

# PLUG-IN SELECTION OF THE NUMBER OF FREQUENCIES IN REGRESSION ESTIMATES OF THE MEMORY PARAMETER OF A LONG-MEMORY TIME SERIES

BY CLIFFORD M. HURVICH AND ROHIT S. DEO

*New York University*

*First version received July 1998*

**Abstract.** We consider the problem of selecting the number of frequencies,  $m$ , in a log-periodogram regression estimator of the memory parameter  $d$  of a Gaussian long-memory time series. It is known that under certain conditions the optimal  $m$ , minimizing the mean squared error of the corresponding estimator of  $d$ , is given by  $m^{(\text{opt})} = Cn^{4/5}$ , where  $n$  is the sample size and  $C$  is a constant. In practice,  $C$  would be unknown since it depends on the properties of the spectral density near zero frequency. In this paper, we propose an estimator of  $C$  based again on a log-periodogram regression and derive its consistency. We also derive an asymptotically valid confidence interval for  $d$  when the number of frequencies used in the regression is deterministic and proportional to  $n^{4/5}$ . In this case, squared bias cannot be neglected since it is of the same order as the variance. In a Monte Carlo study, we examine the performance of the *plug-in* estimator of  $d$ , in which  $m$  is obtained by using the estimator of  $C$  in the formula for  $m^{(\text{opt})}$  above. We also study the performance of a bias-corrected version of the plug-in estimator of  $d$ . Comparisons with the choice  $m = n^{1/2}$  frequencies, as originally suggested by Geweke and Porter-Hudak (The estimation and application of long memory time series models. *J. Time Ser. Anal.* 4 (1983), 221–37), are provided.

**Keywords.** Periodogram; bandwidth.

## 1. INTRODUCTION

In this paper, we will consider a Gaussian long-memory time series whose spectral density is of the form

$$f(\lambda) = |1 - \exp(-i\lambda)|^{-2d} f^*(\lambda) \quad \lambda \in [-\pi, \pi] \quad (1)$$

where the memory parameter  $d \in (-1/2, 1/2)$  and the function  $f^*(\lambda)$  govern the long- and short-term correlation structure respectively of the series. A widely used semiparametric estimator of  $d$  is the log-periodogram regression estimator proposed by Geweke and Porter-Hudak (1983). The estimator is obtained by regressing the log periodogram on log frequency for the first  $m$  Fourier frequencies  $\omega_j = 2\pi j/n$ , where  $n$  is the sample size. The choice of  $m$  is crucial in practice, since it determines the bias, variance and mean squared error of the estimator. Based on simulations, Geweke and Porter-Hudak suggested that  $m = n^{1/2}$  be used, and this choice has been widely adopted in the applied literature. It

is now known, however, that this choice is not optimal in general. Under certain assumptions on  $f^*$  and its first two derivatives at zero, Hurvich *et al.* (1998) showed that the optimal value for  $m$  is  $O(n^{4/5})$ , and the mean squared error of the corresponding estimate of  $d$  is  $O(n^{-4/5})$ . In this case, the choice  $m = n^{1/2}$  would yield the suboptimal convergence rate  $O(n^{-1/2})$ . They also determined that the asymptotically optimal value of  $m$  which minimizes the mean squared error is given by

$$m^{(\text{opt})} = Cn^{4/5} \quad (2)$$

where

$$C = \left( \frac{27}{128\pi^2} \right)^{1/5} \left\{ \frac{f^*(0)}{f^{*''}(0)} \right\}^{2/5}. \quad (3)$$

Note that  $C$  depends on the behavior of the unknown function  $f^*$  at zero. In this paper, we will propose and study the properties of an estimate  $\hat{C}$  of  $C$  in (3), based on log-periodogram regression. Using this value of  $\hat{C}$  yields a corresponding selected value of  $\hat{m} = \hat{C}n^{4/5}$ . We will refer to this selection as a *plug-in* choice of  $m$ , since its motivation is derived from the idea of replacing ('plugging in') an estimated value for  $C$  in the asymptotic relationship (2).

Giraitis *et al.* (1997a, 1997b) have also addressed the problem of selecting the number of frequencies in a regression estimator of  $d$ . They introduce another parameter  $\beta$  which governs the smoothness of the function  $f^*$  around zero. Under their assumptions, which are more general than ours, they derive optimality theory for estimating  $d$ . They also propose a data-driven selection of  $m$ , based on an adaptive estimator of  $\beta$ . Their choice of  $m$  is rate optimal in the sense that its ratio to the optimal value of  $m$  converges to a finite constant. Since they make no attempt to estimate this constant, the resulting estimator of  $d$  will be inefficient in general. By contrast, our assumptions imply that  $\beta = 2$ , so that the optimal value of  $m$  is determined up to an unknown constant, which we seek to estimate. It should be noted that  $\beta$  is in fact 2 for most known parametric long-memory models including the ARFIMA (Granger and Joyeux, 1980; Hosking, 1981), fractional exponential and fractional Gaussian noise models (see Beran, 1994).

Another approach to the selection of  $m$  was considered by Henry and Robinson (1996), who also assumed that  $\beta = 2$ . There, the focus was on the Gaussian semiparametric estimator (GSE), originally proposed by Robinson (1995a), in which  $d$  is estimated by maximizing a Whittle-type quasi-likelihood for Fourier frequencies 1 to  $m$ . GSE enjoys at least two advantages over the regression estimator of Geweke and Porter-Hudak (GPH). First, the properties of GSE have been derived without assuming that the time series is Gaussian, while the theory for GPH (Robinson, 1995b, Hurvich *et al.* 1998) does require the Gaussian assumption. Second, in the Gaussian case, the asymptotic efficiency of GPH relative to GSE is only  $6/\pi^2 \approx 0.67$ . In terms of developing a theory for selecting  $m$ , however, GSE seems potentially problematic. Since

GSE is defined implicitly, it seems difficult, if not impossible, to obtain a rigorous expression for the asymptotically optimal  $m$ . Indeed, the expression given by Henry and Robinson (1996) is admittedly based on somewhat heuristic reasoning. In addition, the resulting formula for the optimal  $m$  depends, unfortunately, on  $d$ , which is unknown and which is in fact the quantity on which the ultimate interest is centered. This dependence necessitates the use of iterative methods in which  $d$  and a corresponding optimal value for  $m$  are estimated in alternating cycles.

Delgado and Robinson (1996) derive an expression for the optimal number frequencies  $m$  in the averaged periodogram estimator of  $d$  proposed by Robinson (1994). This estimator of  $d$  has the advantage that its properties have been established without requiring Gaussianity, although the asymptotic distribution is somewhat complicated. Delgado and Robinson provide a plug-in method of estimating the optimal  $m$ , assuming that  $\beta = 2$ . However, the optimal  $m$  depends again on  $d$  and their plug-in procedure is iterative as in the method of Henry and Robinson (1996).

Another existing approach to the selection of  $m$  is frequency-domain cross-validation, as proposed by Hurvich and Beltrao (1994). There, an asymptotic equivalence was established between the mean squared error of an (arbitrarily defined) estimate of  $d$  and the mean integrated squared error of the corresponding ARFIMA(0,  $d$ , 0) spectral estimator, in a neighborhood around zero frequency. Then it was suggested that  $m$  could be selected so as to minimize an estimate of this mean integrated squared error based on frequency-domain cross-validation. Unfortunately, there is as yet no rigorous justification of the optimality of this selected  $m$ , even for GPH in the Gaussian case.

The remainder of this paper is organized as follows. In Section 2 we propose a plug-in selection of  $m$  and study its properties. Section 3 presents a Monte Carlo study of the associated estimator of  $d$  for certain ARFIMA models. Section 4 contains a brief discussion. We conclude with a mathematical appendix.

## 2. THE PLUG-IN METHOD

We assume that we have  $n$  observations  $y_1, y_2, \dots, y_n$  from a stationary Gaussian time series with spectral density given by (1). We further assume that  $f^*(\cdot)$  is an even continuous function on  $[-\pi, \pi]$  with bounded derivatives up to order 3 near the origin. We also assume that  $f^*(0) > 0$ ,  $f^{*'}(0) = 0$  and  $f^{*''}(0) \neq 0$ . The periodogram at the  $j$ th Fourier frequency is given by

$$I_j = \frac{1}{2\pi n} \left| \sum_{t=1}^n y_t \exp(-i\omega_j t) \right|^2.$$

The regression estimator of  $d$  based on the first  $m$  Fourier frequencies is given by

$$\hat{d}_m = -0.5 \frac{\sum_{j=1}^m (a_j - \bar{a}) \log I_j}{\sum_{j=1}^m (a_j - \bar{a})^2} \quad (4)$$

where  $a_j = \log|1 - \exp(-i\omega_j)| = \log|2 \sin(\omega_j/2)|$  and  $\bar{a} = m^{-1} \sum_{j=1}^m a_j$ .

We will now develop an estimator of  $C$  in Equation (3). It is convenient first to estimate the quantity  $K = f^{*''}(0)/f^*(0)$  directly. Note that

$$\log I_j = -2d \log|2 \sin(\omega_j/2)| + \log f_j^* + \log \frac{I_j}{f_j} \quad (5)$$

where  $f_j = f(\omega_j)$  and  $f_j^* = f^*(\omega_j)$ . A Taylor series expansion of  $\log f^*$  at the origin yields

$$\log f_j^* = \log f_0^* + \frac{\omega_j^2}{2} K + \frac{\omega_j^3}{6} R_j \quad (6)$$

where  $R_j$  is uniformly bounded in the neighborhood of the origin. Substituting (6) in (5) gives

$$\log I_j = \log f_0^* - \gamma - 2d \log|2 \sin(\omega_j/2)| + \frac{\omega_j^2}{2} K + \frac{\omega_j^3}{6} R_j + \varepsilon_j \quad (7)$$

where  $\varepsilon_j = \log I_j/f_j + \gamma$  and  $\gamma = 0.577216 \dots$  is Euler's constant. This suggests the estimator  $\hat{K}$  obtained as the third coefficient in an ordinary linear regression of  $\{\log I_j\}$  on the matrix  $X$  with columns  $(1, \log|2 \sin(\omega_j/2)|, \omega_j^2/2)$  for  $j = 1, 2, \dots, L$  where  $L = An^\delta$  for some arbitrary constant  $A$  and  $0 < \delta < 1$ . The use of  $\delta < 1$  ensures that the regression is confined to a neighborhood of zero. We can express  $\hat{K} = \sum_{j=1}^L b_j \log I_j$  where  $\{b_j\}$  is the third row of  $(X'X)^{-1}X'$ . The following lemma gives the bias and variance of  $\hat{K}$ .

LEMMA 1. For  $\delta > 3/4$ ,  $\text{bias}(\hat{K}) = O(L/n)$  and  $\text{var}(\hat{K}) = O(n^4/L^5)$ .

It can be seen from Lemma 1 that to obtain a consistent estimator of  $K$  it is necessary to have  $L \propto n^6$  frequencies with  $\delta > 4/5$ . This lower bound, which will henceforth be assumed, is required to ensure the asymptotic non-singularity of  $X'X$ . Furthermore, in terms of mean squared error the optimal value of  $\delta$  is  $6/7$ , since then  $\text{bias}^2(\hat{K})$  and  $\text{var}(\hat{K})$  are of the same order,  $O(n^{-2/7})$ . As a consequence of the consistency of  $\hat{K}$ , we obtain the following theorem about the consistency of the estimator of  $C$ ,  $\hat{C} = (27/128\pi^2)^{1/5} \hat{K}^{-2/5}$ .

THEOREM 1. If  $C < \infty$ ,  $\hat{C} \xrightarrow{P} C$ .

This estimator of  $C$  can then be used to construct a regression estimator  $\hat{d}_{\hat{m}}$  of  $d$  by using  $\hat{m} = \hat{C}n^{4/5}$ . It follows from the consistency of  $\hat{C}$  that  $\hat{m}/m^{(\text{opt})}$  is consistent for 1.

One goal of practical interest is the construction of confidence intervals for  $d$ , which requires knowledge of the asymptotic distribution of  $\hat{d}_m$ . The usual confidence interval for  $d$  is based on the assumption that  $m^{1/2}(\hat{d}_m - d)$  is

asymptotically normal with zero mean. Unfortunately, this asymptotic mean is not zero when  $m \propto n^{4/5}$ , which is the optimal rate for estimating  $d$ , since then the squared bias and variance of  $\hat{d}_m$  are both proportional to  $n^{-4/5}$ . This will invalidate the traditional confidence intervals for  $d$ . The following theorem suggests a way of correcting this problem.

THEOREM 2. Let  $m = Bn^{4/5}$  for some arbitrary constant  $B$ . Then

$$m^{1/2}(\hat{d}_m - d) \xrightarrow{D} N\left(\frac{-2\pi^2}{9}KB^{5/2}, \frac{\pi^2}{24}\right).$$

The present result follows immediately from Lemma 1 and arguments similar to those in Theorem 2 in Hurvich *et al.* (1998). It is a generalization of Theorem 2 of Hurvich *et al.* (1998), which was valid for  $m = o(n^{4/5})$ . A bias-corrected version of  $\hat{d}_m$  is given by

$$\hat{d}_{m,U} = \hat{d}_m + \frac{2\pi^2}{9} \hat{K} \frac{m^2}{n^2} \quad (8)$$

where  $\hat{K}$  is any consistent estimator of  $K$ . This suggests that  $\hat{d}_{\hat{m},U}$  may be a useful bias-corrected version of the plug-in estimator  $\hat{d}_{\hat{m}}$ .

### 3. MONTE CARLO STUDY

We conducted a simulation study to examine the performance of our proposed methods of estimating  $d$ . We simulated 1000 realizations of an ARFIMA (1, 0.4, 0) process  $(1 - \phi B)(1 - B)^d x_t = \varepsilon_t$ , where the  $\varepsilon_t$  are independent and identically distributed  $N(0, 1)$  and  $B$  is the backshift operator, for each of the sample sizes  $n = 1024, 2048, 4096$ , and autoregressive parameters  $\phi = 0.1, 0.2, \dots, 0.9$ . For each realization, we estimated  $d$  by several methods:

(i) the method of GPH, i.e. the regression estimator  $\hat{d}_m$  given by Equation (4), with  $m = n^{1/2}$ . (GPH is widely used in practice, and hence provides an important benchmark for comparison with other methods);

(ii) the plug-in method  $\hat{d}_{\hat{m}}$ , where  $\hat{m}$  is given by  $\hat{C}n^{4/5}$  and  $\hat{C}$  is defined below Lemma 1;

(iii) the bias-corrected plug-in method  $\hat{d}_{\hat{m},U}$  (see Equation (8)), where  $\hat{m}$  is as in (ii).

For methods (ii) and (iii), we used three different values of  $L$ :  $0.2n^{6/7}$ ,  $0.25n^{6/7}$  and  $0.3n^{6/7}$ . The exponent in  $n^{6/7}$  was used since it minimizes the mean squared error of  $\hat{K}$  as follows from Lemma 1. The multiplicative constants 0.2, 0.25 and 0.3 were chosen on the basis of a small initial simulation study.

Here, we report on a representative subsample of our results, limiting the sample sizes to  $n = 1024, 4096$ , the values of  $\phi$  to 0.1, 0.3, 0.5, 0.7, 0.9, and

the values of  $L$  to  $0.2n^{6/7}$  and  $0.3n^{6/7}$ . Table I gives  $m^{(\text{opt})}$  as in Equations (2) and (3), together with  $m^{(\text{GPH})} = n^{1/2}$  as well as the average values (over the 1000 realizations) of  $\hat{m}$  obtained with  $L = 0.2n^{6/7}$  and  $L = 0.3n^{6/7}$ . As is to be expected,  $m^{(\text{opt})}$  decreases as  $\phi$  increases. The values of  $m^{(\text{GPH})}$ , which remain fixed for given  $n$ , are clearly too small for small  $\phi$ , but eventually become somewhat too large when  $\phi$  reaches 0.9. By contrast, the average values of  $\hat{m}$  do tend to decrease as  $\phi$  increases, although they tend to underestimate  $m^{(\text{opt})}$  when  $\phi$  is small and overestimate  $m^{(\text{opt})}$  when  $\phi$  is large. In addition, the average values of  $\hat{m}$  are always larger for  $L = 0.3n^{6/7}$  than for  $L = 0.2n^{6/7}$ . Thus, the larger value of  $L$  provides better estimates of  $m^{(\text{opt})}$  than the smaller value when  $\phi$  is small but the situation is reversed when  $\phi$  is large.

Tables II and III give average estimates of  $d$  and mean squared errors,

TABLE I  
AVERAGE OF THE SELECTED NUMBER OF FREQUENCIES

$n$	$\phi$	$m^{(\text{opt})}$	$m^{(\text{GPH})} = n^{1/2}$	Plug-in $\hat{m}$ ( $L = 0.2n^{6/7}$ )	Plug-in $\hat{m}$ ( $L = 0.3n^{6/7}$ )
1024	0.1	207	32	74	112
	0.3	109	32	74	108
	0.5	68	32	73	93
	0.7	39	32	65	69
	0.9	14	32	44	58
4096	0.1	629	64	260	374
	0.3	331	64	250	360
	0.5	206	64	245	277
	0.7	119	64	165	170
	0.9	45	64	101	137

TABLE II  
AVERAGE ESTIMATED  $d$  (TRUE VALUE = 0.4)

$n$	$\phi$	GPH	Plug-in		Bias-corrected plug-in	
			$L = 0.2n^{6/7}$	$L = 0.3n^{6/7}$	$L = 0.2n^{6/7}$	$L = 0.3n^{6/7}$
1024	0.1	0.4171	0.4164	0.4164	0.4171	0.4158
	0.3	0.4110	0.4207	0.4298	0.4164	0.4192
	0.5	0.4217	0.4445	0.4583	0.4320	0.4339
	0.7	0.4535	0.4976	0.5175	0.4648	0.4798
	0.9	0.6627	0.7135	0.7941	0.6662	0.7526
4096	0.1	0.4098	0.4077	0.4068	0.4084	0.4052
	0.3	0.4046	0.4139	0.4218	0.4099	0.4145
	0.5	0.4093	0.4281	0.4331	0.4175	0.4157
	0.7	0.4142	0.4450	0.4535	0.4206	0.4290
	0.9	0.4930	0.5686	0.6375	0.5371	0.6104

TABLE III  
MEAN SQUARED ERROR OF ESTIMATED  $d$

$n$	$\phi$	GPH	Plug-in		Bias-corrected plug-in	
			$L = 0.2n^{6/7}$	$L = 0.3n^{6/7}$	$L = 0.2n^{6/7}$	$L = 0.3n^{6/7}$
1024	0.1	0.0169	0.0099	0.0065	0.0151	0.0096
	0.3	0.0173	0.0104	0.0083	0.0150	0.0109
	0.5	0.0186	0.0146	0.0136	0.0188	0.0144
	0.7	0.0222	0.0300	0.0280	0.0296	0.0221
	0.9	0.0858	0.1201	0.1699	0.0946	0.1398
4096	0.1	0.0088	0.0030	0.0020	0.0045	0.0030
	0.3	0.0084	0.0033	0.0027	0.0048	0.0034
	0.5	0.0083	0.0054	0.0048	0.0064	0.0048
	0.7	0.0084	0.0076	0.0066	0.0067	0.0047
	0.9	0.0162	0.0355	0.0615	0.0261	0.0494

respectively, for methods (i), (ii) and (iii), based on averages over the 1000 realizations, for  $n = 1024$  and  $n = 4096$ .

First, we discuss bias. As  $\phi$  increases, all of the estimates of  $d$  become positively biased, due to the strengthening contamination from the short-memory component  $f^*$ . As  $\phi$  increases, GPH generally becomes much less biased than the other methods, due to the fact (see Table I) that GPH typically uses a smaller value of  $m$  than the other methods, and hence is better able to avoid bias due to contamination from higher frequencies. For small values of  $\phi$ , GPH is still somewhat less biased than the other methods, but the differences are not extremely strong. The bias correction for the plug-in estimator seems to be effective in that, for a given value of  $L$ , the bias-corrected estimator is in fact typically less biased than the plug-in estimator, especially when  $\phi$  is large. Comparing the two values of  $L$ , the use of the larger  $L$  strongly increases bias in the plug-in estimators when  $\phi$  is large. This effect can be explained by noting that the estimates of  $K$  become more biased as  $L$  is increased, particularly when  $\phi$  is large.

Next, we discuss mean squared error. For small values of  $\phi$ , the best method in terms of mean squared error is the plug-in, with  $L = 0.3n^{6/7}$ . This method strongly outperforms GPH, by a factor between 2 and 4.5, for  $\phi = 0.1$  and  $\phi = 0.3$ . In fact, except in one case, all other methods considered outperform GPH for  $\phi$  between 0.1 and 0.5. For a given value of  $L$ , the bias-corrected estimator often has a much larger mean squared error than the plug-in estimator when  $\phi$  is small, but the situation is reversed when  $\phi$  is large. This may be the effect of increased variability induced by the bias correction method, which is only offset by reduction in bias when  $\phi$  is large. Comparing the two values of  $L$ , the larger  $L$  yields smaller mean squared errors when  $\phi$  is small, but larger mean squared errors when  $\phi$  is large. This is consistent with the results in Table I and seems sensible, since the larger value of  $L$  gives less

variable but more biased estimates of  $K$ , and this bias becomes more appreciable as  $\phi$  increases. For  $\phi = 0.9$ , the GPH estimator has the smallest mean squared error of all methods considered, by a wide margin. Again, this can be attributed to the superior bias properties of GPH in this case.

In Table IV, we present the coverage rates of the nominal 95% confidence intervals for  $d$ . These intervals are given by  $\hat{d}_{\sqrt{n}} \pm 1.96(\pi^2/24n^{1/2})^{1/2}$  for GPH,  $\hat{d}_{\hat{m}} \pm 1.96(\pi^2/24\hat{m})^{1/2}$  for plug-in and  $\hat{d}_{\hat{m},U} \pm 1.96(\pi^2/24\hat{m})^{1/2}$  for the bias-corrected plug-in estimator. All of the confidence intervals are anti-conservative, particularly so when  $\phi$  is large. Overall, the coverage rates for GPH are closer to nominal than those for the other methods. For a given value of  $L$ , the intervals based on bias-corrected plug-in have better coverage rates than those based on just plug-in when  $\phi$  is close to 0.9. However, the situation is reversed for small values of  $\phi$ . Given the results of Theorem 2, it is disappointing that, for the situations studied here, the bias-corrected plug-in method does not uniformly outperform and never dramatically improves the coverage rate compared to plug-in alone. This is presumably due to variability in  $\hat{K}$  when  $\phi$  is small as well as bias in  $\hat{K}$  when  $\phi$  is large, particularly when  $L$  is large.

In Table V, we present the average widths over the 1000 realizations of the nominal 95% confidence intervals described above. Note that the widths of the GPH-based intervals remain constant for given  $n$ , while the widths for plug-in and bias-corrected plug-in are equal for a given sample,  $n$  and  $L$ . On average, the GPH-based intervals are wider than those based on the other methods, by a factor of as much as 1.77 in the most extreme case. This partly offsets the apparent advantage of GPH in terms of coverage rates. The patterns observed here are consistent with the fact (see Table I) that the plug-in and bias-corrected plug-in methods use, on average, larger values of  $m$  than GPH.

TABLE IV  
COVERAGE RATES (%) FOR NOMINAL 95% CONFIDENCE INTERVALS

$n$	$\phi$	GPH	Plug-in		Bias-corrected plug-in	
			$L = 0.2n^{6/7}$	$L = 0.3n^{6/7}$	$L = 0.2n^{6/7}$	$L = 0.3n^{6/7}$
1024	0.1	89.6	90.4	89.8	82.3	82.9
	0.3	90.3	88.8	82.5	80.5	75.8
	0.5	89.3	78.7	71.9	72.8	73.8
	0.7	86.3	64.2	59.1	68.4	68.9
	0.9	38.1	24.5	4.8	34.6	8.6
4096	0.1	89.7	89.4	89.8	81.2	82.5
	0.3	91.3	85.3	80.1	78.3	76.1
	0.5	92.9	77.6	75.1	74.2	79.0
	0.7	92.9	73.7	72.5	81.8	81.4
	0.9	75.5	30.4	4.6	43.9	8.8



TABLE V  
AVERAGE WIDTH OF NOMINAL 95% CONFIDENCE INTERVALS

$n$	$\phi$	GPH	Plug-in		Bias-corrected plug-in	
			$L = 0.2n^{6/7}$	$L = 0.3n^{6/7}$	$L = 0.2n^{6/7}$	$L = 0.3n^{6/7}$
1024	0.1	0.4444	0.3120	0.2517	0.3120	0.2517
	0.3	0.4444	0.3109	0.2559	0.3109	0.2559
	0.5	0.4444	0.3134	0.2750	0.3134	0.2750
	0.7	0.4444	0.3373	0.3126	0.3373	0.3126
	0.9	0.4444	0.3910	0.3350	0.3910	0.3350
4096	0.1	0.3142	0.1694	0.1391	0.1694	0.1391
	0.3	0.3142	0.1710	0.1414	0.1710	0.1414
	0.5	0.3142	0.1740	0.1596	0.1740	0.1596
	0.7	0.3142	0.2038	0.1937	0.2038	0.1937
	0.9	0.3142	0.2502	0.2148	0.2502	0.2148

#### 4. DISCUSSION

Overall, our results indicate that plug-in selection of  $m$  may be advantageous when  $f^*$  is not too peaked near zero frequency. Thus, these methods may be useful for econometric data, where one generally works with the first differences which typically exhibit only weak short-term correlation. Inevitably, in small enough samples, a simple deterministic choice of  $m$ , say  $m = n^{1/2}$ , as advocated in GPH can yield estimates of  $d$  which outperform those based on a plug-in choice of  $m$ . However, asymptotically the performance of GPH will be inferior to that of the plug-in estimator.

#### APPENDIX

PROOF OF LEMMA 1. We will assume throughout the proof that all summations are over  $j = 1, 2, \dots, L$ . Let  $X_{j2} = \log|2 \sin(\omega_j/2)|$  and  $X_{j3} = 0.5\omega_j^2$ . Then it follows from standard regression theory that  $\hat{K} = (\sum r_j^2)^{-1} \sum r_j \log I_j$ , where  $r_j$  is the  $j$ th residual in the regression of  $\{X_{j3}\}$  on  $\{(1, X_{j2})\}$ . More specifically,

$$r_j = X_{j3} - \bar{X}_3 - \hat{\nu}(X_{j2} - \bar{X}_2) \quad (9)$$

where

$$\hat{\nu} = \frac{\sum (X_{j2} - \bar{X}_2) X_{j3}}{\sum (X_{j2} - \bar{X}_2)^2} \quad (10)$$

and  $\bar{X}_k = L^{-1} \sum_j X_{jk}$  for  $k = 2, 3$ . From Hurvich *et al.* (1998), it is known that

$$X_{j2} - \bar{X}_2 = \log \frac{j}{L} + 1 + o(1)$$

and

$$\sum (X_{j2} - \bar{X}_2)^2 = L + o(L). \quad (11)$$

Using the fact that for any integer  $s \geq 0$

$$\sum j^s \log \frac{j}{L} = -\frac{L^{s+1}}{(s+1)^2} + o(L^{s+1}) \quad (12)$$

we obtain

$$\begin{aligned} \sum (X_{j2} - \bar{X}_2) X_{j3} &= \frac{2\pi^2}{n^2} \sum j^2 \left\{ \log \frac{j}{L} + 1 + o(1) \right\} \\ &= \frac{4\pi^2}{9} \frac{L^3}{n^2} + o\left(\frac{L^3}{n^2}\right). \end{aligned} \quad (13)$$

Using Equations (10), (11) and (13) in (9), we get

$$r_j = \frac{\pi^2 L^2}{n^2} \left( 2 \frac{j^2}{L^2} - \frac{10}{9} - \frac{4}{9} \log \frac{j}{L} + o(1) \log \frac{j}{L} \right). \quad (14)$$

Applying Equation (12) together with the fact that  $\sum \{\log(j/L)\}^2 = 2L + o(L)$ , we get

$$\sum r_j^2 = \frac{64}{405} \frac{\pi^4 L^5}{n^4} + o\left(\frac{L^5}{n^4}\right).$$

Letting  $b_j = (\sum r_j^2)^{-1} r_j$ , it follows from Equation (7) that

$$\text{bias}(\hat{K}) = \sum \frac{b_j \omega_j^3 R_j}{6} + \sum b_j E(\varepsilon_j). \quad (15)$$

Using the uniform boundedness of  $R_j$ , Equation (12) and the fact that  $|\sum \omega_j^3 \log(jL^{-1})| = -\sum \omega_j^3 \log(jL^{-1})$ , some simple manipulation yields

$$\sum b_j \omega_j^3 R_j = O(Ln^{-1}).$$

To handle the second term on the right-hand side of Equation (15), we use Lemmas 5 and 6 of Hurvich *et al.* (1998) to conclude that for  $\delta > 3/4$

$$\begin{aligned} \left| \sum b_j E(\varepsilon_j) \right| &\leq A \sum_{j=1}^{\log^2 L} |b_j| + A \sum_{j=\log^2 L+1}^L |b_j| \frac{\log j}{j} \\ &= O(Ln^{-1}) \end{aligned}$$

where  $A$  is some finite positive constant. Hence, we have

$$\text{bias}(\hat{K}) = O(Ln^{-1}).$$

Using Lemmas 2, 3 and 5 of Hurvich *et al.* (1998), an argument similar to the one presented in the proof of their Theorem 1 also gives

$$\begin{aligned} \text{var}(\hat{K}) &= \frac{\pi^2}{6} \frac{1}{\sum r_j^2} + o(n^4 L^{-5}) \\ &= \frac{135n^4}{128\pi^2 L^5} + o(n^4 L^{-5}). \end{aligned}$$

## REFERENCES

- BERAN, J. (1994) *Statistics for Long-memory Processes*. New York: Chapman and Hall.
- DELGADO, M. and ROBINSON, P. M. (1996) Optimal spectral bandwidth for long memory. *Stat. Sin.* 6, 97–112.
- GEWEKE, J. and PORTER-HUDAK, S. (1983) The estimation and application of long memory time series models. *J. Time Ser. Anal.* 4, 221–37.
- GIRAITIS, L., ROBINSON, P. M. and SAMAROV, A. (1997a) Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long-range dependence. *J. Time Ser. Anal.* 18, 49–60.
- , — and — (1997b). Adaptive rate-optimal semiparametric estimation of the long memory parameter. Preprint.
- GRANGER, C. W. J. and JOYEUX, R. (1980) An introduction to long memory time series models and fractional differencing. *J. Time Ser. Anal.* 1, 15–29.
- HENRY, M. and ROBINSON, P. M. (1996) Bandwidth choice in Gaussian semiparametric estimation of long range dependence. *Proc. Athens Conference on Applied Probability and Time Series* (eds P. M. Robinson and M. Rosenblatt), pp. 220–32. Taipei: Academia Sinica.
- HOSKING, J. R. M. (1981) Fractional differencing. *Biometrika* 68, 165–76.
- HURVICH, C. M. and BELTRAO, K. I. (1994) Automatic semiparametric estimation of the memory parameter of a long memory time series. *J. Time Ser. Anal.* 15, 285–302.
- , DEO, R. S. and BRODSKY, J. (1998) The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series. *J. Time Ser. Anal.* 19, 19–46.
- ROBINSON, P. M. (1994) Semiparametric analysis of long-memory time series. *Ann. Stat.* 22, 515–39.
- (1995a) Gaussian semiparametric estimation of long range dependence. *Ann. Stat.* 23, 1630–61.
- (1995b) Log-periodogram regression of time series with long range dependence. *Ann. Stat.* 23, 1048–72.