an approximate Lip(2-2H) condition outside a neighbourhood of the origin, thereby ruling out long-memory behaviour at nonzero frequencies. Under these assumptions, when 3/4< H< 1, Robinson (1991a) established that

$$(3.2) \sim A_1\ (2\pi m)^{4H-4} + A_2\ (2\pi m)^{2H-2+\alpha}\ n^{-\alpha} + A_3\ (2\pi m)^{2\alpha}\ n^{-2\alpha}, \tag{3.11}$$

where

$$A_1 = 2D_H^2\ (1-H)^2 \left\{\frac{1}{(4H-3)(2H-1)} + \frac{1}{2H^2(4H-1)} + \frac{1}{H^2(4H-1)} + \frac{4\Gamma(2H-1)^2}{\Gamma(4H)}\right\},$$

$$A_2 = -\frac{4D_H\ E_\alpha(H)\ (1-H)^2}{H(2H-1)(2-2H+\alpha)},\quad A_3 = \frac{4E_\alpha(H)^2(1-H)^2}{(2-2H+\alpha)^2},$$

which is minimized with respect to m by

$$m_{opt}(H) \sim A(m_{opt}(H))\ \frac{n^{\frac{\alpha}{2-2H+\alpha}}}{2\pi}\left\{\frac{H^{(2H-2+\alpha)}}{|E_\alpha(H)|}\left[\frac{2-2H+\alpha}{4\alpha(2H-1)}\left(\frac{(2H-2+\alpha)^2}{H^2(2H-1)^2}\right.\right.\right.$$

$$\left.\left.\left. + 16\alpha\ (1-H)\left\{\frac{1}{(4H-3)(2H-1)} - \frac{1}{H^2(4H-1)} - \frac{4\Gamma(2H-1)^2}{\Gamma(4H)}\right\}\right)^{1/2}\right]\right\}^{\frac{1}{2-2H+\alpha}}, \tag{3.12}$$

where $A(m) = L(n)/L(\lambda_m^{-1})$. In general there is no closed form expression for $m_{opt}(H)$ in this case. However, in ARFIMA models discussed in the next section, $A(m) = 1$ all m, and $m_{opt}(H)$ is a function of H and n. When $A(m) \longrightarrow 1$ as $n \longrightarrow$

∞, e.g. $L(\lambda) = |\log \lambda|$ and $m \sim \psi n$, $0 < \psi < 1$, (3.12) also simplifies. Unlike in (3.8), the power of n in (3.12) is a function of H.

4. FRACTIONAL ARIMAS.

In a fractional differenced model we have

$$f(\lambda) = |1 - e^{i\lambda}|^{1-2H} h(\lambda), \tag{4.1}$$

where $0 < h(0) < \infty$. In particular, this class of models includes the ARFIMA(p, H- 1/2, q) model in which

$$h(\lambda) = \frac{\sigma^2}{2\pi} \frac{|b(e^{i\lambda})|^2}{|a(e^{i\lambda})|^2}, \tag{4.2}$$

where

$$a(z) = 1 - \sum_{j=1}^{p} a_j z^j, \quad b(z) = 1 - \sum_{j=1}^{q} b_j z^j, \tag{4.3}$$

all zeros of a and b are outside the unit circle in the complex plane, and $\sigma^2 > 0$.

In general, and as is the case in the ARIMA model, assume that $h(\lambda)$ has first derivative $h'(0) = 0$, and second derivative $h''(0)$. Then, $g_H(\lambda) = C\lambda^{1-2H}$, where $0 < C < \infty$, and

$$\frac{f(\lambda)}{g_H(\lambda)} = \frac{h(\lambda)}{C} \left| \frac{1 - e^{i\lambda}}{\lambda} \right|^{1-2H}$$

$$= \frac{h(\lambda)}{C} \left(\frac{\sin(\lambda/2)}{\lambda/2} \right)^{1-2H}$$

$$\sim C^{-1}\left\{ h(0) + h''(0)\,\lambda^2/2 \right\}\left\{1 - (\lambda/2)^2/6 \right\}^{1-2H}$$

$$\sim C^{-1}\left\{ h(0) + h''(0)\,\lambda^2/2 \right\}\left\{ 1 - (1-2H)\,\lambda^2/24 \right\}$$

$$\sim 1 + \left\{ h''(0)/\,2h(0) + (2H-1)/24 \right\}\lambda^2, \tag{4.4}$$

on taking $C = h(0)$. Thus

$$E_2(H) = h''(0)/\,2h(0) + (2H-1)/24. \tag{4.5}$$

The second component of $E_2(H)$ is positive and takes values zero (when H = 1/2), 1/48 (when H= 3/4), and 1/24 (when H= 1). Notice that $E_2(H)= (2H-1)/24$ in the ARIMA(0, H- 1/2, 0) case. The first component of (4.5) can be positive or negative and it can be large or small.

We can get a more useful picture of the variability in $E_\alpha(H)$ by studying the ARFIMA case. Put

$$a = a(1) = 1- \sum_{j=1}^{p} a_j, \tag{4.6}$$

$$b = b(1) = 1- \sum_{j=1}^{q} b_j, \tag{4.7}$$

$$a' = d\, a(e^{i\lambda})/d\,\lambda\big|_{\lambda=0} = -\,i \sum_{j=1}^{p} j\, a_j, \tag{4.8}$$

$$b' = d\, b(e^{i\lambda})/d\,\lambda\big|_{\lambda=0} = -\,i \sum_{j=1}^{q} j\, b_j, \tag{4.9}$$

$$a'' = d^2\, a(e^{i\lambda})/d\,\lambda^2\big|_{\lambda=0} = \sum_{j=1}^{p} j^2\, a_j, \tag{4.10}$$

$$b'' = d^2\, b(e^{i\lambda})/d\,\lambda^2\big|_{\lambda=0} = \sum_{j=1}^{q} j^2\, b_j. \tag{4.11}$$

It may be shown that

$$\frac{h''(0)}{h(0)} = \left(\frac{\bar{b}''}{\bar{b}} + \frac{b''}{b} + 2\frac{|b'|^2}{|b|^2} \right) - \left(\frac{\bar{a}''}{\bar{a}} + \frac{a''}{a} + 2\frac{|a'|^2}{|a|^2} \right)$$

$$= 2\left\{ \frac{\sum_{j=1}^{q} j^2 b_j}{1-\sum_{j=1}^{q} j^2 b_j} + \left(\frac{\sum_{j=1}^{q} j\, b_j}{1-\sum_{j=1}^{q} j\, b_j} \right)^2 - \frac{\sum_{j=1}^{q} j^2 a_j}{1-\sum_{j=1}^{q} j^2 a_j} - \left(\frac{\sum_{j=1}^{q} j\, a_j}{1-\sum_{j=1}^{q} j\, a_j} \right)^2 \right\}. \quad (4.12)$$

In the AR(1) case we have

$$\frac{h''(0)}{2\,h(0)} = -\left\{ \frac{a_1}{1-a_1} + \left(\frac{a_1}{1-a_1} \right)^2 \right\} = -\frac{a_1}{(1-a_1)^2}, \quad (4.13)$$

and in the AR(2) case

$$\frac{h''(0)}{2\,h(0)} = -\left\{ \frac{a_1 + 4\,a_2}{1-a_1-a_2} + \left(\frac{a_1 + 2\,a_2}{1-a_1-a_2} \right)^2 \right\} = \frac{a_1 - a_1 a_2 + 4\,a_2}{(1-a_1-a_2)^2}. \quad (4.14)$$

Corresponding MA formulae are obtained by replacing a's by b's and changing sign. For the ARMA(1, 1) case

$$\frac{h''(0)}{2\,h(0)} = \frac{b_1}{(1-b_1)^2} - \frac{a_1}{(1-a_1)^2}. \quad (4.15)$$

In the AR(1) case $h''(0)/2h(0)$ approaches minus infinity when a_1 approaches 1, for example, it is -90 when $a_1 = 0.9$ and -990 when $a_1 = 0.99$. For large or moderate a_1, E_{2H} will be dominated by the $h''(0)/\ 2h(0)$ component.

In Figure 1, we plot $m_{opt}(H)n^{-2\alpha/(2\alpha+1)}$, $\alpha= 2$, versus H, for $1/2 < H< 3/4$, i.e. $m_{opt}(H)$ in (3.8). When $a= 0$, $E_\alpha(H)$ is very small and then $m_{opt}(H)n^{-4/5}$ takes very large values; and $E_\alpha(H) \longrightarrow 0$ as $H \longrightarrow 1/2$ (i.e. $m_{opt}(H) \longrightarrow \infty$ as $H \longrightarrow 1/2$). For other values of a different from zero, $m_{opt}(H)n^{-4/5}$ suffers little variation with respect to H. As (3.8) indicates, for any a, $m_{opt}(H)$ increases quickly when H is closed to 3/4.

FIGURE 1 ABOUT HERE

Figure 2 plots $m_{opt}(H)$ against H for two different sample sizes when $3/4< H< 1$, i.e. $m_{opt}(H)$ in (3.12).

FIGURE 2 ABOUT HERE

In the AR(2) case, for $a_1^2 + 4\, a_2 < 0$ the roots of the characteristic polynomial are complex conjugate, and this may correspond to a finite peak in $h(\lambda)$ at a nonzero frequency, and hence in $f(\lambda)$ at a nonzero frequency. Figure 3 plots the spectral density for different H values for the ARFIMA(2, H-1/2, 0) model. We present two examples where a peak at $\lambda\neq0$ is present. The peaks are not located at the same λ value for different H values. In the short memory case ($H= 1/2$), the peak is located at $\lambda= \pi/4$ if $a_1(a_2-1)/4a_2 = 1/\sqrt{2}$, which can happen if $a_1= 1.172$ and $a_2= -0.707$; and the peak is located at $\pi/6$ if $a_1(a_2-1)/4a_2 = \sqrt{3}/2$, which can happen if $a_1= 1.268$ and $a_2= -0.577$.

FIGURE 3 ABOUT HERE

If m is chosen large enough that the peak fals to the left of λ_m then an estimate of H based on the $I(\lambda_j)$ for $j \le m$ might have a serious negative bias.

5. FEASIBLE APPROXIMATIONS TO THE OPTIMAL BANDWIDTH.

In order to approximate the optimal bandwidth, we need an estimate of H. Robinson(1991 a) has suggested an estimate based on the averaged periodogram, which is consistent even when $L(.)$ is of unknown functional form. He noted that (3.1) implies, for any $q>0$,

$$\frac{F(q\lambda)}{F(\lambda)} \sim q^{2(1-H)} \frac{L(1/q\lambda)}{L(1/\lambda)} \sim q^{2(1-H)} \text{ as } \lambda \longrightarrow 0^+, \tag{5.1}$$

suggesting the H estimate

$$\hat{H}_{mq} = 1- \frac{log\{\hat{F}(q\lambda_m)/\hat{F}(\lambda_m)\}}{2\ log\ q}, \tag{5.2}$$

where $q \in (0,1)$ because $\hat{H}_{mq} = \hat{H}_{m,1/q}$.

In order to illustrate the behaviour of $\hat{H}_{mq}$ evaluated at the optimal bandwidth values derived, we have performed a small Monte Carlo experiment. We have generated data according to an ARFIMA(1, H-1/2, 0) model, i.e.

$$(1- L)^{H-1/2}\ (1- La_1)X_t = \varepsilon_t, \tag{5.3}$$

where $LX_t = X_{t-1}$ and $\varepsilon_t \sim$ iid $N(0,1)$. Figure 4 and 5, presents plots of sample root mean squared errors (RMSE) and biases of $\hat{H}_{mq}$ (q= 1/2), in 5000 replications of model (5.3), with a_1= 0.5, versus m, for various values of H. Figure 4, presents results with n= 400, and Figure 6 with n= 800.

FIGURES 4 AND 5 ABOUT HERE

The m values which minimize the Monte Carlo MSE differ from $m_{opt}(H)$. Even the theoretical MSE of $\hat{H}_{mq}$ will differ from $m_{opt}(H)$, depending, among other things, on q. Table 1 below compares m values minimizing Monte Carlo RMSE of $\hat{H}_{mq}$, $\tilde{m}$ say, and corresponding $m_{opt}(H)$, for different values of H.

TABLE 1

m values minimizing the RMSE of $\hat{H}_{mq}$ in the Monte Carlo, $\tilde{m}$, versus $m_{opt}(H)$ in model (5.11), Monte Carlo RMSE in parenthesis.

H	n = 400		n = 800	
	$\tilde{m}$	$m_{opt}(H)$	$\tilde{m}$	$m_{opt}(H)$
0.6	37 (0.1044)	26 (0.1218)	65 (0.0830)	46 (0.0920)
0.7	41 (0.0809)	32 (0.0885)	67 (0.0650)	56 (0.0692)
0.8	51 (0.0552)	33 (0.0800)	85 (0.0453)	60 (0.0544)
0.9	75 (0.0285)	45 (0.0593)	127 (0.0237)	84 (0.0405)

So, the m's minimizing the RMSE of $\hat{H}_{mq}$ are greater than $m_{opt}(H)$. However, the RMSE of $\hat{H}_{m_{opt}(H)q}$ are fairly close to the minimum achievable RMSE.

Once H has been estimated we need to approximate $E_{\alpha}(H)$, which depends on H and, possibly, the parameters explaining the short memory part of the model. For instance, in ARFIMA models, the formula for $E_2(H)$ is given in (4.5). Given a preliminary value of $h''(0)/2h(0)$, $E_2(H)$ can be estimated according to

$$E_2(\hat{H}_{mq}) = h''(0)/2h(0) + (2\hat{H}_{mq} - 1)/24. \qquad (5.4)$$

Given a pilot value of m, $\hat{m}^{(0)}$ say, $m_{opt}(H)$ and H can be estimated by the following iterative procedure,

$$\hat{H}_q^{(k+1)} = \hat{H}_{\hat{m}^{(k)}q}, \text{ where } \hat{m}^{(k+1)} = m_{opt}(\hat{H}_q^{(k+1)}),\ k = 0,\ 1,\ 2,\ \ldots, \qquad (5.5)$$

where (5.4) in the computation of $m_{opt}(H)$.

Table 2 and 3 below summarize Monte Carlo results for the iterative procedure (5.5), taking $h''(0)/2h(0) = -a_1/(1-a_1)^2$ as known.

TABLE 2

Monte Carlo mean values of $\hat{m}^{(k)}$ in procedure (5.5) based on 5000 replications of model (5.3), with $h''(0)/2h(0)$ known. Starting value $\hat{m}^{(0)} = n^{4/5}$.

	n = 400				n = 800			
H	0.6	0.7	0.8	0.9	0.6	0.7	0.8	0.9
$\hat{m}^{(0)}$	36	41	48	54	63	74	88	105
$\hat{m}^{(1)}$	31	35	38	43	53	63	67	82
$\hat{m}^{(2)}$	31	35	37	41	52	63	66	77
$\hat{m}^{(\infty)}$	30	33	36	40	51	60	64	76
$m_{opt}(H)$	26	32	33	45	46	56	60	84

TABLE 3

Monte Carlo RMSE and BIAS of $\hat{H}_q^{(k)}$ in (5.5) based on 5000 replications of model (5.3). Starting value $\hat{m}^{(0)} = n^{4/5}$.

		n = 400				n = 800			
	H	0.6	0.7	0.8	0.9	0.6	0.7	0.8	0.9
$\hat{H}_q^{(1)}$	RMSE	0.2347	0.1798	0.1175	0.0477	0.2207	0.1702	0.1121	0.0456
	BIAS	-0.232	-0.178	-0.116	-0.045	-0.219	-0.169	-0.111	-0.043
$\hat{H}_q^{(2)}$	RMSE	0.1092	0.0869	0.0614	0.0457	0.0856	0.0700	0.0502	0.0328
	BIAS	-0.049	-0.038	-0.018	0.0175	-0.044	-0.038	-0.024	-0.008
$\hat{H}_q^{(3)}$	RMSE	0.1273	0.0970	0.0734	0.0723	0.0975	0.0745	0.0559	0.0523
	BIAS	-0.034	-0.020	-0.008	0.0406	-0.031	-0.024	-7×10^{-4}	-0.027
$\hat{H}_q^{(\infty)}$	RMSE	0.1287	0.1074	0.0842	0.0809	0.0967	0.0810	0.0595	0.0600
	BIAS	-0.018	-0.009	-0.013	0.0478	-0.019	-0.016	-0.004	-0.035

Convergence is typically achieved after two iterations. The estimate of $m_{opt}(H)$ values are fairly close to the true ones, and the RMSE are also close to the minimum achievable ones. However, it is not automatic since the true value $h''(0)/2h(0)$ is unknown.

It is possible to obtain a more "automatic" m by using an expansion of the semiparametric spectral density

$$f(\lambda) = |1 - e^{i\lambda}|^{1-2H} h(\lambda).$$

Given a pilot m value $\hat{m}^{(0)}$, estimate H by $\hat{H} = \hat{H}_{\hat{m}^{(0)}q}$. Then perform the least squares regression

$$I(\lambda_j) = \sum_{k=0}^{2} Z_{jk}(\hat{H})\, \hat{\beta}_k + \hat{\varepsilon}_j, \quad j=1,\ldots,\hat{m}^{(0)}, \tag{5.6}$$

where $Z_{jk}(H) = |1 - e^{i\lambda_j}|^{1-2H} \lambda_j^k/k!$. $\hat{\beta}_0$ and $\hat{\beta}_2$ are estimates of $h(0)$ and $h''(0)$ respectively. Hence, $h''(0)/2h(0)$ is estimated by $\hat{\beta}_2/2\hat{\beta}_0$. This estimate is plugged in (5.4) in order to implement the iterative procedure (5.5).

Tables 4 and 5 summarize Monte Carlo results for the feasible estimates of $m_{opt}(H)$ and corresponding H estimates based on the algorithm (5.5). The $h''(0)/2h(0)$ estimate is not updated at each iteration. The $m_{opt}(H)$ estimates in Table 4 are more biased than those using the infeasible procedure (Table 2), and the H estimates are more inefficient (compare Tables 5 and 3). However, Monte Carlo results seem sensible enough to us to recomend consideration of the automatic iterative procedure in practice, or at least to warrant further study directed at theoretically justifying and refuting it.

TABLE 4

Monte Carlo mean values of $\hat{m}^{(k)}$ in procedure (5.5) based on 5000 replications of model (5.3), with $h''(0)/2h(0)$ estimated by $\hat{\beta}_2/2\hat{\beta}_0$. Starting value $\hat{m}^{(0)} = n^{4/5}$.

	n = 400				n = 800			
H	0.6	0.7	0.8	0.9	0.6	0.7	0.8	0.9
$\hat{m}^{(0)}$	36	41	48	54	63	74	88	105
$\hat{m}^{(0)}$	35	49	66	85	47	65	105	136
$\hat{m}^{(1)}$	31	41	56	71	39	56	84	121
$\hat{m}^{(\infty)}$	29	43	55	70	39	54	82	110
$m_{opt}(H)$	26	32	33	45	46	56	60	84

TABLE 5

Monte Carlo RMSE and BIAS of $\hat{H}^{(k)}$ in (5.5) based on 5000 replications of model (5.3), with $h''(0)/2h(0)$ estimated by $\hat{\beta}_2/2\hat{\beta}_0$. Starting value $\hat{m}^{(0)} = n^{4/5}$.

		n = 400				n = 800			
	H	0.6	0.7	0.8	0.9	0.6	0.7	0.8	0.9
$\hat{H}_q^{(1)}$	RMSE	0.2347	0.1798	0.1175	0.0477	0.2207	0.1702	0.1121	0.0456
	BIAS	-0.232	-0.178	-0.116	-0.045	-0.219	-0.169	-0.111	-0.043
$\hat{H}_q^{(2)}$	RMSE	0.1520	0.1369	0.1201	0.1185	0.1123	0.1028	0.0946	0.0895
	BIAS	-0.014	-0.013	-0.006	0.0246	-0.005	-0.005	-0.002	0.019
$\hat{H}_q^{(3)}$	RMSE	0.1881	0.1673	0.1444	0.1214	0.1332	0.1180	0.1033	0.0914
	BIAS	-0.019	-0.004	-0.013	0.0356	0.019	0.006	0.012	0.032
$\hat{H}_q^{(\infty)}$	RMSE	0.1957	0.1844	0.1657	0.1560	0.1365	0.1290	0.1136	0.1141
	BIAS	-0.019	-0.021	-0.027	0.0517	0.020	0.019	0.117	0.042

References

Brillinger, D.R. (1975), *Time Series, Data Analysis and Theory.* Holden-Day1111, San Francisco.

Parzen, E. (1957), "On consistent estimates of the spectrum of a stationary time series", *Ann. Math. Statist.*, **28**, 329-348.

Robinson, P.M. (1983), "Review of various approaches to power spectrum estimation" in *Handbook of Statistic, Vol. 3*, edited by D.R. Brillinger and P.R. Krishnaiah, 343-368, Elsevier Science Publishers B.V., New York.

Robinson, P.M. (1991 a), "Semiparametric analysis of long-memory time series", forthcoming in *Ann. Statist.*

Robinson, P.M. (1991 b), "Rates of convergence and optimal spectral bandwidth for long range dependence", revision submitted to *Probab. Theory Rel. Fields.*

Departamento de Estadística y Econometría,
Universidad Carlos III de Madrid,
C./ Madrid 126,
Getafe, Madrid 28903 (Spain).

Department of Economics,
London School of Economics,
Houghton Street,
London WCA 2AE (U.K.).

FIGURE 1

Plots of $m_{opt}(H)\, n^{-4/5}$ for the ARFIMA(1,H−1/2,0) model

$$(1-aL)(1-L)^{H-1/2} X_t = \varepsilon_t.$$

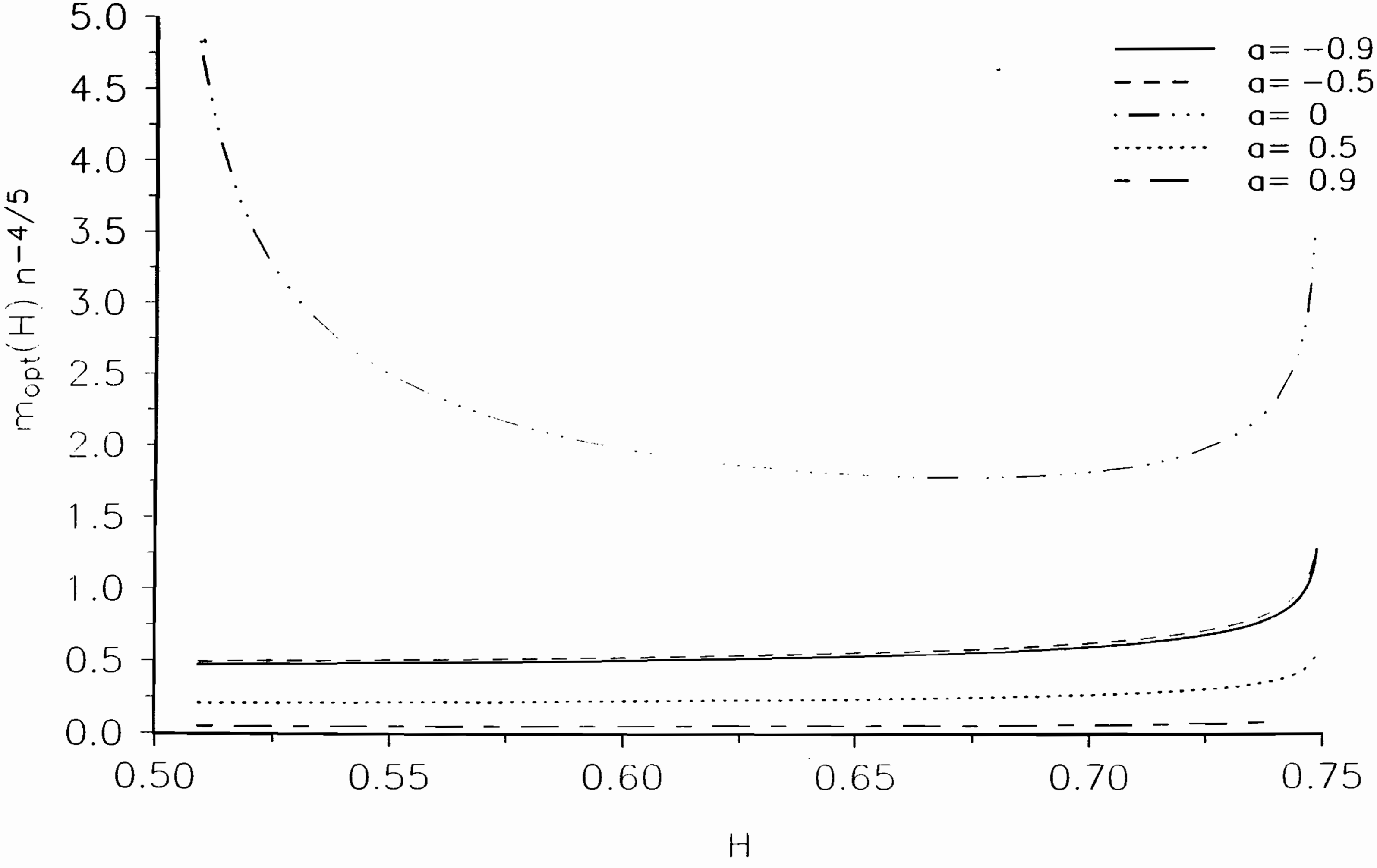

FIGURE 2

Plots of $m_{opt}(H)$ for the ARFIMA(1,H−1/2,0) model

$$(1-L)^{H-1/2}(1-aL)X_t = \varepsilon_t$$

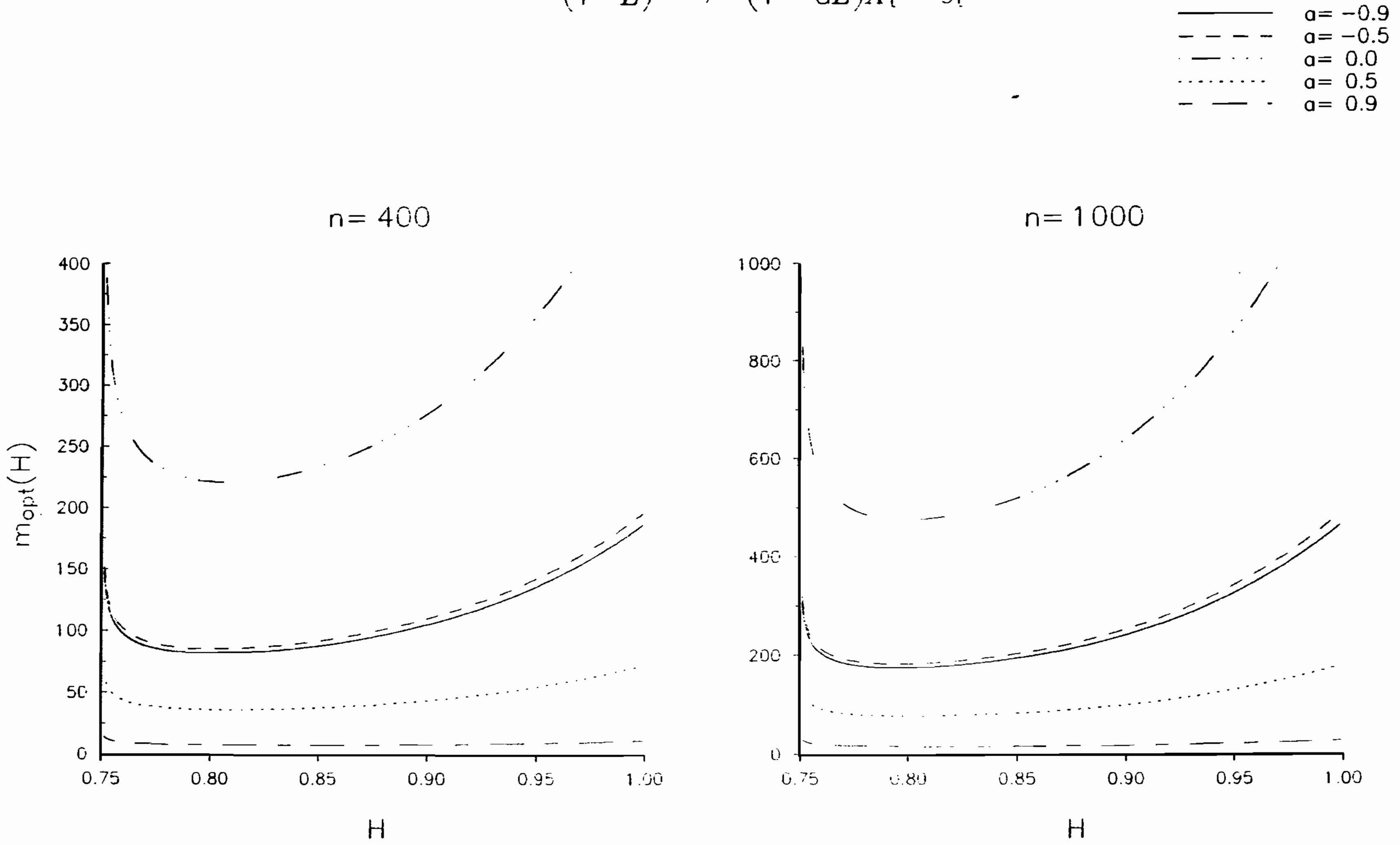

FIGURE 3

Log spectral density of the ARFIMA(2,H−1/2,0) model with a peak at $\lambda \neq 0$

$$(1-L)^{H-1/2}(1-a_1L-a_2L^2)X_t=\varepsilon_t$$

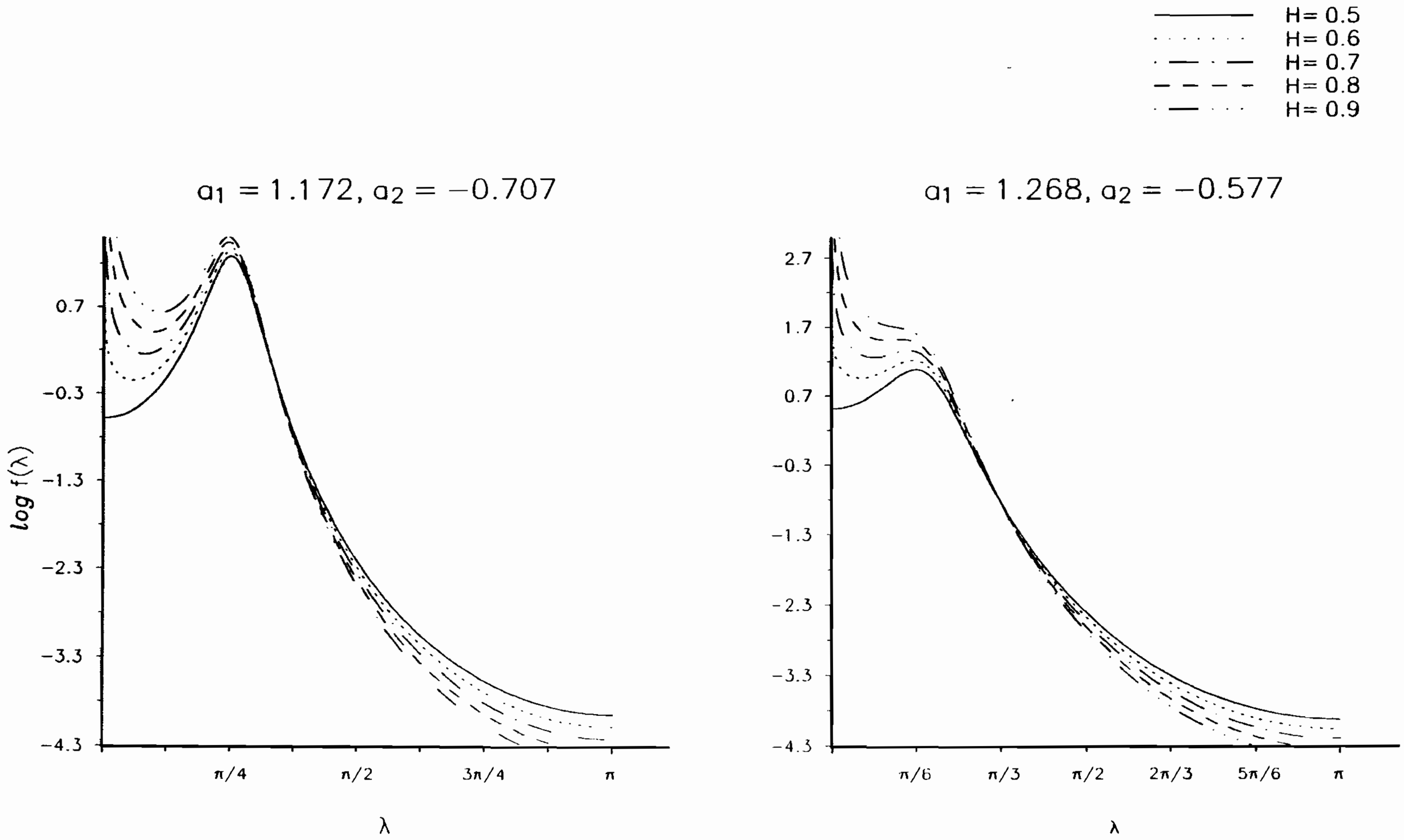

FIGURE 4

Monte Carlo Bias and RMSE of H estimates in an ARFIMA(1/2,H−1/2,0)

Sample size n=400, based on 5000 replications

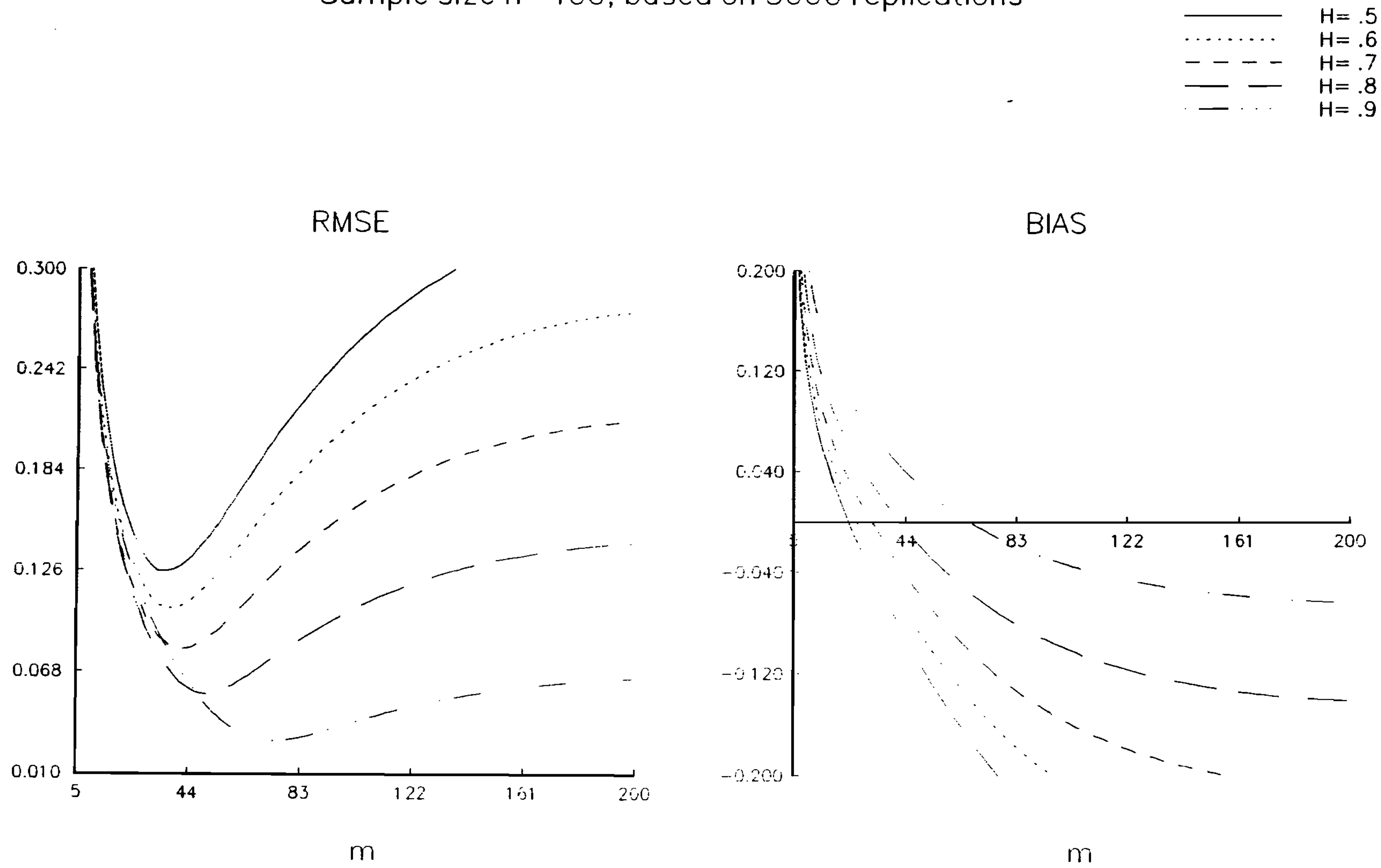

FIGURE 5

Monte Carlo Bias and RMSE of H estimates in an ARFIMA(1/2,H−1/2,0)

Sample size n=800, based on 5000 replications