

# DOCUMENTOS DE TRABAJO

# BILTOKI

D.T. 2008.01

Selection of the number of frequencies using bootstrap techniques  
in log-periodogram regression.

Josu Arteche y Jesus Orbe

eman ta zabal zazu



Universidad Euskal Herriko  
del País Vasco Unibertsitatea

Facultad de Ciencias Económicas.  
Avda. Lehendakari Aguirre, 83  
48015 BILBAO.

**Documento de Trabajo BILTOKI DT2008.01**

Editado por el Departamento de Economía Aplicada III (Econometría y Estadística)  
de la Universidad del País Vasco.

Depósito Legal No.: BI-576-08

ISSN: 1134-8984

URN: RePEc:ehu:biltok:200801

# Selection of the number of frequencies using bootstrap techniques in log-periodogram regression

Josu Arteche and Jesus Orbe \*  
University of the Basque Country, Bilbao, Spain

12th February 2008

## Abstract

The choice of the bandwidth in the local log-periodogram regression is of crucial importance for estimation of the memory parameter of a long memory time series. Different choices may give rise to completely different estimates, which may lead to contradictory conclusions, for example about the stationarity of the series. We propose here a data driven bandwidth selection strategy that is based on minimizing a bootstrap approximation of the mean squared error and compare its performance with other existing techniques for optimal bandwidth selection in a mean squared error sense, revealing its better performance in a wider class of models. The empirical applicability of the proposed strategy is shown with two examples: the widely analyzed in a long memory context Nile river annual minimum levels and the input gas rate series of Box and Jenkins.

*Keywords:* Bootstrap, long memory, log-periodogram regression, bandwidth selection.

## 1 Introduction

Over the last years, log periodogram regression has become one of the most popular tools for the estimation of the memory parameter in long memory time series. It has been widely applied for statistical inference in empirical research due to its simple implementation, pivotal asymptotic normality and robustness thanks to its semiparametric or local condition. The log periodogram regression estimation (LPE hereafter) is based on a simple least squares regression of the logarithm of the periodogram over the logarithm of the  $m$  Fourier frequencies closest to the origin, providing that  $m$  goes to infinity but more slowly than the sample size such that the band of frequencies used in the estimation shrinks to zero. The parameter  $m$ , known as the bandwidth in a local or semiparametric memory parameter estimation context, plays an important role on the performance of the LPE. A large  $m$  reduces the variance at

---

\**Address for correspondence:* Josu Arteche, Dpt. Econometría y Estadística, University of the Basque Country, Avda. Lehendakari Agirre 83, 48015 Bilbao, Spain. Tl: 34 94 601 3852. Fax: 34 94 601 3754. Email: josu.artech@ehu.es

the cost of a higher bias, which in some situations can render meaningless estimates, as for example in the presence of a significant short memory component. On the contrary a low  $m$  guarantees a small bias but with larger variability. From an empirical perspective, the estimates of the memory parameter usually vary significantly with the choice of  $m$ . Figures 9b) and 10b) show the LPE estimates for the Nile river annual minimum series and the input gas rate series of Box and Jenkins (1976, series J), both analyzed in detail in Section 4, for a grid of bandwidths. Depending on the bandwidth choice we can get different conclusions about the persistence of the series or even about their stationarity since long memory series are only stationary for a memory parameter lower than 0.5.

Optimal bandwidth selection techniques have been proposed as a way to balance bias and variance in a compromise to minimize some approximation of the mean squared error (MSE). Hurvich and Deo (1999) introduced a plug in version of the theoretical bandwidth that minimizes an asymptotic approximation of the MSE whereas Giraitis et al. (2000) suggest an adaptive LPE that adapts to an unknown local to zero spectral smoothness. While the former is only valid for a particular local to zero spectral smoothness, the latter does not give unique choices of the bandwidth and only offers bandwidths with an optimal growth rate that can be arbitrarily changed by a multiplicative constant without affecting its rate of increase. To our knowledge, there is not any other formal procedure for the choice of an optimal bandwidth, despite the great dependence of the estimates on the bandwidth choice. We propose here a nonparametric and fully data driven bandwidth selection strategy based on choosing the bandwidth that minimizes a bootstrap mean squared error. This strategy is justified by the likeness of the bootstrap MSE to the Monte Carlo MSE so that it can be safely used for bandwidth selection.

We consider long memory series  $x_t$  with a spectral density satisfying

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

where the memory parameter  $d \in (-0.5, 0.5)$  guarantees stationarity and invertibility. Usually the function  $g(\lambda)$  controls the weak dependence and is assumed to be positive and bounded over all the frequencies such that

$$g(\lambda) = g(0) + \Delta(\lambda), \quad |\Delta(\lambda)| \leq C_1|\lambda|^\alpha \quad (2)$$

for some constant  $C_1$  and a positive local spectral smoothness parameter  $\alpha$ . For  $\alpha \leq 1$  condition (2) holds if  $g(\lambda)$  satisfies a Lipschitz condition of degree  $\alpha$  and for  $1 < \alpha$  if  $g(\lambda)$  is

$[\alpha]$  times differentiable around zero with zero derivatives at  $\lambda = 0$  and the  $[\alpha]$ -th derivative satisfies a Lipschitz condition of degree  $\alpha - [\alpha]$  around zero. Fractional ARIMA processes fall on this category with  $\alpha = 2$ .

The LPE of the memory parameter  $d$  is based on the periodogram of the series  $x_t$ ,  $t = 1, \dots, n$ , defined as

$$I(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^n x_t \exp(-it\lambda) \right|^2.$$

Taking logs in (1) and considering only Fourier frequencies of the type  $\lambda_j = 2\pi j/n$  we have the following linear regression model

$$y_j = c + dz_j + u_j, \quad j = 1, 2, \dots, m, \quad (3)$$

where  $y_j = \log I_j$ ,  $z_j = -2 \log \lambda_j$ ,  $I_j = I(\lambda_j)$ ,  $c = \log g_0 + Ev_j$ ,  $u_j = v_j + \varepsilon_j - Ev_j$ ,  $\varepsilon_j = \log(g_j/g_0)$  and  $v_j = \log(I_j/f_j)$  for  $f_j$  and  $g_j$  defined similarly to  $I_j$ . The LPE,  $\hat{d}(m)$ , is obtained by ordinary least squares and  $\varepsilon_j$  represents the error committed by assuming that the function  $g(\cdot)$  is constant in the interval  $[0, \lambda_m]$ . This is the main source of the bias of the LPE, which depends on the smoothness of the function  $g(\cdot)$  around frequency zero, that is on the deviation of this function from a constant. Usually  $g(\cdot)$  is an even function with bounded second derivative around zero such that  $g(\lambda) = g(0) + O(\lambda^2)$  (e.g. in ARFIMA models). In this case the bandwidth has to satisfy  $m^{-1} + n^4 m^{-5} \rightarrow 0$  as  $n \rightarrow \infty$  for the bias to be negligible with respect to the variance and the LPE to have the asymptotic distribution

$$\sqrt{m}(\hat{d}(m) - d) \xrightarrow{d} N\left(0, \frac{\pi^2}{24}\right) \quad \text{as } n \rightarrow \infty. \quad (4)$$

Balance of variance and bias is achieved with a choice of  $m = Cn^{4/5}$ , for a positive constant  $C$ , which is the optimal choice in a MSE sense. The practical application of this optimal bandwidth is however not feasible since the constant  $C$  is unknown. Hurvich and Deo (1999) proposed a plug-in version of the optimal bandwidth based on a prior estimation of  $C$ . However, even though  $\alpha = 2$  in the more common parametric long memory models (but not in other cyclical long memory context), in practice the spectral smoothness is unknown and the optimal bandwidth is generally of order  $O(n^{2\alpha/(1+2\alpha)})$ . Giraitis et al. (2000) -see also the version of Moulines and Soulier (2003)- introduced an adaptive LPE that adapts the choice of the bandwidth to an unknown  $\alpha$  such that it selects a bandwidth with the optimal growth rate  $O(n^{2\alpha/(1+2\alpha)})$  for an unknown  $\alpha$ . We propose here a bootstrap based bandwidth selection strategy that does not require estimation of any nuisance parameter

nor knowledge of the spectral smoothness of  $g$ , and is fully data driven. We compare its performance with the plug-in bandwidth and the adaptive LPE which are to our knowledge the only two rigorous bandwidth selection strategies proposed to date.

Section 2 describes our proposal together with the plug-in bandwidth selection of Hurvich and Deo (1999) and the adaptive log periodogram version of Moulines and Soulier (2003). The performance of the three strategies is analyzed in a Monte Carlo in Section 3. Section 4 shows two empirical applications of our bandwidth selection procedure. Finally Section 5 concludes and suggests further extensions.

## 2 Bandwidth selection

### 2.1 Plug-in bandwidth selection

Hurvich et al. (1998) determine the optimal value of  $m$  that minimizes an asymptotic approximation of the mean squared error when the function  $g(\lambda)$  is an even continuous function on  $[-\pi, \pi]$  with bounded derivatives up to order three near the origin such that the smoothness of  $g(\lambda)$  corresponds to  $\alpha = 2$ . It has the form

$$m^{(opt)} = Cn^{4/5}, \quad \text{for } C = \left( \frac{27}{128\pi^2} \right)^{1/5} \left\{ \frac{g(0)}{g''(0)} \right\}^{2/5} \quad (5)$$

where  $g''(0)$  is the second derivative of the function  $g$  at the origin. Since this function is not specified  $m^{(opt)}$  is not feasible for practical purposes. A plug-in version based on an estimate of  $C$  was proposed by Hurvich and Deo (1999),  $\hat{m}_{hd} = \hat{C}n^{4/5}$  where the constant  $C$  is consistently estimated by

$$\hat{C} = \left( \frac{27}{128\pi^2} \right)^{1/5} \hat{K}^{-2/5}$$

where  $\hat{K}$  is obtained as the third coefficient in an ordinary linear regression of  $\log I_j$  on  $(1, \log \lambda_j, \lambda_j^2/2)$  for  $j = 1, 2, \dots, An^\delta$ , for  $4/5 < \delta < 1$  and  $A$  an arbitrary constant.

### 2.2 Adaptive estimation

The plug-in bandwidth choice of Hurvich and Deo (1999) is designed for long memory series where  $g(\cdot)$  is an even function with bounded second derivative around the origin such that  $\alpha = 2$ . The local spectral smoothness determines the exponent of the optimal bandwidth such that, for a general  $\alpha$ ,  $m^{(opt)} = C_2 n^{2\alpha/(2\alpha+1)}$  for some positive constant  $C_2$ . However, in practice  $\alpha$  is not known and some estimation is thus required for plug-in optimal bandwidth selection. Giraitis et al. (2000) proposed instead an adaptive version of the LPE that

adapts the bandwidth choice to the unknown local spectral smoothness around zero but does not estimate the multiplicative constant  $C_2$ . Thus, as they already pointed out, this is not strictly speaking an optimal bandwidth selection criterion but a confirmation of the existence of an estimator that achieves an optimal (up to a logarithmic term) adaptive rate of convergence in a minimax sense. We consider here the version of Moulines and Soulier (2003) that we found empirically more attractive because compares estimates based on arbitrary values of  $m$  and does not depend on prior bounds of  $\alpha$ . The adaptive estimator of the bandwidth,  $\hat{m}_{ad}(\kappa)$  is obtained as the largest integer  $m$  such that

$$\forall m' < m, \quad |\hat{d}(m') - \hat{d}(m)| < \kappa \sqrt{\log n} \frac{\pi^2}{24m'} \quad (6)$$

for some positive constant  $\kappa$ .

### 2.3 Local Bootstrap bandwidth selection

We propose a fully data driven bandwidth selection strategy based on a bootstrap approximation of the MSE that does not require estimation either of the local spectral smoothness nor of any multiplicative constant depending on unknown nuisance parameters. The optimal bandwidth is selected as the bandwidth that minimizes a bootstrap MSE, which we use as an approximation of the finite sample MSE. As in Arteche and Orbe (2005), the bootstrap is applied to the residuals in the regression (3) based on the pivotal character of the asymptotic distribution of  $\hat{d}(m)$ . Since the errors are a function of the Studentized periodogram ordinates  $I_j/f_j$  this strategy is related to other frequency domain bootstraps that use this ratio in different contexts (see Dahlhaus and Janas, 1996 or Franke and Härdle, 1992).

A blind bootstrap however should not be applied since the residuals are not i.i.d. but show some structure due to the  $\varepsilon_j$  term. Consider for example an *ARFIMA*(1,  $d$ , 0) defined as  $(1 - \phi L)(1 - L)^d x_t = \xi_t$  for  $\xi_t \sim iid(0, 1)$ . The spectral density of  $x_t$  satisfies (1) with

$$g(\lambda) = \frac{1}{2\pi} \frac{[2\lambda^{-1} \sin(\lambda/2)]^{-2d}}{1 + \phi^2 - 2\phi \cos \lambda} = \frac{1}{2\pi(1 - \phi)^2} \left[ 1 + \left( \frac{d}{12} - \frac{\phi}{(1 - \phi)^2} \right) \lambda^2 + O(\lambda^4) \right]$$

as  $\lambda \rightarrow 0$ , such that

$$\varepsilon_j = \log \left( \frac{g_j}{g_0} \right) = \left( \frac{d}{12} - \frac{\phi}{(1 - \phi)^2} \right) \lambda_j^2 + O(\lambda_j^4).$$

Then  $\varepsilon_j$ , and consequently also the errors in the LPE regression, increase for frequencies far from the origin, and this enlargement can be quite significant if  $\phi$  approaches 1, when the negative  $\varepsilon_j$  causes a positive bias on the LPE that is asymptotically negligible with

an appropriate bandwidth choice, but can be significantly large in finite samples if a large bandwidth is used.

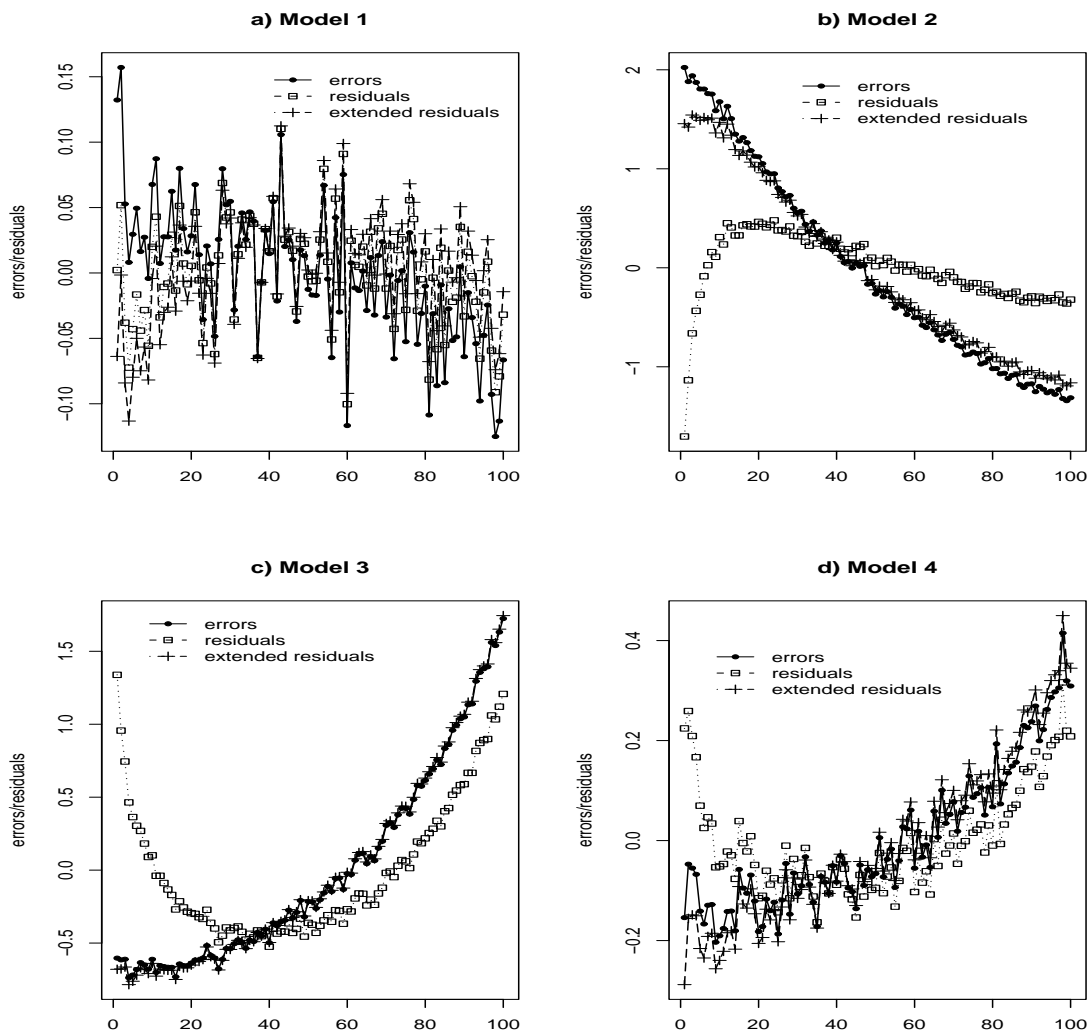
Figure 1 shows the centered errors  $u_j - \bar{u}_j$  (solid-- line) for  $j = 1, \dots, 100$  of the four models analyzed in the Monte Carlo below for a sample size  $n = 512$ . Instead of showing the results of one single simulation, which can be affected for any kind of randomness, we show the average errors obtained with 1000 simulations. The distinct structure that turns up in some cases renders the bootstrap techniques based on i.i.d. errors rather inappropriate for application here. Consider for example Figure 1b) that shows the errors in an *ARFIMA*(1,0.4,0) with  $\phi = 0.8$ . The significant short memory component, ignored in the LPE regression, gives rise to a marked structure of the errors such that the bootstrap procedure should preserve that structure over the bootstrap samples. For that purpose we propose to use the local bootstrap suggested by Paparoditis and Politis (1999) and applied for the estimation of the memory parameter in ARFIMA models by Silva et al. (2006), which maintains the global structure of the bootstrapped series by resampling in a neighborhood of each frequency. Our proposal differs from theirs in that we resample residuals instead of the periodogram of the series. The local procedure applied directly to the periodogram is based on the property that for a smooth spectral density the distribution of adjacent periodogram ordinates is very similar and they are independent (at least asymptotically). This is not the case in long memory series. However the regression errors are functions of the Studentized periodogram ordinates, which show a more stable behaviour such that bootstrapping these quantities gives more reliable results than a direct bootstrap applied to the periodogram or its logarithm in the regression model (3). See Dahlhaus and Janas (1996) and Franke and Härdle (1992) to that respect in a weak dependent context.

But our proposal requires a prior estimation of the memory parameter since the regression bootstraps are based on residuals rather than errors. We then first need residuals whose behaviour approach that of the errors. The dotted-□- line in Figure 1 shows the average (over the 1000 simulations) residuals obtained with a bandwidth  $m = 100$ . They do not approximate at any extent the behaviour of the true errors, represented by a continuous-- line, in three out of four of the models considered, and any bootstrap based on them would be misleading. A bandwidth  $m = 100$  is too large and raises a positive bias which transmits to the residuals. The dashed-+- line in Figure 1 displays the extended centered residuals over the first 100 Fourier frequencies but obtained with LPE estimates based on a bandwidth  $m = 10$ . We clearly avoid the large bias with this low bandwidth and get residuals



whose behaviour largely resembles that of the true errors.

Figure 1: LPE errors and residuals



With all these considerations we propose the following local bootstrap bandwidth selection procedure:

1. Estimate the model (3) for a prior bandwidth  $m_1$ . Obtain  $\hat{c}_1$  and  $\hat{d}_1 = \hat{d}(m_1)$ .
2. For  $m_1 < m_2 < \lfloor n/2 \rfloor$  calculate the extended residuals  $\hat{u}_j = y_j - \hat{c}_1 - \hat{d}_1 z_j$ ,  $j = 1, 2, \dots, m_2$ .
3. Select a resampling width  $k_n \in \mathcal{N}$  and  $k_n < \lfloor m_2/2 \rfloor$ .
4. Define i.i.d. discrete random variables  $S_1, \dots, S_{m_2}$  taking values in the set  $\{0, \pm 1, \dots, \pm k_n\}$  with equal probability  $1/(2k_n + 1)$ .

5. Generate  $B$  bootstrap series  $\hat{u}_{bj}^* = \hat{u}_{|j+S_j|}$  if  $|j + S_j| > 0$ ,  $\hat{u}_{bj}^* = \hat{u}_1$  if  $j + S_j = 0$  for  $b = 1, 2, \dots, B$ ,  $j = 1, \dots, m_2$ .
6. Generate  $B$  bootstrap samples  $y_{bj}^* = \hat{c}_1 + \hat{d}_1 z_j + \hat{u}_{bj}^*$  for  $b = 1, 2, \dots, B$ ,  $j = 1, \dots, m_2$ .
7. Estimate  $d$  for the  $B$  bootstrap samples and different bandwidths  $m \in [\underline{m}, m_2]$  such that  $\underline{m} < m_1$  and calculate the bootstrap MSE

$$MSE^*(m) = \frac{1}{B} \sum_{b=1}^B (\hat{d}(m) - \hat{d}_1)^2$$

8. Chose  $\hat{m}_1^*$  such that  $MSE^*(\hat{m}_1^*) \leq MSE^*(m)$  for all  $m \in [\underline{m}, m_2]$ .
9. With  $\hat{m}_1^*$  instead of  $m_1$  repeat the procedure from step 1 until

$$\frac{MSE^*(\hat{m}_i^*) - MSE^*(\hat{m}_{i-1}^*)}{MSE^*(\hat{m}_{i-1}^*)} < \delta$$

for some small (in absolute value)  $\delta < 0$  stopping criterion.

*Remark 1:* We only consider residuals for positive  $j + S_j$  in step 5 because the regression is only defined for positive frequencies. Also symmetry of the periodogram and spectral density implies symmetry of the residuals if negative frequencies were included. However the frequency zero can not be included because the spectral density is infinity at the origin and the relation between periodogram and spectral density that motivates the LPE does not hold. This implies a double probability of appearance of  $\hat{u}_1$  in some bootstrap samples, but this situation is not very frequent and we believe that its effect is negligible.

*Remark 2:* In other residual based bootstraps a prior centering and heteroscedasticity correction is usually carried out by dividing the centered residuals by  $\sqrt{1 - h_j}$  for  $h_j$  the elements in the diagonal of the matrix  $I - X(X'X)^{-1}X'$  where  $X$  is a matrix with typical  $k$ -th row  $[X]_k = (1, z_k)$ . In our extended residuals setup the precise correction should imply dividing the centered extended residuals in step 2 by the square root of the elements in the diagonal of the matrix  $MM'$  for  $M = I - X(X'_1X_1)^{-1}X'_1D$ , where  $X$ , and  $X_1$  are  $m_2 \times 2$  and  $m_1 \times 2$  matrices defined as  $X$  above,  $I$  is the identity matrix and  $D$  is a  $m_1 \times m_2$  matrix of zeros except ones in the  $m_1$  first diagonal elements. However, due to the local nature of our resample strategy, we found this correction unnecessary (simulations not reported and available upon request show that this correction does not affect the results obtained hereafter).

*Remark 3:* The procedure starts with a user chosen bandwidth  $m_1$ . We found in the simulations that a low value of  $m_1$  is more adequate since we reduce in that way the probability of a highly biased first estimate and the iterative bootstrap based procedure has a faster convergence. For all the simulations in the next section the process converged in less than 8 iterations.

*Remark 4:* We could have used other set of probabilities in step 4 but, as noted by Paparoditis and Politis (1999), the choice of the probability scheme is not very relevant (similar to the choice of the kernel in a nonparametric density estimation).

*Remark 5:* However the choice of the resampling width  $k_n$  is more important. Silva et al. (2006) suggest the use of a very low value  $k_n = 1$  or  $2$  in their local periodogram bootstrap. However we have found that a larger  $k_n$  gives better results in many situations due to the higher stability of the residuals. We show below that the choice of the resampling width is important up to a certain extent. A blind selection deteriorates significantly the results but we can safely use different widths in a sensible region without affecting significantly the performance of our strategy for the choice of the bandwidth. Simple data driven criteria for resampling width selection are also described in next sections.

### 3 Monte Carlo

We generate 1000 replications of Gaussian series of length 512 satisfying equation (1) with  $d = 0.4$ . We think these values are quite representative for the time series where the LPE is usually applied. We have also consider other sample sizes and values of  $d$  in the stationary region and similar conclusions apply. We consider four different type of models:

- Model 1:  $(1 - 0.1L)(1 - L)^d x_t = u_t$
- Model 2:  $(1 - 0.8L)(1 - L)^d x_t = u_t$
- Model 3:  $(1 - 0.1 + 0.9L)(1 - L)^d x_t = u_t$
- Model 4:  $(1 - L)^d(1 + L^2)^{0.2} x_t = u_t$

for  $u_t$  a standard normal variable. The plug-in method of Hurvich and Deo (1999) is designed for processes with  $\alpha = 2$  such as Models 1, 2 and 3. The first model is an  $ARFIMA(1, d, 0)$  with a rather weak short memory component such that a large bandwidth would be here appropriate. The opposite situation arises in the second model. The third one is an  $ARFIMA(2, d, 0)$  with a short memory component that shows a spectral peak at

frequency 1.518. Inclusion of neighbouring Fourier frequencies in the estimation generates a large bias and the optimal strategy is to choose a frequency band closer to zero. The last model shows a similar cyclical behaviour now at frequency  $\pi/2$  but in this case the cycle is strongly persistent such that the spectral density function diverges not only at zero but also at  $\pi/2$ . There is here a combination of a cyclical long memory component together with standard long memory at the origin. As in the previous models the standard memory parameter is 0.4 and we consider a cyclical long memory parameter 0.2. This kind of processes implies the presence of a persistent cycle of period four and fits adequately the behaviour of many quarterly economic time series (Arteche and Robinson, 2000).

For the Hurvich and Deo (1999) plug-in bandwidth selection we use  $\delta = 6/7$  and  $A = 0.25$  which are the values suggested by the authors for practical purposes. Regarding the adaptive procedure, Moulines and Soulier (2003) proposed a value of  $\kappa \geq 6$ . However such a choice provides poor finite sample performance and we have found that a lower constant gives better results. We follow here Andrews and Sun (2004) and  $\kappa$  is tuned to the Gaussian *ARFIMA*(1,  $d$ , 0) model with autoregressive parameter  $\phi = 0.6$  and  $d = 0.4$ . That is,  $\kappa$  is chosen as the value that minimizes the Monte Carlo MSE of the adaptive LPE in such a model over 1000 replications with  $n = 512$  for a grid of values  $\{0.05, 0.1, \dots, 6\}$ . The constant chosen in that way is  $\kappa = 1.1$ .

For the local bootstrap bandwidth selection we take  $\underline{m} = 5$ ,  $m_1 = 10$ , and different values of  $m_2$  for each model. Thus, for Model 1 with low dependent short memory component, we know that the best results are obtained with large bandwidths, therefore we consider  $m_2 = 256$  that corresponds to the frequency  $\lambda_{m_2} = \pi$ . For the other three models the best results are obtained with smaller bandwidths so we use  $m_2 = 130$ . For Model 2, with a highly dependent short memory component, a much lower value could have been considered. From a practical point of view, the choice of  $m_2$  can be based on the plot of the extended residuals. If they show a marked structure such as those of Models 2, 3 and 4 in Figure 1 the optimal bandwidth should be low such that there is not need to consider large bandwidths in the iterative procedure. In order to reduce the computational time of each iteration we only consider odd values of the bandwidths. When applied to real series all bandwidths between  $\underline{m}$  and  $m_2$  could (and should) be used.

We consider four values of the resampling width for each model since the performance of the local bootstrap depends on the choice of  $k_n$ . Models with a weak short memory component, such as Model 1, show a better performance with a large resampling width,

since the errors and residuals do not have a marked structure (Figure 1a)). In fact, we have found that in this case the naive regression bootstrap as described in Arteche and Orbe (2005) and Franco and Reisen (2004), which can be compared with a local bootstrap with sufficiently large resampling width, gives better results. On the contrary, when the short memory component shows a higher dependence as in Figure 1b), a smaller  $k_n$  gives better results since the errors and extended residuals show a marked structure and a small  $k_n$  is needed to keep this structure over the bootstrap samples. The structure of the errors and residuals in Models 3 and 4 are halfway between the previous ones such that a medium-large resampling width is more appropriate. With these considerations we analyse resampling widths  $k_n = 1, 2, 4, 8$  for Model 2 and  $k_n = 25, 40, 55, 70$  for Models 1, 3 and 4. The stopping criterion for step 9 is  $\delta = -0.02$  such that the iterative bootstrap only continues if we get an MSE improvement higher than 2%. The number of bootstrap samples considered is  $B = 200$  which is large enough for the approximation of the MSE (Efron and Tibshirani, 1993).

The basic condition for the local bootstrap to be used for bandwidth selection is that the bootstrap MSE should be a good approximation of the Monte Carlo MSE (taken as the best approximation to the finite sample MSE) for the range of bandwidths considered, or at least around the true optimal bandwidth, such that both achieve its infimum around the same  $m$ . Figure 2 shows the results for the four models using four different resampling widths, except for Model 1 where the basic residual bootstrap is used for the reasons explained above. The resemblance of the bootstrap MSE and the Monte Carlo MSE is quite remarkable, even though it can be improved with a more elaborated selection of the resampling width as explained below.

The analogy of the bootstrap MSE and Monte Carlo MSE suggests that the local bootstrap could be used to approximate the MSE for optimal bandwidth selection. Table 1 compares this strategy with the Hurvich and Deo (1999) and adaptive proposals described above. For each model, Table 1 shows the Monte Carlo MSE of the LPE with each bandwidth selection strategy and its infimum, the Monte Carlo mean, median and standard deviation of the different bandwidth selections together with the bandwidth that minimizes the Monte Carlo MSE over the 1000 replications, which represents here the best possible situation. This information is complemented with Figures 3-5 displaying the histograms of the different bandwidths selected over the 1000 replications for Model 2, 3 and 4. The histograms corresponding to Model 1 are not shown to save space because it accumulates, as expected, on the values close to  $m = 255$ , except for the Hurvich and Deo selection strategy

Table 1: Optimal bandwidth selection under different strategies

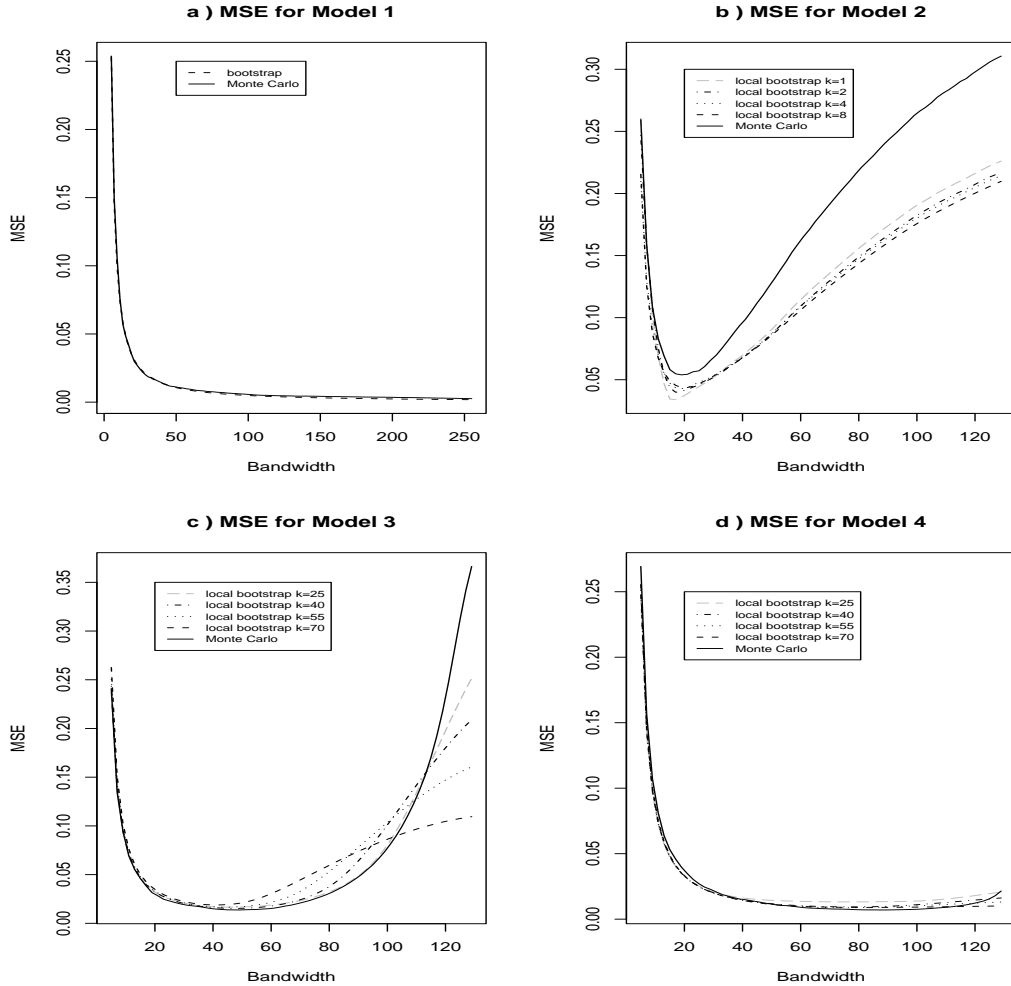
Model type	selection strategy	MSE	m-opt (mean)	m-opt (dev)	m-opt (med)
Model 1	Monte Carlo	0.00255	255		
	Adaptive	0.05969	176.2	106.8	255
	Hurvich-Deo	0.01652	52.6	31.2	43
	bootstrap	0.00260	253.2	3.42	255
Model 2	Monte Carlo	0.05393	19		
	Adaptive	0.19553	59.4	38.3	55
	Hurvich-Deo	0.07967	36.5	17.6	31
	local bootstrap $k_n = 1$	0.06295	33.4	22.8	27
	local bootstrap $k_n = 2$	0.06680	35.7	25.0	27
	local bootstrap $k_n = 4$	0.06965	36.5	24.9	29
	local bootstrap $k_n = 8$	0.07710	38.3	25.8	31
Model 3	Monte Carlo	0.01368	49		
	Adaptive	0.08178	68.7	35.3	83
	Hurvich-Deo	0.03655	48.7	24.8	40
	local bootstrap $k_n = 25$	0.02568	60.4	19.4	59
	local bootstrap $k_n = 40$	0.02005	54.4	14.7	53
	local bootstrap $k_n = 55$	0.01742	47.1	10.0	45
	local bootstrap $k_n = 70$	0.01937	42.3	10.7	41
Model 4	Monte Carlo	0.00700	89		
	Adaptive	0.07967	90.2	49.5	125
	Hurvich-Deo	0.01968	50.6	24.8	42
	local bootstrap $k_n = 25$	0.01692	96	28.9	101
	local bootstrap $k_n = 40$	0.01367	95.7	26.1	93
	local bootstrap $k_n = 55$	0.01311	95.8	27.5	91
	local bootstrap $k_n = 70$	0.01581	101.6	30.4	121

where the values range mainly from  $m = 25$  to  $m = 100$ .

With all this information we can emphasize the following conclusions:

- The four models considered cover a complete range of values for optimal bandwidth, Model 1 with a weak short memory component corresponds to a very large optimal bandwidth, Model 2 on the contrary agrees with a low  $m$  and Models 3 and 4 have medium optimal bandwidths.
- The MSE of the LPE estimates obtained with the bootstrap bandwidth selection strategy are always lower than those obtained with the Hurvich and Deo and adaptive selection criteria for any resampling width. The proposal of Hurvich and Deo tends to

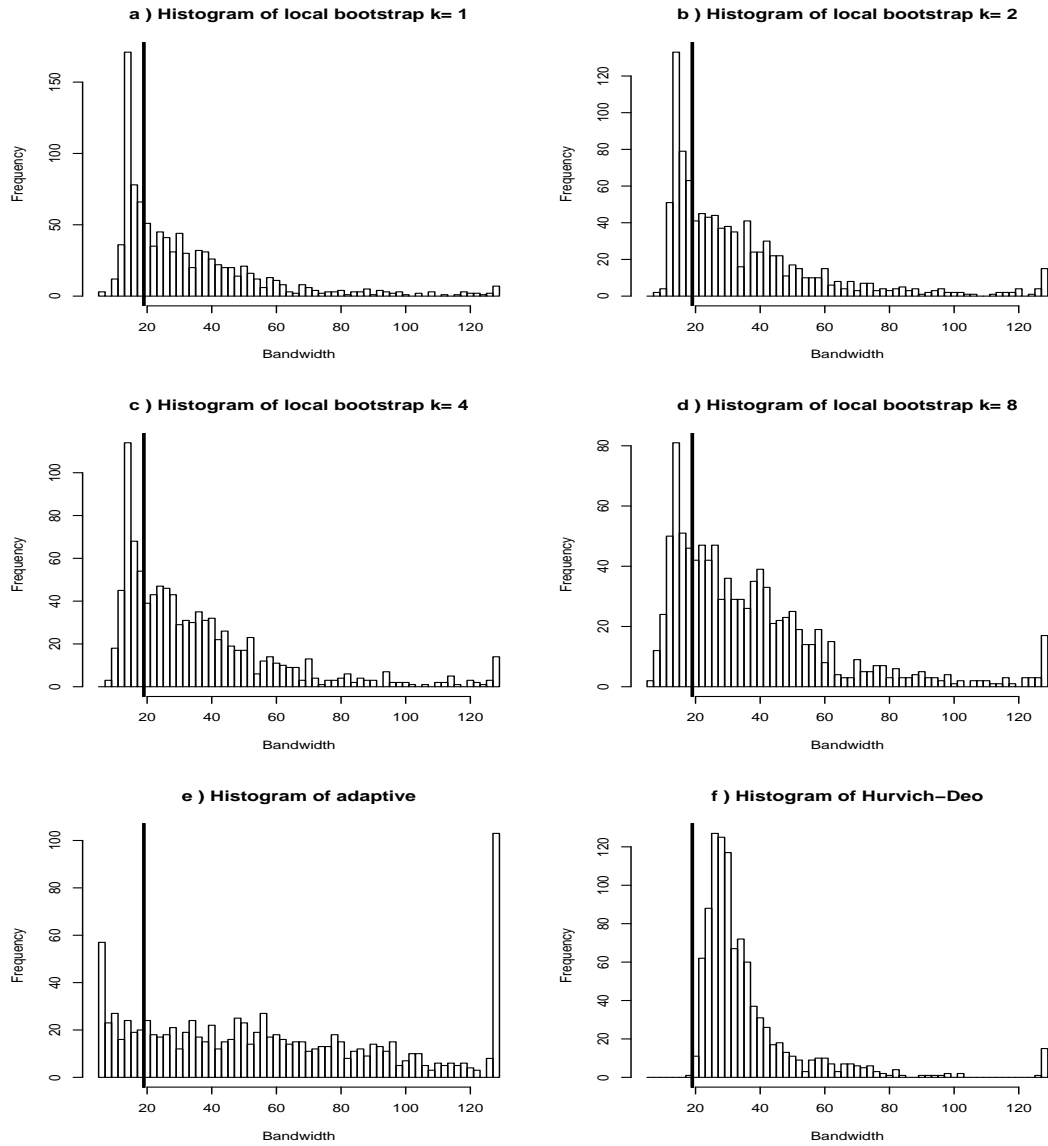
Figure 2: Monte Carlo MSE and bootstrap MSE



be positively biased in Model 2, which corresponds to a low optimal bandwidth, and negatively biased in the other models, which need a larger bandwidth. The adaptive has the worst behaviour overall.

- The performance of the local bootstrap depends on the choice of the resampling width  $k_n$ , a low  $k_n$  being more adequate for those cases with a marked structure in the residuals. For Model 2 a resampling width  $k_n = 1$  works better while in Models 3 and 4 larger resampling widths seems more appropriate. Model 1 shows very little structure in the residuals such that a large  $k_n$  is more appropriate. In fact the typical naive residual bootstrap that considers all the residuals in the resampling procedure

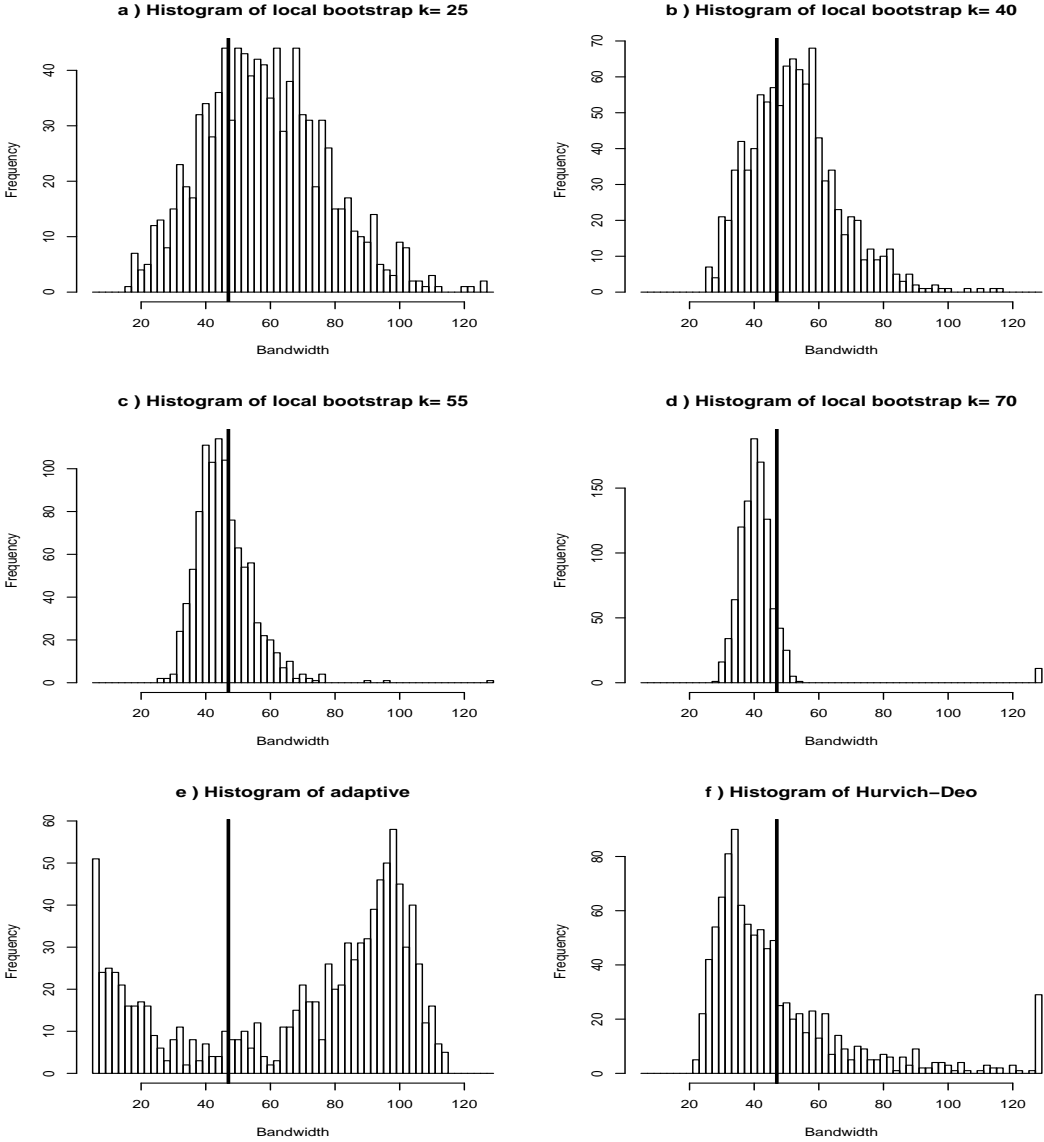
Figure 3: Histogram of optimal  $m$  under different selection strategies in Model 2



The straight vertical line represents the bandwidth that minimizes the Monte Carlo MSE.

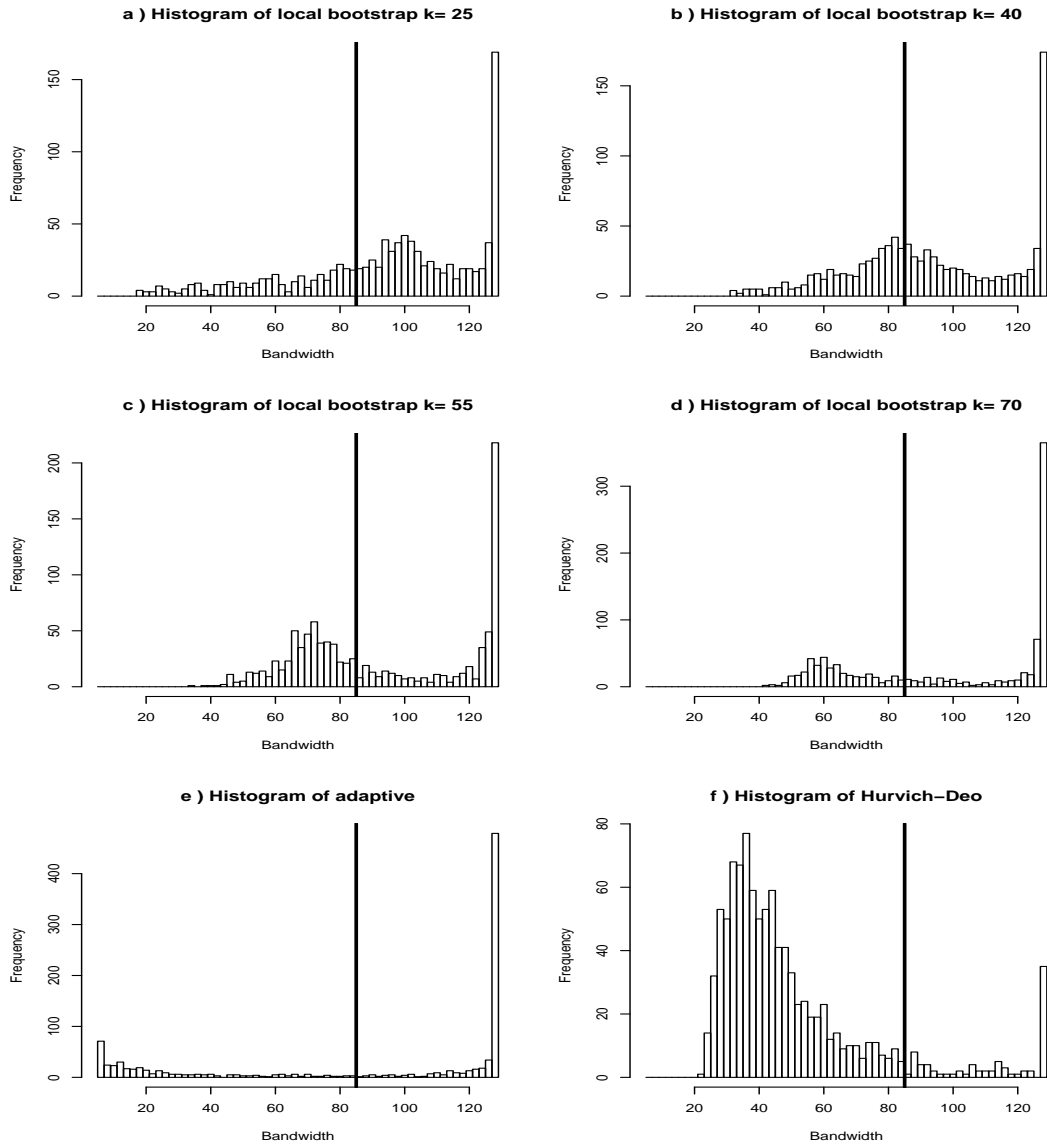


Figure 4: Histogram of optimal  $m$  under different selection strategies in Model 3



The straight vertical line represents the bandwidth that minimizes the Monte Carlo MSE.

Figure 5: Histogram of optimal  $m$  under different selection strategies in Model 4



The straight vertical line represents the bandwidth that minimizes the Monte Carlo MSE.

gives here better results.

- The bootstrap optimal bandwidths tend to be closer in mean and median to the Monte Carlo optimal bandwidth than the other selection techniques. In addition, in Models 1 and 3 the bootstrap gives rise to the smallest standard deviations whereas in Models 2 and 4 the standard deviation is not much larger than the lowest one that corresponds to the (biased) Hurvich and Deo selection strategy. The histograms of the optimal bandwidths in Figures 3-5 displays further information. In general, we find an accumulation of the bootstrap selected bandwidths around the Monte Carlo optimal bandwidth represented by a thick straight line.
- The bootstrap bandwidth selection performs satisfactorily well in all cases, especially in Models 1 to 3, which correspond to a local spectral smoothness of  $\alpha = 2$ . In Model 4 we observe many cases where the selected optimal bandwidth is the superior bound of the values considered, but this happens in the three strategies here analyzed and in general the bootstrap bandwidth selection overcomes the others in terms of MSE, with a choice closer to the minimizer of the Monte Carlo MSE.
- It is noteworthy the performance of the bootstrap in Model 3 that corresponds to a medium optimal bandwidth. Whereas the Hurvich and Deo selection shows a significant bias and the behaviour of the adaptive is really poor, our bootstrap proposal shows lower MSE and an empirical distribution concentrated around the optimal value with a dispersion that decreases with the resampling width. This behaviour was also found by Paparoditis and Politis (1999) in estimating the spectral density function, increasing  $k_n$  increases the bias and decreases the variance. This satisfactory behaviour is explained by the good approximation of the bootstrap MSE to the Monte Carlo MSE around those values where the minimum is achieved (see Figure 2). A resampling width of 70 is definitely too large and the poorer approximation of the MSE generates the negative bias in Table 1 and Figure 4.

As it stands our procedure is not fully data driven since it requires the participation of the user in the selection of the resampling width. Although the bootstrap bandwidth selection overcomes other strategies for all the different resampling widths here considered, different choices of  $k_n$  lead to different results and some criterion should be established. We propose two rules that can be used for a prior selection of  $k_n$ :

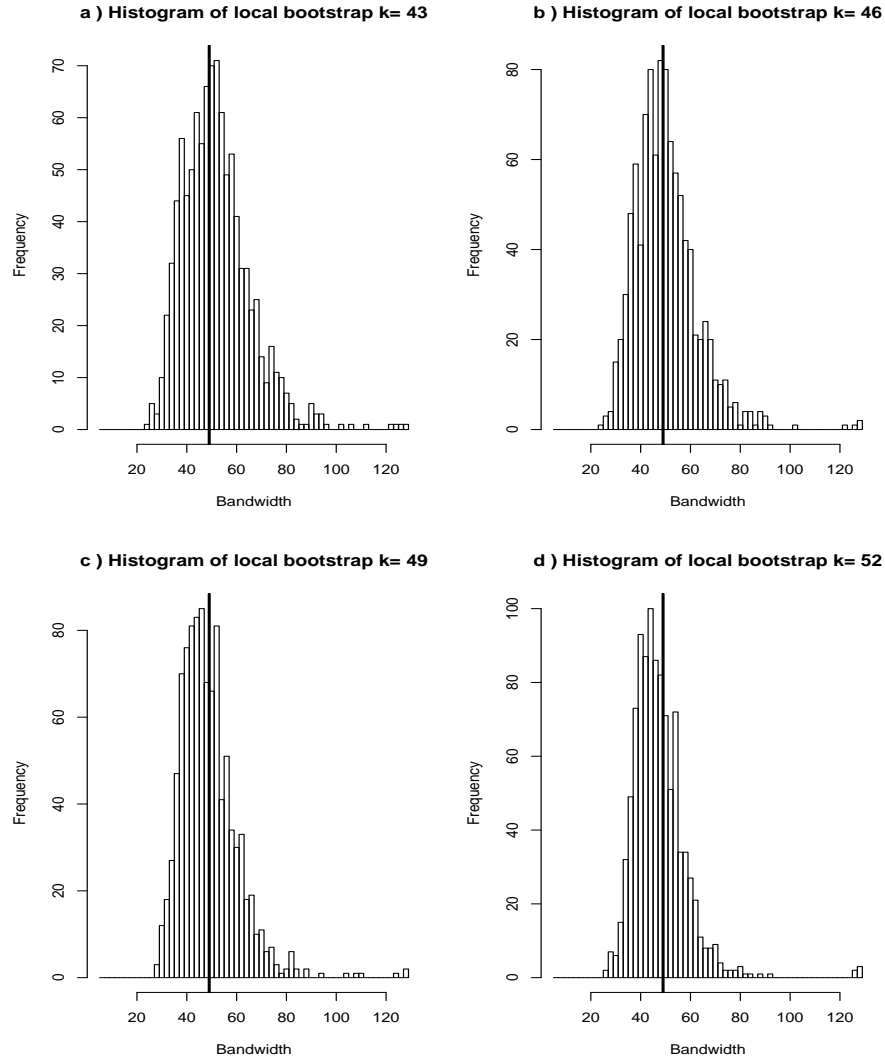
- Based on the graph of the extended residuals, a large  $k_n$  should be chosen if they show a stable behaviour. The more marked structure in the residuals the lower  $k_n$  should be chosen.
- Based on the graph of the bootstrap MSE with different resampling widths, a high  $k_n$  should be used if their shape do not vary with  $k_n$  and they get a minimum at a large  $m$ . Otherwise choose a lower  $k_n$ .

In this manner we can chose a resampling width that will give rise to satisfactory results, bearing in mind that we do not need a sharp selection of  $k_n$  since the conclusions are not altered by the use of similar widths. Further refinements on the resampling width selection can also be used. Table 2 shows the results for Model 3 considering a thinner grid of values for  $k_n$  in the range  $[40, 55]$ . Some MSE reduction is achieved but in general the results are quite robust to the resampling width selection in this range of values. The histograms of the selected bandwidths in Figure 6 indicate that a further refinement in the choice of  $k_n$  has a slight improvement in the bias of the selected bandwidth. The bootstrap MSEs in Figure 7a) show also a similar pattern compared with the Monte Carlo MSEs.

Table 2: Further refinements of the resampling width

Model type	selection strategy	MSE	m-opt (mean)	m-opt (dev)	m-opt (med)
Model 3	local bootstrap $k_n = 25$	0.02568	60.4	19.4	59
	local bootstrap $k_n = 40$	0.02005	54.4	14.7	53
	local bootstrap $k_n = 43$	0.01899	52.8	13.7	51
	local bootstrap $k_n = 46$	0.01780	51.1	12.5	49
	local bootstrap $k_n = 49$	0.01727	49.8	11.6	47
	local bootstrap $k_n = 52$	0.01748	48.4	10.8	47
	local bootstrap $k_n = 55$	0.01742	47.1	10	45
	local bootstrap $k_n = 70$	0.01937	42.3	10.7	41
Model 4	local bootstrap $k_n = 25$	0.01692	96	28.9	101
	local bootstrap $k_n = 40$	0.01367	95.7	26.1	93
	local bootstrap $k_n = 55$	0.01311	95.8	27.5	91
	local bootstrap $k_n = 70$	0.01581	101.6	30.4	121
	local bootstrap moving $k_n$	0.00961	101.5	20.6	109

Figure 6: Further refinements of the resampling width in Model 3: Histograms



The straight vertical line represents the bandwidth that minimizes the Monte Carlo MSE.

We may also consider the possibility of a moving resampling width, as proposed by Paparoditis and Politis (1999). This strategy could be adequate when we observe an unstable behaviour in the extended residuals. For example, the residuals in Model 4 show a flat behaviour at low frequencies and then increase, quite rapidly at the end of the range of frequencies. Table 2 shows the results with a resampling width  $k_n = 70$  for the first 50 frequencies,  $k_n = 40$  for  $50 < j < 100$ , and  $k_n = 1$  for frequencies  $j > 100$ . The improvement here is quite remarkable, with a reduction in the MSE over 25% and a bootstrap MSE that closely resembles the Monte Carlo MSE (see Figure 7b)). Finally the histogram in Figure 8 displays a smaller concentration at extreme values than with the previous resampling

schemes.

As a conclusion, Table 3 shows the increments in the Monte Carlo MSE of the LPE due to the use of different bandwidth selection criteria compared with the MSE of the LPE using the best bandwidth in a minimum Monte Carlo MSE sense. We only report the results of the bootstrap bandwidth selection linked with the best selection of the resampling. The improvement of this strategy over the other criteria is noteworthy with a MSE that is much closer to the optimal.

Table 3: LPE-Bandwidth selection MSE increment w.r.t. Monte Carlo optimal MSE

Model type	selection strategy	diff. MSE (%)	diff. MSE
Model 1	Adaptive	2235.77	0.05713
	Hurvich-Deo	546.54	0.01396
	bootstrap	1.84	0.00004
Model 2	Adaptive	262.53	0.14159
	Hurvich-Deo	47.72	0.02574
	local bootstrap $k_n = 1$	16.73	0.00902
Model 3	Adaptive	497.60	0.06810
	Hurvich-Deo	167.12	0.02287
	local bootstrap $k_n = 49$	26.25	0.00359
Model 4	Adaptive	847.20	0.07266
	Hurvich-Deo	157.15	0.01268
	local bootstrap moving $k_n$	37.30	0.00261

## 4 Empirical application

This section applies the different bandwidth selection strategies discussed above to two real time series, namely the Nile river annual minimum levels recorded between 622 and 1284 A.D. and the input gas rate which corresponds to series J in Box and Jenkins (1976). The first series has been widely analyzed in the long memory literature and also for bandwidth selection in the local Whittle estimation by Henry and Robinson (1996), and comprises a total of 663 observations. The second series represents the methane rate incorporated every nine seconds in a furnace to form a mixture of gases containing  $CO_2$ . This last series corresponds to series J in the seminal Box and Jenkins (1976) book for time series analysis

Figure 7: Monte Carlo MSE and bootstrap MSE

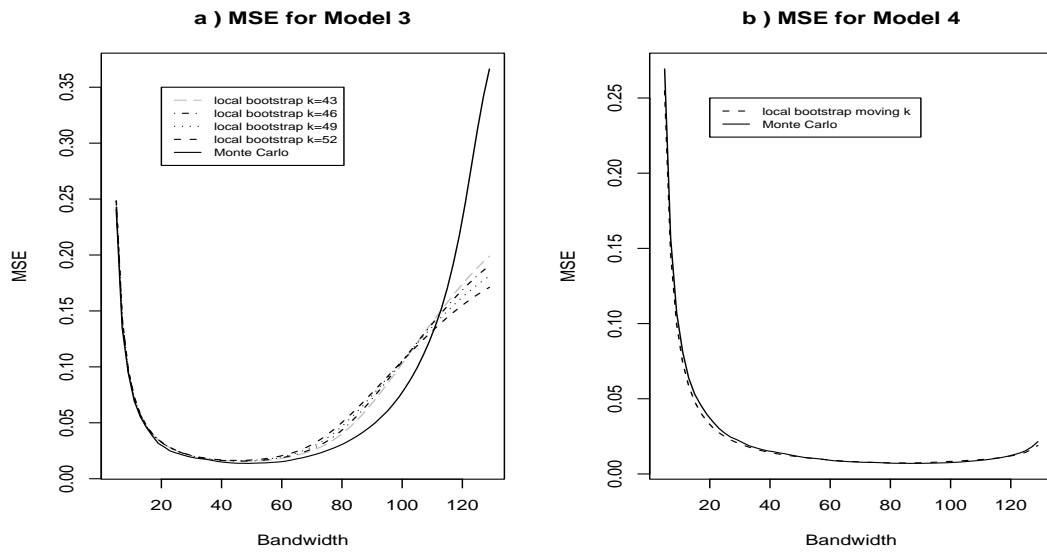
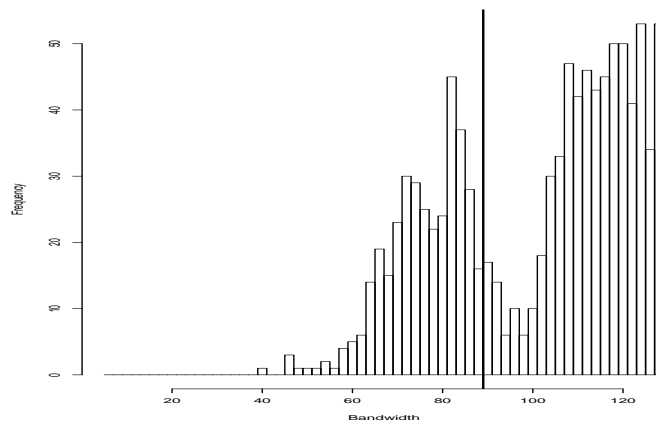


Figure 8: Further refinements of the resampling width in Model 4: Histogram



The straight vertical line represents the bandwidth that minimizes the Monte Carlo MSE.

and consists of 296 observations. Both series are displayed in the first panel of Figures 9 and 10. Panel b) in both figures shows the LPE estimates of the memory parameter over a grid of bandwidths. The variability of these estimates is quite remarkable and the statistical conclusions about the stationarity of the series depend heavily on the bandwidth selected.

Table 4 shows the bandwidth selected with the bootstrap, Hurvich and Deo and adaptive criteria, together with the LPE estimate with the corresponding bandwidth selection and the standard deviation calculated as  $\pi^2(6 \sum(z_j - \bar{z})^2)^{-1}$  which has been proven to give better results in finite samples than the standard deviation in the asymptotic distribution in formula (4). Regarding the Nile river series, the stable behaviour of the extended residuals in Figure 9c) suggests that a large resampling width for the local bootstrap should be used. This is confirmed by the bootstrap MSE in Figure 9d) whose shape is quite robust to the resampling width selection. The results displayed in the upper half of Table 4 correspond to the basic naive bootstrap, but similar choices are attained with the local bootstrap with a large resampling width. According to this selection the memory parameter of the series falls in the stationary region with a high confidence, and of course the whole 95% Gaussian based confidence interval belongs to this region. However the bandwidth selected by the Hurvich and Deo strategy is much smaller and arises some uncertainty about the stationarity of the series, the 95% confidence interval including values of the memory parameter corresponding to both the stationary and nonstationary region. This bandwidth choice is close to the values found in Henry and Robinson (1996) who proposed a bandwidth selection technique for the local Whittle estimator of the memory parameter which is very similar in spirit to the proposal of Hurvich and Deo. They obtained an optimal bandwidth estimation between 53 and 71, depending on the number of iterations of their strategy, which seems to be too low according to our analysis.

The extended residuals of the input gas rate series in Figure 10c) show a marked structure such that a lower resampling width seems here more adequate for the local bootstrap. Figure 10d) shows the bootstrap MSE for different resampling widths. While  $k_n = 70$  and  $k_n = 40$  seems to be too large, we get similar bandwidth selections for a resampling width between 1 and 10, with bandwidth choices between 27 and 30 (we have also proved different moving resampling widths as suggested in the previous sections and the minimum of the bootstrap MSE was always attained between those two values). Table 4 shows the results for a bandwidth choice of 27 which was the most frequent selection. In this case the bootstrap selects a value quite close to the bandwidth selected by the criteria of Hurvich and Deo,



Table 4: LPE-Bandwidth selection and estimation of the memory parameter

		Bootstrap	Hurvich-Deo	Adaptive
Nile river	$m - opt$	250	40	259
	$\hat{d}(m - opt)$	0.372	0.475	0.367
	$\hat{sd}(m - opt)$	0.042	0.118	0.042
Gas rate	$m - opt$	27	24	46
	$\hat{d}(m - opt)$	0.412	0.467	0.732
	$\hat{sd}(m - opt)$	0.149	0.161	0.108

but the adaptive gives a too large selection with an LPE estimate that clearly falls in the nonstationary region.

## 5 Conclusion and possible extensions

The finite sample performance of the LPE depends considerably on the bandwidth used in the estimation, such that for the same series we can get different conclusions. For example we can get stationarity or nonstationarity depending on the bandwidth choice such that estimates below and above 0.5 are possible just by using a different number of frequencies. We propose here a data driven minimum MSE bandwidth selection criterion that clearly outperforms other existing techniques such as the plug-in version of Hurvich and Deo (1999) or the adaptive LPE of Moulines and Soulier (2003). The technique is based on the local bootstrap such that a prior selection of the resampling width is necessary. We have also proposed data driven strategies for the choice of this resampling width and have shown, via Monte Carlo and with empirical examples, that our proposal is quite robust to the resampling width selection as long as you do not misinterpret the signals extracted from the data and make a pervert choice. The technique is easy to apply and can be extended to more general regression contexts such as the non linear versions of the LPE of Sun and Phillips (2003) and Arteché (2006) or to the bias reduced LPE of Andrews and Guggenberger (2003). Further research about the theoretical properties of the proposed bandwidth selection procedure and its applicability in other more general contexts seem to be worthy.

## Acknowledgements:

Research supported by the Ministerio de Ciencia y Tecnología and FEDER grant SEJ2007-61362, and by the Department of Education of the Basque Government grant IT-334-07 (UPV/EHU Econometrics Research Group). We also thank Marc Henry for providing information about the series in the empirical analysis.

## References

- [1] Andrews, D.W.K. and Guggenberger, P. (2003) A bias-reduced log-periodogram regression estimator for the long-memory parameter. *Econometrica* 71, 675-712.
- [2] Andrews, D.W.K. and Sun, Y. (2004) Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica* 72, 569-614.
- [3] Arteche, J. (2006) Semiparametric estimation in perturbed long memory series. *Comput. Statist. Data Anal.* 52, 2118-2141.
- [4] Arteche, J. and Orbe, J. (2005) Bootstrapping the log-periodogram regression. *Econ. Letters* 86, 79-85.
- [5] Arteche, J. and Robinson, P.M. (2000) Semiparametric inference in seasonal and cyclical long memory processes. *J. Time Ser. Anal.* 21, 1-27.
- [6] Box, G. E. P. and Jenkins, G. M. (1976) *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- [7] Efron, B. and Tibshirani, R.J. (1993) *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- [8] Dahlhaus, R. and Janas, D. (1996) A frequency domain bootstrap for ratio statistics in time series analysis. *Ann. Statist.* 24, 1934-1963.
- [9] Franco, G.C. and Reisen, V.A. (2004) Bootstrap techniques in semiparametric estimation methods for ARFIMA models: a comparison study. *Comput. Statist.* 19, 243-259.
- [10] Franke, J. and Härdle, W. (1992) On bootstrapping kernel spectral estimates. *Ann. Statist.* 20, 121-145.
- [11] Giraitis, L., Robinson, P.M. and Samarov, A. (2000) Adaptive semiparametric estimation of the memory parameter. *J. Multiv. Anal.*, 72, 183-207.

- [12] Henry, M. and Robinson, P.M. (1996) Bandwidth choice in Gaussian semiparametric estimation of long range dependence. In *Athens Conference on Applied Probability and Time Series, Vol. II. Lecture Notes in Statistics* (eds P.M. Robinson and M. Rosenblatt), vol. 115, pp. 220-232. New York: Springer-Verlag.
- [13] Hurvich, C.M. and Deo, R.S. (1999) Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long-memory time series. *J. Time Ser. Anal.* 20, 331-341.
- [14] Hurvich, C.M., Deo, R.S. and Brodsky, J. (1998) The mean squared error of Geweke and Porte-Hudak's estimator of the memory parameter of a long memory time series. *J. Time Ser. Anal.* 19, 19-46.
- [15] Moulines, E. and Soulier, P. (2003) Semiparametric spectral estimation for fractional processes. In *Theory and Applications of Long-Range Dependence* (eds P. Doukhan, G. Oppenheim and M-S. Taqqu), pp. 251-301. Boston: Birkhäuser.
- [16] Paparoditis, E. and Politis, D. (1999) The local bootstrap for periodogram statistics. *J. Time Ser. Anal.* 20, 193-222.
- [17] Silva, E.M., Franco, G.C., Reisen, V.A. and Cruz, F.R.B. (2006) Local bootstrap approaches for fractional differential parameter estimation in ARFIMA model. *Comput. Statist. Data Anal.* 51, 1002-1011.
- [18] Sun, Y. and P.C.B. Phillips (2003) Nonlinear log-periodogram regression for perturbed fractional processes. *J. Econometrics* 115, 355-389.

Figure 9: Nile river annual minimum

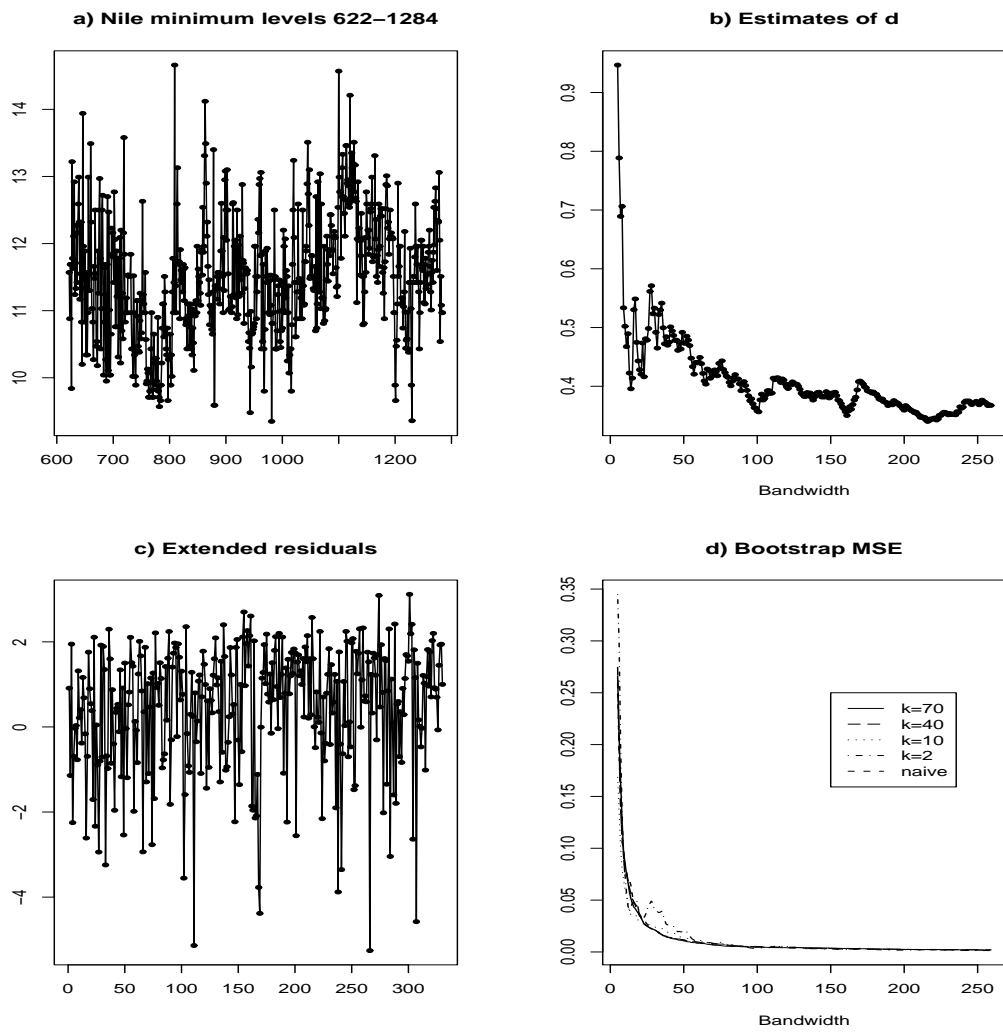


Figure 10: Input gas rate (series J of Box and Jenkins)

