Bootstrapping the Log-periodogram Estimator of the Long-Memory Parameter: is it Worth Weighting?

S. M. Heravi, K. D. Patterson
Cardiff Business School, University of Cardiff
Department of Economics, University of Reading

Abstract

Estimation of the long-memory parameter from the log-periodogram (LP) regression, due to Geweke and Porter-Hudak (GPH), is a simple and frequently used method of semi-parametric estimation. However, the simple LP estimator suffers from a finite sample bias that increases with the dependency in the short-run component of the spectral density. In a modification of the GPH estimator, Andrews and Guggenberger, AG (2003) suggested a bias-reduced estimator, but this comes at the cost of inflating the variance. To avoid variance inflation, Guggenberger and Sun (2006) suggested a weighted LP (WLP) estimator using bands of frequencies, which potentially improves upon the simple LP estimator. In all cases a key parameter in these methods is the need to choose a frequency bandwidth, m, which confines the chosen frequencies to be in the 'neighbourhood' of zero. GPH suggested a 'square-root' rule of thumb that has been widely used, but has no optimality characteristics. An alternative, due to Hurvich and Deo (1999), is to derive the root mean square error (rmse) optimising value of m, which depends upon an unknown parameter, although that can be consistently estimated to make the method feasible. More recently, Arteche and Orbe (2009a,b), in the context of the GPH estimator, suggested a promising bootstrap method, based on the frequency domain, to obtain the rmse value of m that avoids estimating the unknown parameter. We extend this bootstrap method to the AG and WLP estimators and to consideration of bootstrapping in the frequency domain (FD) and the time domain (TD) and, in each case, to 'blind' and 'local' versions. We undertake a comparative simulation analysis of these methods for relative performance on the dimensions of bias, rmse, confidence interval width and fidelity.

Keywords: Long memory; bootstrap; log-periodogram regression; variance inflation; weighted LP regression; time domain; frequency domain.

JEL classification: C1; C14

Corresponding author: Saeed Heravi; Heravis@cardiff.ac.uk

I. Introduction

There have been a number of suggestions to improve the basic log-periodogram, LP, estimator of the long-memory parameter, d, due to Geweke and Porter-Hudak (1983), (hereafter GPH) which, although popular, is known to be biased in finite samples, see for example Agiakloglou et al (1993). Andrews and Guggenburger (2003), hereafter AG, suggested improving the GPH estimator by basing an estimator of d on a higher order Taylor series expansion of the log periodogram. This method, however, inflates the estimator variance and, to avoid this undesirable feature, Guggenberger and Sun (2004, 2006), hereafter GS, suggested a method that weights N component LP estimators of increasing bandwidths into a single estimator, referred to as a weighted LP, WLP, estimator; by an appropriate choice of parameters, the bias can be controlled without inflating the variance; thus, the WLP estimator offers the possibility of dominating the simple LP and AG estimators.

Whilst our preliminary simulations indicated that the WLP procedure offered improvements over the simple LP estimator, recent research has suggested that bootstrapping can be a fruitful way of improving the LP estimator, especially in determining the bandwidth, m, which is the number of spectral frequencies included in the LP pseudo-regression. Whilst the variance is driven down as m increases, the inclusion of too many short-run frequencies increases the bias; on the other hand, reducing the number of included frequencies reduces the bias, but increases the variance. Hence, as Arteche and Orbe (2009b), hereafter AO, have shown, m is a key parameter in determining the actual, rather than theoretical, properties of the resulting estimator.

There have been a number of suggestions to overcome the problem of choosing m in the context of the standard LP regression. 1) GPH (1983) suggested the square root rule, $m = \sqrt{n}$, where n is the number of observations; this idea has led to the use of a mechanical rule of the form $m = n^{\gamma}$, but without any clear uniformity on the choice of γ. 2) Hurvich and Deo (1999), hereafter HD, suggested choosing m to minimise the root mean squared error, rmse, of the LP estimator; although this is not quite a 'plug-in' rule, it is relatively easy to implement with prior estimation of an extended LP

regression. 3) For the standard case, AO (2009b) suggested a bootstrap version of the rmse rule, which performed well relative to the HD method.

In order to undertake a comparison of the leading LP-based estimators, we consider a number of issues in this paper, as follows:

1. The design of a bootstrap scheme for the WLP estimator that respects the underlying sequential nature of the N component LP estimators. This is important in order to mimic the way that the data is generated. (A bootstrap is also undertaken for the AG estimator, but that is a trivial extension of that for the LP estimator.)
2. The design of bootstraps for the frequency domain (FD) and, alternately, the time domain (FD) for the various LP estimators considered here.
3. Bootstrapping using local resampling in both the FD and TD cases.
4. Simulating the rmse of the WLP estimator (and other estimators) in order to select the operational bandwidth and compare that choice with others, including a mechanical rule and a feasible implementation of the HD rule.
5. We also undertake a comparative analysis of the LP, AG and WLP estimators across these various dimensions in terms of an rmse comparison and the fidelity of the resulting confidence intervals.

Some preliminary comments motivate these issues. On point 1), care has to be taken in the design of the WLP bootstrap procedure in order to preserve the sequential way in which the data is generated for the N component LP estimators. A residual based bootstrap is used here but, for example, an unrelated resampling scheme for each of the N LP estimators does not respect the generation of the data in the actual WLP procedure. On point 2), the bootstrap can be based on resampling in the frequency domain, as in AO (2005), (2009a, b), or in the time domain, as in Franco and Reisen (2004, 2007), Poskitt (2008) and Kapetanios and Papailias (2011). On point 3), resampling in either the FD or TD may be 'blind' or 'local' in some form see, for example, Paparoditis and Politis (1999, 2001, 2003) and AO (2009b). The bootstrap scheme used here for the WLP estimator may be described as a 'sighted' bootstrap in the sense of respecting the sequential nature of the component LP estimators.

The differences between 'blind', 'local' and 'sighted' resampling schemes are described in detail in section 4, but some motivation for these terms can be simply described at this point. In a standard bootstrap from a sample of length T (for time) or M (for frequency), resampling can occur from anywhere in the sample, that is $t = 1, \ldots, T$ or $j = 1, \ldots, M$, respectively; however, to mimic the properties of an estimator that does not

use all of the time or frequency observations, a (fixed) window of limited length around some usually centred point is specified; for example, sample from adjacent observations or frequencies, in which case the resampling process is described as 'local', that is from the 'window' of choice. The situation is more complex in the case of the WLP estimator, which uses an increasing sequence of frequencies for several LP estimators, weighting the end result to obtain an overall estimator; thus, it would not be appropriate to use a either a 'blind' or fixed 'local' window resampling scheme. Rather, we allow the resampling window to increase in length to reflect the increasing number of frequencies that are used in calculating the component LP estimators: heuristically, it is able to 'see' how the choice of 'local' is not fixed but expands with the required number of frequencies.

On point 4), and developing the method in AO (2009b), we use a simulation method to select the value of m that minimises the rmse and also compare that choice with an operational version of the rule of selecting m by minimising the asymptotic rmse and with the often used square-root rule. On point 5) there are a number of dimensions to choosing an LP-type estimator, specifically: i) which estimator, ii) which bootstrap design, iii) should the bootstrap be in the frequency domain or the time domain and iv) how is the number of included frequencies chosen? Thus, we provide some simulation guidance on these issues. Moreover, as an issue often of practical concern is to use the estimates to form confidence intervals, we also consider the fidelity of the confidence interval (CI) coverage compared to the nominal coverage and the coverage width. Ideally, one seeks the best fidelity and the shortest interval, but some trade-off may be necessary in practice.

This paper is organised as follows. In order to make the paper reasonably self-contained, the derivation of the LP and WLP estimators, and their properties, are briefly outlined in sections 2 and 3, respectively. Section 4 details the bootstrap routine for the WLP estimator and section 5 reports the results from a number of simulation experiments. Section 6 contains concluding remarks.

2. The log-periodogram estimator (LPE): GPH and AG estimators

We consider a data series $\{y_i\}_1^n$ with spectral density function given by:

$$f_y(\lambda_j) = |(1 - e^{-i\lambda_j})|^{-2d} f_u(\lambda_j) \qquad (1)$$

where $d \in (-½, ½)$ is the 'long-memory' parameter and $f_u(\lambda_j)$ captures the short-run frequency components of the series, such that $f_u(\lambda_j) \to G$, a finite constant, as

$\lambda_j \to 0^+$; other assumptions are as in AG (2003) and GS (2004, 2006). The LP estimator, $\hat{d}_{LP}$, is obtained from least squares (LS) estimation of d in the following pseudo-regression, for example, Robinson (1995):

$$\ln[I_y(\lambda_j)] = \beta_0 - 2d[\ln(\lambda_j)] + \varepsilon_j \qquad j = 1, \ldots, m. \qquad (2)$$

where $I_y(\lambda_j) = |\omega(\lambda_j)|^2$ is the periodogram of y and $\omega(\lambda_j)$ is the discrete Fourier transform evaluated at the Fourier frequencies $\lambda_j = 2\pi j/n$ for j = 1, ..., [m]; m is a user-selected integer truncation parameter confining the frequencies to the neighbourhood of the origin from the right ($\lambda_j \to 0^+$). The asymptotic variance of $\hat{d}_{LP}$ is $\sigma^{2,asym}(\hat{d}_{LP}) = \pi^2/(24m)$, whilst the finite sample standard error of $\hat{d}_{LP}$ is obtained in the usual way from the LS estimation of the regression (2).

To overcome the finite sample bias of the LP estimator, AG (2003) suggested a bias-reduced estimator, obtained by augmenting the simple LP regression with even powers of λ_j, which arises from a Taylor series expansion of $\log[f_u(\lambda_j)]$, that is:

$$\ln[I_y(\lambda_j)] = \beta_0 - 2d[\ln(\lambda_j)] + \sum_{k=1}^{r} \varphi_k \lambda_j^{2k} + \breve{\varepsilon}_j \qquad j = 1, \ldots, m \qquad (3)$$

where $\varphi_k = b_{2k}/(2k)!$ and $b_k = \dfrac{\partial^k \{\ln[f_u(\lambda_j)]\}}{\partial \lambda_j^k}\Big|_{\lambda_j = 0}$.

The resulting LS estimator is denoted $\hat{d}_{LP}(r)$, of which the LP estimator is seen to be the special case $\hat{d}_{LP}(0)$, and the AG estimator for r = 1 is denoted $\hat{d}_{LP}(1)$.

The asymptotic bias, asymptotic variance of $\hat{d}_{LP}(r)$ and the asymptotically optimal value of m, obtained by minimising the rmse, are given, respectively, by:

$$AB(\hat{d}_{LP}(r)) = \tau_r b_{2+2r} \frac{m^{2+2r}}{n^{2+2r}} \qquad (4)$$

$$\sigma^2(\hat{d}_{LP}(r)) = \frac{\pi^2}{24m} c_r \qquad (5)$$

$$m_{LP}^{opt}(r) = K(r) n^{(4+4r)/(5+4r)} \qquad (6a)$$

where $K(r) = \left(\frac{\pi^2 c_r}{24(4 + 4r)\tau_r^2 b_{2+2r}^2} \right)^{1/(5+4r)}$. (6b)

AG (2003) provide the sets $\{c_r\}$ and $\{\tau_r\}$ of known constants; for example, $c_0 = 1$, $c_1 = 2.25$, $c_2 = 3.52$, $c_3 = 4.79$ and $\tau_0 = -2.19$, $\tau_1 = 2.23$, $\tau_2 = -0.793$, $\tau_3 = 0.146$. For given m, $\sigma^2(\hat{d}_{LP}(r))$, $r \geq 1$, is inflated relative to $\sigma^2(\hat{d}_{LP}(0))$ by the constant c_r. The unknown coefficient in $m_{LP}^{opt}(r)$ is b_{2+2r}, see K(r), which can be estimated from the regression (3), but with r of one higher order than required for $\hat{d}_{LP}(r)$; this regression uses $\overline{m}$ frequencies, where $\overline{m} = Kn^{\alpha}$, $\alpha = (8 + 4r)/(9 + 4r)$, $K > 0$. (The choice of a value for K is considered below.) Although not fully automatic, the resulting estimator, $\hat{m}_{LP}^{opt}(r)$, is sometimes referred to as a 'plug-in' estimator and is based on Hurvich and Deo (1999). The AG estimator, despite its improvement in bias reduction, has not, however, displaced use of the basic LP estimator, partly because of the variance inflation.

3. The weighted LP (WLP) estimator

The estimator suggested by GS (2004, 2006), hereafter denoted $\hat{d}_{WLP}(r)$, controls the variance inflation in the AG method. In summary, the GS estimator involves splitting the overall number of included frequencies into a finite number N of bands of increasing length m_i and estimating d using the LP estimator, $\hat{d}_{LP}(0, m_i)$, for each of the frequency intervals. The overall estimator is then a weighted average of the N component estimators, where the weights are chosen to ensure asymptotic unbiasedness by eliminating the first r dominant terms in the bias. In practice, r is a choice parameter, with r fairly small, for example, r = (0, 1, 2), as the variance of the resulting estimator increases with r.

The bandwidths of the N component LP estimators are $m_i = \alpha_i m$, $i = 1, \ldots, N$, for an increasing sequence of α_i, where $[\alpha_N m] \leq \text{floor}[n/2]$. These bandwidths, $\alpha = (\alpha_1, \ldots, \alpha_N)'$, are used to obtain the N estimates $\hat{d}_{LP}(0, m_i)$, which are components in the weighted estimator with corresponding weights $w_{(1)} = (w_1, \ldots, w_N)'$. The bandwidths and weights satisfy the following conditions (see GS, 2004, 2006):

C1) $\sum_{i=1}^{N} w_i = 1$; C2) $\sum_{i=1}^{N} w_i \alpha_i^{2j} = 0$, for $j = 1, \ldots, r$; C3) $\sum_{i=1}^{N} w_i \alpha_i^{2+2r} = \delta$.

C1 ensures that the weighted LP estimator is asymptotically unbiased; C2 follows from eliminating the r-th order bias involving terms in $(m/n)^{2r}$; and C3 controls the

multiplicative constant of the dominant bias term, $(m/n)^{2+2r}$, enabling $\hat{d}_{LP}(r)$ and $\hat{d}_{WLP}(r)$ to have the same term, (δ = [1, –0.429, 0.1515] for r = [0, 1, 2]); for further details see GS (2004, 2006, Theorem 2, Lemma 3.1, and remark c, therein). Condition 3 also provides the basis of obtaining an estimator of b_{2+2r}, see equation (14) below, as an input to the 'plug-in' selection of m.

These three conditions are satisfied by the following regression:

$$\hat{d}_{LP}(0, m_i) = d + \sum_{j=1}^{r} \beta_{2j} \alpha_i^{2j} + \beta_{2+2r}(\alpha_i^{2j} - \delta) + u_j \quad i = 1, \ldots, N. \qquad (7)$$

The WLP estimator of $(d, \beta')'$, where $\beta = (\beta_2, ..., \beta_{2r}, \beta_{2+2r})'$, is obtained by an application of GLS to (7), resulting in:

$$(\hat{d}_{WLP}(r), \hat{\beta}')' = W\hat{d}_{LP}(0, \alpha) \qquad (8)$$

where $W = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}$.

The notation is as follows:
$\hat{d}_{LP}(0, \alpha) = (\hat{d}_{LP}(0, m_1), \ldots, \hat{d}_{LP}(0, m_N))'$ is the vector of LP estimators for frequency bands $\lambda_j, j \in (1,..., m_1; m_1 + 1, ..., m_2; ...., ; m_{N-1} + 1, ..., m_N)$;
$Z_i = (1, \alpha_i^2, ..., \alpha_i^{2r}, \alpha_i^{2+2r} - \delta)', Z = (Z_1', \ldots, Z_N')'$;
$\Omega_{ij} = 1/\max(\alpha_i, \alpha_j)$;
$w_{(i)}'$ is the i-th row of W.

The WLP estimator of d of order r is a weighted combination of the N $\hat{d}_{LP}(0, m_i)$ estimators, conveniently written as follows:

$$\hat{d}_{WLP}(r) = w_{(1)}'\hat{d}_{LP}(0, \alpha) \qquad (9)$$

The relevant properties of $\hat{d}_{WLP}(r)$ are, see GS (2004, 2006):

$$AB(\hat{d}_{WLP}(r)) = \tau b_{2+2r} \frac{m^{2+2r}}{n^{2+2r}} \qquad (10)$$

$$\sigma^2(\hat{d}_{WLP}(r)) = \frac{\pi^2}{24m} c_r^* \qquad (11)$$

$$m_{WLP}^{opt}(r) = K_{WLP}(r) n^{(4+4r)/(5+4r)}. \qquad (12a)$$

$$K_{WLP}(r) = \left(\frac{\pi^2 c_r^*}{24(4+4r)\tau^2 b_{2+2r}^2} \right)^{1/(5+4r)}. \qquad (12b)$$

The parameters τ and c_r^* in (10) and (11) govern the asymptotic properties of the estimator and relate to the corresponding parameters in AG (2003). The variance $\sigma^2(\hat{d}_{WLP}(r))$ is as $\sigma^2(\hat{d}_{LP}(r))$, but with c_r replaced by c_r^*; hence, for $\sigma^2(\hat{d}_{WLP}(r))$ to dominate $\sigma^2(\hat{d}_{LP}(r))$, it must be the case that $c_r^* < c_r$ for given m. GS show that it is possible to choose α such that $c_r^* < c_r$. For example, when $r = 1, \tau = \tau_1$ and $\alpha =$ (0.5, 0.55, …, 2.0) then $c_1^* = 1.38$, which compared to $c_1 = 2.25$ of $\hat{d}_{LP}(1)$ represents a variance reduction of approximately 39%; whilst c_1^* can be further reduced, for example choosing $\alpha_N = 3$ results in $c_1^* = 1$, larger values of α_N imply that the underlying LP regressions use more frequencies, which will increase the bias although they reduce the variance (see GS, 2004, Remark 6).

The HD equivalent method for selecting m by estimating the rmse-driven optimal value, but for the WLP estimator, is obtained as follows. The unknown coefficient in (12b) is b_{2+2r}, which can be estimated by:

$$\hat{b}_{2+2r} = \frac{\hat{\beta}_{2+2r}}{\tau_r} \frac{n^{2+2r}}{\bar{m}^{2+2r}} \qquad (13)$$

see GS (2004), where $\bar{m} = Kn^{4(r+2)/(4(r+2)+1)}$. N auxiliary regressions are run first using $\alpha_i \bar{m}$ frequencies to obtain, say, $w_{(r+2)}(\bar{m})'$ and $\hat{d}_{LP}(0, \alpha, \bar{m})$; $\hat{\beta}_{2+2r}$ in (13) is then the weighted combination $\hat{\beta}_{2+2r} = w_{(r+2)}(\bar{m})'\hat{d}_{LP}(0, \alpha, \bar{m})$. The resulting estimator $\hat{b}_{2+2r}$ is used to form $\hat{m}_{WLP}^{opt}(r)$, see (12a). As in standard LP estimation, K is a user-chosen constant and, from previous work, we considered $K \in [0.1, 0.2, 0.3, 0.4]$, finding the choice to lie between K = 0.1 and K = 0.2; the former is favoured for a selection motivated by minimising the bias, whereas the latter is preferred for minimising the rmse and is used in the simulations reported below. Although dependent on the user-chosen value of K, this method is referred to as 'plug-in'. The α sequence was (as in GS, 2004) $\alpha = (\alpha_1, \alpha_1 + \Delta\alpha, \ldots, \alpha_N)$, $\alpha_1 = 0.5$, $\alpha_N = 2$, $\Delta\alpha = 0.05$, which implies that $c_1^* = 1.38$, (see ibid., Table 1).

4. Bootstrapping the LP and WLP estimators

Bootstrapping the LP and WLP estimators can take place in either the frequency domain (FD) or the time domain (TD) and we consider each in turn. A residual-based frequency domain bootstrap was suggested for the standard LP estimator by AO (2005, 2009a, b), whereas guidance for the time domain bootstrap is provided by Franco and Reisen (2004, 2007) and Kapetanios and Papailias, hereafter KP, (2011). For related work see also Bisaglia and Procidano (2003), Bisaglia et al (2008), Silva et al., (2006), Murphy and Izzeldin (2009) and Buhlmann (2002), and for a theoretical justification for the use of the bootstrap for long-memory processes, with further simulation results, see Poskitt (2007) and Kapetanios (2010).

In the context of the standard LP regression, AO (2009b) suggested a bootstrap method to choose the value of m, noting that, and as confirmed in our preliminary simulations, the key to good performance of the LP estimator is in choosing the value of m. We extend the AO procedure to the WLP case and to the time domain. The general bootstrap procedure involves an iterative scheme to choose the value of m that minimizes the bootstrap rmse, the principle of which is as follows. Let $(m_1,\ m_2) \in M$, where m_1 and m_2 are the minimum and maximum values of m for each iteration; $m_2 \leq \text{floor}[n/2]$ for the LP procedure and $\alpha_N m_2 \leq \text{floor}[n/2]$ for the WLP procedure. $m^{(1)}$ is the initial value of m that starts the process, where $m_1 \leq m^{(1)} < m_2$ and, although not critical, a sensible choice of $m^{(1)}$ errs on the side of a relatively small value of m, so as to control the bias.

The design of the algorithm is to start with $m^{(1)}$, and project the regression based on $m^{(1)}$ frequencies to obtain a set of residuals extended to include all m_2 frequencies, from which to resample and construct the bootstrap data. The initial value of m, $m^{(1)}$, is then sequentially updated through an iterative procedure until convergence is achieved based on minimizing the bootstrap rmse. We first describe 'blind' and 'local' bootstraps for the LP and AG estimators and, subsequently, a 'sighted' bootstrap for the WLP estimator; we then consider variations arising from bootstrapping in the time domain.

4.1. The LP bootstrap algorithm

4.1.1. Frequency domain

i). Estimate the (initial) LP regression $Y_j = \beta_0 + dX_j + \varepsilon_j$, where $Y_j \equiv \ln[I_y(\lambda_j)]$ and $X_j \equiv -2\ln(\lambda_j)$; the estimation of this regression requires choosing an initial value of m, $m^{(1)}$, which will be updated through the looping process and referred to in general as $m^{(i)}$. This choice of $m^{(1)}$ is not critical as it is updated in the algorithm. We

start the LP algorithm with $m^{(1)} = n^{0.5}$. Let $\hat{\beta}_0^{(1)}$ and $\hat{d}_{LP}^{(1)}$ denote the (initial) estimates from the LP regression. In general let $\hat{d}_{LP}^{(i)}$ denote the estimate of $\hat{d}_{LP}$ in the i-th iterative round.

ii). Generate the residuals from the initial regression with $m = m^{(1)}$, that is $\hat{e}_j = Y_j - \hat{Y}_j$, where $\hat{Y}_j^{(1)} = \hat{\beta}_0^{(1)} + \hat{d}_{LP}^{(1)} X_j$, $j = 1, \ldots, m^{(1)}$, and then project these over frequencies, $j = m^{(1)} + 1, \ldots, m_2$; denote the complete set of residuals as $\hat{\varepsilon}^{(i)} = \{\hat{\varepsilon}_j\}_{j=1}^{m_2}$, where i = 1 for the initial round.

iii). For a fixed value of $m \in M$, construct the bootstrap replicates (in the frequency domain) of Y_j, that is $Y_j^b = \hat{Y}_j^{(1)} + e_j^{(1)}$, $j = 1, \ldots, m$, where, in the 'blind' bootstrap, the $e_j^{(1)}$ are resamples from anywhere in $\hat{\varepsilon}^{(1)}$, with equal probability.

iv). For each value of $b = 1, \ldots, B$, estimate $\hat{d}_{LP}$, denoted $\hat{d}_{LP}^{b,m}$; then loop b over 1, …, B, and obtain the rmse of $\{\hat{d}_{LP}^{b,m}\}_{b=1}^{B}$ relative to the starting value $\hat{d}_{LP}^{(i)}$.

v). Repeat steps iii) and iv) for all $m \in M$.

vi). Determine the m that minimizes the rmse for $m \in M$; denote this value of m as $m_{min}^{(i)}$, with associated $rmse_{min}^{(i)}$; this is the value of m with the minimum rmse on the i-th iteration conditional on the i-th starting value of m, $m^{(i)}$.

vii). Use $m_{min}^{(i)}$ as the initial value in the next iterative round and repeat steps i) to vi).

viii). Stop the process when:

$$\text{abs}\left(\frac{rmse_{min}^{(i+1)} - rmse_{min}^{(i)}}{rmse_{min}^{(i)}} \right) < \delta,$$

for some small value δ, taken to be 0.01 in the simulations reported below. The process is also stopped if the upper bound, in this case $\text{floor}[n/2]$, is reached. The selected value of m_{min} is then $m_{min}^{(i)}$, corresponding to the overall minimum estimate of the rmse minimizing value of m, which is used for LP estimation.

viii). Using $m_{min}^{(i)}$, estimate d_{LP}, with the estimate denoted $\hat{d}_{LP}(m)$, $m = m_{min}^{(i)}$.

ix). Construct the percentile-t confidence interval[1] (CI):

$$CI(1-\alpha) = [\hat{d}_{LP}(m) \pm 1.96\hat{\sigma}(\hat{d}_{LP}(m))]$$

where $\hat{\sigma}(\hat{d}_{LP}(m))$ is the LS estimator of $\sigma(\hat{d}_{LP}(m))$ from the LP regression using m = $m_{min}^{(i)}$.

4.1.2. Time domain

An alternative to the frequency domain bootstrap is a time domain bootstrap, which has the following motivation: by hypothesis $y_t \sim I(d)$, hence, $\Delta^d y_t = u_t \sim I(0)$, that is y_t is integrated of order d, so that its d-th difference is weakly dependent; for convenience of notation define the d-th difference of y_t by $x_t \equiv \Delta^d y_t$. The random variable x_t is estimated by replacing the unknown d by an estimate $\hat{d}$, and then constructing $\hat{x}_t \equiv \Delta^{\hat{d}} y_t$ for t = 1 , …, T, using the binomial expansion of $(1 - L)^{\hat{d}}$ applied to y_t, setting pre-sample values of y_t to zero.

In the TD bootstrap, resampling is from the sequence of x_t, that is $x = \{\hat{x}_t\}_{t=1}^{T}$; however, note that $u_t \sim I(0)$ and x is not, therefore, necessarily an i.i.d sequence. For example, suppose that y_t is generated by the ARFIMA(1, d, 0) process, $(1 - \phi_1 L)\Delta^d y_t = \varepsilon_t$, where ε_t is weak white noise and $|\phi_1| < 1$, then $u_t \sim I(0)$, where $u_t \equiv (1 - \phi_1 L)^{-1}\varepsilon_t$. Thus, in principle, a 'blind' bootstrap is not appropriate because of the correlation between u_t and u_s for $t \neq s$; the same line of argument applies to the application of a 'blind' bootstrap in the frequency domain, but in that case to ε_i and ε_j, $i \neq j$. There are a number of bootstrap designs based on restricting the domain from which the resampled values are taken and we consider one of these below. In the present context, the adopted time domain bootstrap is as follows.

1). As in the frequency-based bootstrap, estimate the (initial) LP regression $Y_j = \beta_0 + dX_j + \varepsilon_j$ for each m; as before $\hat{d}_{LP}^{(1)}$ denotes the LP estimate of d from this regression.

2). Construct $\hat{x}_t$ based on $\hat{d}_{LP}^{(1)}$, where: $\hat{x}_t \equiv (1 - L)^{\hat{d}_{LP}^{(1)}} y_t$.

3). For each m ∈ M, construct a bootstrap sample $\hat{x}^b = \{\hat{x}_t^b\}_1^T$ by resampling $\hat{x}_t^b$ from $\hat{x} = \{\hat{x}_t\}_1^T$. This is a 'blind' bootstrap if the resampling is from any element of $\hat{x}$, whereas, a local bootstrap restricts the resampling to be from some subset of $\hat{x}$.

4). Estimate d from the LP regression applied to each bootstrap sample and obtain the rmse of the resulting bootstrap estimate relative to $\hat{d}_{LP}^{(1)}$.

5). As in the FD bootstrap, update the starting value of m and continue the iterative scheme until convergence is achieved.

Justification for the validity of this time-domain bootstrap is provided in KP (2011), which is based on weak conditions on the estimator used in step 1). Both the FD and TD bootstraps can be combined with a method of local resampling to reduce the dependence in the frequency regression residuals underlying the FD resampling scheme or in the reconstructed time series observations for the TD bootstrap.

The central feature of such schemes is to reduce the dependence by resampling from a subset of $\{\hat{\varepsilon}_j\}_{j=1}^{m}$ in the FD bootstrap or $\{\hat{x}_t\}_{t=1}^{T}$ in the TD bootstrap, usually by using a 'window' that moves through the sample – see, for example, Politis and White (2004)) and KP (2011). The particular form used here is a moving window bootstrap, where resampling is from a centered window; for example in the case of the TD bootstrap, $\hat{x}_t^b$ is a resample, with equal probability, from the 'local' window $(\hat{x}_{t-s}, \ldots, \hat{x}_{t-1}, \hat{x}_t, \hat{x}_{t+1}, \ldots, \hat{x}_{t+s})$, $t - s \geq 1$ and $t + s \leq T$. The selected indexes are constrained so as not to fall outside the boundaries given by the beginning and end of the sample.

The TD procedure requires selection of the resampling width, k = 2s + 1. The theoretical justification of Paparoditis and Politis (1999), and the Monte-Carlo experiments of Silva et al (2006), suggest that although in principle s should depend on the weak dependency in the process, good results are obtained from small s. A similar local resampling scheme can also be used in the frequency domain with resampling from a moving window of frequencies centred on $\hat{\varepsilon}_j$.

4.2. AG bootstrap: frequency and time domains

The FD bootstrap algorithm for the AG estimator $\hat{d}_{LP}(1)$, is a straightforward development of that for the simple LP estimator, where the initial regression becomes $Y_j = \beta_0 + dX_j + \varphi_1\lambda_j^2 + \varepsilon_j$, with the residuals and the bootstrap replicates defined accordingly. The TD bootstrap algorithm reconstructs data in the time dimension using $\hat{d}_{LP}(1)$ rather than $\hat{d}_{LP}(0)$.

4.3. The WLP bootstrap algorithm

4.3.1. Frequency domain

As in the LP case, $(m_1, m_2) \in M$, but here $\alpha_N m_2 \leq \text{int}[n/2]$. The algorithm is outlined for the case in which the N component estimators are $\hat{d}_{LP}(0)$. Note that the proposed bootstrap is not 'blind', but uses resampling of a local form; in particular it is referred to here as a locally sequenced or 'sighted' bootstrap, in the sense that the

resampling respects the nature of the increasing frequency subsets that are used in obtaining the component LP estimators.

i). $m^{(1)}$ is again the initial value of m that starts the process, $m_1 \leq m^{(1)} < m_2$; in this case $m^{(1)}$ serves to determine the set of N frequency bands: $M^{(1)} = \{m_h^{(1)}\}_{h=1}^{N}$ where $m_h^{(1)} = (1, \ldots, \alpha_h m^{(1)})$, h = 1, …, N; these are the bands for estimation of the N component $\hat{d}_{LP}(0)$ estimators. $m^{(1)}$ is updated through the iterative process and, in general, is referred to as $m^{(i)}$.

ii). Obtain the initial estimate $\hat{d}_{WLP}^{(1)}$, as in (8), from the N bands in $M^{(1)}$. This estimate will be used to evaluate the rmse in the bootstrap loop.

ii). Consider the first round of the iteration with the first value of $m \in M$ to be considered, that is m_1, and the first set of frequencies, $m_1^{(1)}$. Estimate the LP regression with frequencies $(1, \ldots, \alpha_1 m^{(1)})$; project the residuals from this regression over the full set of frequencies in *M*, that is $(1, \ldots, \alpha_1 m_2)$; next increment the α_j index by 1 to consider $\alpha_2 m^{(1)}$ and estimate the LP regression with frequencies $(1, \ldots, \alpha_2 m^{(1)})$, again project this regression, but now over $(1, \ldots, \alpha_2 m_2)$; continue until h = N. In general, this gives the extended set of fitted values $\hat{Y}^{(1)} = \{\hat{Y}_j^{(1)}\}_{j=1}^{\alpha_h m_2}$, h = 1, …, N, and residuals $\hat{\varepsilon}_h^{(1)} = \{\hat{\varepsilon}_j^{(1)}\}_{j=1}^{\alpha_h m_2}$, h = 1, …, N, from which resampling will take place through the α sequence. The sets of residuals so obtained will form the sets from which resampling will take place in a sequence that reflects the marginal addition to the fitted values and residuals as h increases. (The initial value of m, $m^{(1)}$, will be updated to $m^{(i)}$ through the iterative loop described in step v.)

iii)a. The bootstrap is a loop based on respecting the sequential nature of the increasing bandwidths from which the N component LP estimates are obtained; specifically, resampling is from within the N marginal subsets of residuals $\hat{\varepsilon}_{(h)}^{(1)}$ for the h-th band.

iii)a. For the first value m_1 of $m \in M$, construct the set of bootstrap values $Y_{(h)}^{b}$, h = 1, that is: $Y_{(1)}^{b} = \{\hat{Y}_j^{b}\}_{j=1}^{\alpha_1 m_1}$, where $\hat{Y}_j^{b} = \hat{Y}_j^{(1)} + e_j^{(1)}$, j = 1, …, $\alpha_1 m_1$, and $e_j^{(1)}$ is a random draw with replacement from $\hat{\varepsilon}_{(1)}^{(1)}$. Note that this a form of local resampling as an alternative would be to take all of the resamples from the LP regression with $\alpha_N m^{(1)}$ frequencies for all α_i.

iii)b. Obtain the bootstrap LP estimate, denoted $\hat{d}_{LP,1}^{b}$, for this set of frequencies.

iii)c. For h = 2, extend $Y^{b}_{(1)}$ by (only) adding the marginal frequencies: $\hat{Y}^{b}_{j} = \hat{Y}^{(1)}_{j} + e^{(1)}_{j}$, $j = \alpha_1 m_1 + 1, \ldots, \alpha_2 m_1$, where $e^{(1)}_{j}$ is a random draw from $\hat{\varepsilon}^{(1)}_{(2)}$ over the same (marginal) frequency range; next obtain the LP estimate $\hat{d}^{b}_{LP,2}$ for the frequencies 1, …, $\alpha_2 m_1$.

iii)d. Continue the sequence of adding the marginal sub-sets of bootstrapped observations with draws from the frequencies in the marginal bandwidths. This sequence results in the N component estimates of $\hat{d}^{b}_{LP,h}$, h = 1, …, N, which are weighted into the bootstrap WLP estimate $\hat{d}^{b}_{WLP}$, as in (8).

iii)e. Repeat this procedure B times with initial value $m^{(1)}$ and, hence, obtain the set of bootstrap estimates: $\hat{d}^{bs,m}_{WLP} = \{\hat{d}^{b}_{WLP}\}^{B}_{b=1}$, where $m = m_1$.

iv). Obtain the rmse of $\hat{d}^{b,m}_{WLP}$ relative to $\hat{d}^{(1)}_{WLP}$.

v). Repeat steps iii) and iv) for all $m \in M$ (note that $m^{(1)}$ is fixed for this looping).

The next steps, vi) to viii), are as in the case of the simple LP estimator, with the minor modification that the upper bound is $\text{floor}[n/2\alpha_N]$. In summary, given $m^{(i)}$, the two loops ($m = m_1, \ldots, m_2$ and b = 1, …, B), serve to determine the value of m that results in a minimum over m for the rmse of the bootstrapped WLP estimator; this minimising value of m then becomes the new starting value, $m^{(i+1)}$, for the loops and the process is stopped when the difference $m^{(i+1)} - m^{(i)}$ is small in some well-defined sense. The overall minimizing value of m, $m^{(i)}_{min}$, is the rmse minimized value that is used in the ultimate WLP estimator, $\hat{d}_{WLP}(m)$, $m = m^{(i)}_{min}$ and in constructing the confidence intervals.

ix). The 95% CI based on asymptotic considerations is:

$$CI(1-\alpha) = [\hat{d}_{WLP}(m) \pm 1.96\sigma^2(\hat{d}_{WLP}(m))],$$

where $\sigma^2(\hat{d}_{WLP}(m))$ is given by (9) with $m = m^{(i)}_{min}$.

Note that the bootstrap mimics the data generation process in respect of the generation of the values of Y_j and the included frequencies. The WLP estimator is constructed from the N LP estimators, $\hat{d}_{LP,h}$, but note that the observations for the underlying regressions only change at the margins of the frequency bands. Schematically, what is in effect constructed for each value of $m \in M$, is a 3-dimensional array G, of

dimensions N, B and $\alpha_N m$, with typical values denoted h, b and m, respectively; now fix b and obtain the 2-dimensional array over N (rows) and $\alpha_N m$ (columns), and denote this as D, which is a 2-dimensional 'slice' of G along the B axis. The rows of D sequentially repeat the non-zero elements in the preceding rows, the case with N = 5 is illustrated below:

$$\begin{bmatrix}
d_{1,1} & \cdots & d_{1,\alpha_1 m} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
d_{1,1} & \cdots & d_{1,\alpha_1 m} & d_{2,\alpha_1 m+1} & \cdots & d_{2,\alpha_2 m} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
d_{1,1} & \cdots & d_{1,\alpha_1 m} & d_{2,\alpha_1 m+1} & \cdots & d_{2,\alpha_2 m} & d_{3,\alpha_3 m+1} & \cdots & d_{2,\alpha_4 m} & 0 & \cdots & 0 & 0 & \cdots & 0 \\
d_{1,1} & \cdots & d_{1,\alpha_1 m} & d_{2,\alpha_1 m+1} & \cdots & d_{2,\alpha_2 m} & d_{3,\alpha_3 m+1} & \cdots & d_{3,\alpha_4 m} & d_{4,\alpha_4 m+1} & \cdots & d_{4,\alpha_5 m} & 0 & \cdots & 0 \\
d_{1,1} & \cdots & d_{1,\alpha_1 m} & d_{2,\alpha_1 m+1} & \cdots & d_{2,\alpha_2 m} & d_{3,\alpha_3 m+1} & \cdots & d_{3,\alpha_4 m} & d_{4,\alpha_4 m+1} & \cdots & d_{4,\alpha_5 m} & d_{5,\alpha_4 m+1} & \cdots & d_{5,\alpha_5 m}
\end{bmatrix}$$

Note that this structure implicitly gives rise to a form of sequenced local resampling. Consider the case used in the simulations reported below, where $\alpha_1 = 0.5$ and $\alpha_N = 2$ with N = 21 and n = 512. Starting with the square root rule, $m^{(1)} = 22$, $\alpha_1 m^{(1)} = 11$ and $\alpha_N m^{(1)} = 44$, so the component LP estimators are based on the frequencies 1 to11 and then frequencies 12 to 44 in 20 steps, so that apart from the first, the marginal frequencies are either 1 or 2 in number at each margin and, hence, the random draws of the residuals are confined to very narrow bands.

4.3.2. Time domain WLP bootstrap

The TD version of the WLP algorithm is as in the LP case, but with the reconstructed data in the time domain based on $\hat{d}^{(i)}_{WLP}$ rather than $\hat{d}^{(i)}_{LP}$. This algorithm proceeds by constructing a bootstrap sample $\hat{x}^b = \{\hat{x}^b_t\}_1^T$ by resampling $\hat{x}^b_t$ from $\hat{x} = \{\hat{x}_t\}_1^T$, based on $\hat{d}^{(i)}_{WLP}$ (see the description the LP TD bootstrap). As resampling is, in this case, in the time domain, rather than the frequency domain, a local resampling scheme may have value-added and we adopt that used for the LP TD bootstrap.

5. Simulations

5.1. The simulation design

Data for the simulation experiments were generated by the following models:

Model 1. The ARFIMA(1, d, 0) model[2], $(1 - \phi_1 L)\Delta^d y_t = \varepsilon_t$, where $\varepsilon_t \sim$ niid(0, 1), d = 0.4, $\phi_1 \in (0.0, 0.5)$. This model is widely chosen as the basic benchmark model.

Model 2. The ARFIMA(2, d, 0) model, $(1 - \phi_1 L)(1 - \phi_2 L)\Delta^d y_t = \varepsilon_t$, d = 0.4, $\phi_1 = 0.4$ and $\phi_2 = -0.7$, this model has imaginary roots reasonably close to the unit root circle with an approximate period of five time units, Priestley (1981).

The sample size for the reported results is n = 1024 and B = 499, with 500 replications (the results for n = 512 were qualitatively similar)[3]. The estimators to be compared fall into three classes: the (simple) LP estimator; the AG estimator, which is estimated from the LP regression augmented by the squared frequency; the WLP estimator combining the N LP estimators. The infeasible estimators that use the Monte-Carlo rmse optimal value of m are included for comparison. Bootstrapping alternately uses the frequency domain and time domain versions and, where appropriate each version also considers 'blind' and 'local' resampling. Guided by AO (2009) and our own simulations, the local forms of the bootstraps use a centered window of length 5, being 2 either side of the respective frequency for FD or time index for TD.

The LP estimators are:
i). $\hat{d}_{LP}(m_{LP}^{opt})$: the LPE, $\hat{d}_{LP}$, using m = m_{LP}^{opt}, as in (5). This estimator is infeasible, but provides guidance as to what could be achieved.
ii). $\hat{d}_{LP}(\sqrt{n})$: the LPE with m = $\sqrt{n}$, that is the often used GPH square root rule.
iii). $\hat{d}_{LP}(\hat{m}_{LP}^{opt})$: the LPE with the 'plug-in' version of m_{LP}^{opt}, that is $\hat{m}_{LP}^{opt}$.
iv). $\hat{d}_{LP}(m_{min})$: the bootstrapped versions of $\hat{d}_{LP}$, with the rmse minimising value of m.

The estimators in iv) are alternately in frequency domain (FD) and time domain (TD) versions, with 'blind' and 'local' bootstrapped versions.

The WLP estimator comparators are, in each case for r = (0, 1, 2):
ix). $\hat{d}_{WLP}(m_{WLP}^{opt})$: the WLPE using m = m_{WLP}^{opt}, as in (10). Again, whilst infeasible it is a useful guide.
x). $\hat{d}_{WLP}(\sqrt{n})$: the WLPE with m = $\sqrt{n}$.
xi). $\hat{d}_{WLP}(\hat{m}_{WLP}^{opt})$: the WLPE $\hat{d}_{WLP}$, using the 'plug-in' estimate $\hat{m}_{WLP}^{opt}$.
xii). $\hat{d}_{WLP}(m_{min})$: the bootstrapped version of $\hat{d}_{WLP}$ with the rmse minimising value of m, this is based on the 'sequenced' FD bootstrap.

A TD version of the estimator in xii) is also reported based on 'blind' and 'local' forms of the bootstrap and in each case the weighting allows for the r = 0, 1, 2 cases.

5.2. Results and discussion

The results are summarised in Tables 1 – 5 and organised by estimator, with each table reporting simulation means, simulation rmse and simulation coverage rates and widths, which are in each case averages over the number of simulations. There are three rows for the results relating to the fidelity of the confidence intervals, being the overall rejection proportion and the proportions rejected in the left and right-hand tails, respectively; ideally, the first of these should be 95%, with 2.5% in each of the tails. The best result for each estimator is indicated in bold, with the second best indicated in bold italics. The column headed MC refers to the average over the outer Monte-Carlo repetition loop, here 500.

LP

Whilst the $\sqrt{n}$ rule does well in terms of bias, there is a cost in terms of rmse and the wide confidence intervals. We can also rule out TD 'local' for its poor bias, rmse and interval widths. Of the remaining methods TD 'blind' has the best performance in terms of rmse, although it does not dominate in terms of bias; however, it does well overall with the shortest confidence intervals whilst maintaining a relatively good fidelity in its confidence intervals.

AG

Generally, AG does produce a benefit relative to LP/GPH in terms of reducing bias across all models and methods, but it does so at a fairly substantial cost in terms of increasing the width of the confidence interval and is likely to be ruled out practically for that reason. A comparison of methods suggests that $\hat{m}_{LP}^{opt}$ gives the best overall performance with a good performance in terms of rmse, the shortest confidence intervals and very good fidelity of the confidence inervals.

WLP

There are two questions to consider in this case. One relates to an evaluation of the selection of m for a given r and the other to the choice of r. In the case of r = 0, the choice narrows to be between $\hat{m}_{WLP}^{opt}$ and FD 'seq', with the latter favoured given its slightly better performance in terms of bias and coverage. The TD methods tend to be inferior to the FD method, particularly with wide confidence intervals. The same general picture emerges for r = 1 and r = 2. Both TD methods are inferior to FD 'seq', which whilst dominated by $\hat{m}_{WLP}^{opt}$ for rmse and interval widths, dominates $\hat{m}_{WLP}^{opt}$ in terms of bias and confidence interval fidelity.

In comparing different values of r, bias tends to decrease with r, but the rmse and confidence interval widths tend to increase with r, although much less so for $\hat{m}_{WLP}^{opt}$, and

with no particular gain the confidence interval fidelity. Overall, the simulation results favour choosing r = 0 combined with FD 'seq' or, possibly $\hat{m}^{opt}_{WLP}$.

Comparison between methods

Comparing LP with AG, whilst AG has some benefit in terms of reducing bias, the reduction is quite slight and, although there is some improvement in fidelity, this comes at the cost of a substantial increase in the interval widths. Hence, the critical comparison is whether it is better to weight the LP estimators as in the WLP estimator. The general answer is yes, where the benefit comes in substantially shorter interval widths, whether using $\hat{m}^{opt}_{WLP}$ of FD 'seq', with no deterioration in bias.

Table 1 $\hat{d}_{LP}$, n = 1024

	m^{opt}_{LP}	$\sqrt{n}$	$\hat{m}^{opt}_{LP}$	FD: 'blind'	FD: 'local'	TD: 'blind'	TD: 'local'	MC
Mean								
$\phi_1 = 0.0$	0.366	0.415	***0.406***	**0.398**	0.411	0.415	0.334	0.368
$\phi_1 = 0.2$	0.432	***0.416***	***0.416***	**0.400**	***0.416***	0.426	0.310	0.434
$\phi_1 = 0.5$	0.450	**0.426**	0.451	0.437	***0.434***	0.475	0.278	0.456
AR2	0.283	0.409	***0.375***	0.370	**0.393**	0.362	0.344	0.311
Rmse								
$\phi_1 = 0.0$	0.043	0.151	***0.106***	0.119	0.125	**0.085**	0.246	0.042
$\phi_1 = 0.2$	0.064	0.152	***0.110***	0.122	0.123	**0.088**	0.269	0.063
$\phi_1 = 0.5$	0.104	0.152	0.132	***0.129***	0.132	**0.115**	0.283	0.100
AR2	0.130	0.150	***0.130***	0.136	0.134	**0.119**	0.193	0.108
Width								
$\phi_1 = 0.0$	0.114	0.521	***0.319***	0.388	0.352	**0.311**	0.439	0.116
$\phi_1 = 0.2$	0.222	0.522	***0.318***	0.393	0.352	**0.314**	0.500	0.214
$\phi_1 = 0.5$	0.336	0.522	***0.320***	0.390	0.356	**0.312**	0.598	0.317
AR2	0.231	0.521	***0.334***	0.391	0.358	**0.314**	0.365	0.244
Coverage								

$\phi_1 = 0.0$	0.803	***0.923***	0.887	0.917	0.867	**0.943**	0.853	0.843
l-h	0.000	0.037	0.057	0.023	0.057	0.033	0.030	0.000
r-h	0.197	0.040	0.057	0.060	0.077	0.023	0.117	0.157
$\phi_1 = 0.2$	0.903	***0.923***	0.860	0.913	0.863	**0.930**	0.810	0.907
l-h	0.087	0.037	0.090	0.030	0.067	0.047	0.047	0.090
r-h	0.010	0.040	0.050	0.057	0.070	0.023	0.143	0.003
$\phi_1 = 0.5$	0.887	**0.917**	0.757	***0.840***	0.803	0.803	0.820	0.893
l-h	0.103	0.040	0.217	0.120	0.140	0.190	0.047	0.103
r-h	0.010	0.043	0.027	0.040	0.057	0.007	0.133	0.003
AR2	0.473	**0.920**	0.813	0.810	0.810	0.790	***0.870***	0.707
l-h	0.000	0.033	0.027	0.033	0.047	0.017	0.010	0.000
r-h	0.527	0.047	0.160	0.157	0.143	0.193	0.120	0.293

Table 2 AG , $\hat{d}_{LP}(1)$, n = 1024

	m_{LP}^{opt}	$\sqrt{n}$	$\hat{m}_{LP}^{opt}$	FD: 'blind'	FD: 'local'	TD: 'blind'	TD: 'local'	MC
Mean								
$\phi_1 = 0.0$	0.410	0.411	***0.410***	0.415	0.414	***0.410***	**0.406**	0.410
$\phi_1 = 0.2$	0.421	0.411	0.411	***0.405***	0.411	0.407	**0.403**	0.427
$\phi_1 = 0.5$	0.427	***0.411***	0.423	0.414	**0.410**	0.415	0.413	0.443
AR2	0.430	**0.411**	***0.412***	0.427	0.416	0.417	0.416	0.436
Rmse								
$\phi_1 = 0.0$	0.108	0.217	**0.125**	0.169	0.189	***0.148***	0.151	0.083
$\phi_1 = 0.2$	0.063	0.218	**0.125**	0.171	0.193	***0.147***	0.155	0.062
$\phi_1 = 0.5$	0.103	0.219	**0.129**	0.170	0.190	***0.150***	0.154	0.092
AR2	0.088	0.215	**0.132**	0.165	0.187	***0.150***	0.151	0.087
Width								
$\phi_1 = 0.0$	0.390	0.890	**0.401**	0.508	0.555	***0.498***	0.506	0.318
$\phi_1 = 0.2$	0.243	0.892	**0.401**	0.517	0.567	***0.506***	0.513	0.224
$\phi_1 = 0.5$	0.373	0.892	**0.401**	0.511	0.559	***0.500***	0.518	0.319
AR2	0.321	0.889	**0.403**	0.512	0.560	***0.505***	0.506	0.306
Coverage								
$\phi_1 = 0.0$	0.920	0.963	**0.903**	0.887	0.887	0.890	***0.900***	0.947
l-h	0.043	0.013	0.040	0.050	0.047	0.047	0.037	0.023
r-h	0.037	0.023	0.057	0.063	0.067	0.063	0.063	0.030
$\phi_1 = 0.2$	0.950	**0.960**	***0.900***	0.887	0.883	0.883	0.877	0.930
l-h	0.040	0.013	0.043	0.047	0.050	0.053	0.050	0.063
r-h	0.010	0.027	0.057	0.067	0.067	0.063	0.073	0.007
$\phi_1 = 0.5$	0.933	**0.957**	0.867	***0.890***	0.883	0.883	0.880	0.903
l-h	0.043	0.013	0.087	0.053	0.047	0.053	0.053	0.083
r-h	0.023	0.030	0.047	0.057	0.070	0.063	0.067	0.013
AR2	0.927	**0.967**	0.893	0.893	0.887	0.900	***0.903***	0.913
l-h	0.053	0.010	0.050	0.060	0.043	0.047	0.043	0.070
r-h	0.020	0.023	0.057	0.047	0.070	0.053	0.053	0.017

Table 3 $\hat{d}_{WLP}$, r = 0, n = 1024

	m_{WLP}^{opt}	$\sqrt{n}$	$\hat{m}_{WLP}^{opt}$	FD: 'seq'	TD: 'blind'	TD: 'local'	MC
Mean							
$\phi_1 = 0.0$	0.397	0.414	**0.406**	**0.409**	0.437	0.425	0.397
$\phi_1 = 0.2$	0.420	**0.415**	0.433	**0.417**	0.441	0.422	0.433
$\phi_1 = 0.5$	0.422	**0.425**	0.483	0.457	0.461	***0.434***	0.454
AR2	0.404	***0.408***	**0.397**	0.382	0.421	0.425	0.365
Rmse							
$\phi_1 = 0.0$	0.035	0.137	**0.058**	***0.079***	0.129	0.155	0.035
$\phi_1 = 0.2$	0.070	0.138	**0.071**	***0.081***	0.142	0.149	0.060
$\phi_1 = 0.5$	0.163	0.138	***0.115***	**0.096**	0.139	0.148	0.096
AR2	0.128	0.137	***0.086***	**0.081**	0.127	0.127	0.074
Width							
$\phi_1 = 0.0$	0.145	0.410	**0.189**	***0.275***	0.371	0.409	0.145
$\phi_1 = 0.2$	0.245	0.410	**0.194**	***0.276***	0.380	0.406	0.197
$\phi_1 = 0.5$	0.484	0.410	**0.234**	***0.276***	0.368	0.404	0.279
AR2	0.387	0.410	**0.286**	***0.276***	0.384	0.380	0.242
Coverage							
$\phi_1 = 0.0$	0.967	0.850	***0.910***	**0.913**	0.873	0.867	0.967
l-h	0.013	0.073	0.057	0.050	0.083	0.063	0.013
r-h	0.020	0.077	0.033	0.037	0.043	0.070	0.020
$\phi_1 = 0.2$	0.933	0.857	0.787	**0.917**	***0.863***	0.850	0.910
l-h	0.050	0.070	0.193	0.057	0.100	0.087	0.087
r-h	0.017	0.073	0.020	0.027	0.037	0.063	0.003
$\phi_1 = 0.5$	0.853	**0.863**	0.663	***0.853***	**0.863**	0.847	0.860
l-h	0.077	0.077	0.330	0.143	0.107	0.097	0.133
r-h	0.070	0.060	0.007	0.003	0.030	0.057	0.007
AR2	0.850	0.850	***0.883***	**0.897**	**0.897**	0.867	0.890
l-h	0.060	0.067	0.030	0.030	0.050	0.073	0.003
r-h	0.090	0.083	0.087	0.073	0.053	0.060	0.107

Table 4 $\hat{d}_{WLP}$, r = 1, n = 1024

	m_{WLP}^{opt}	$\sqrt{n}$	$\hat{m}_{WLP}^{opt}$	FD: 'seq'	TD: 'blind'	TD: 'local'	MC
Mean							
$\phi_1 = 0.0$	0.409	0.414	**0.409**	***0.411***	0.463	0.434	0.409
$\phi_1 = 0.2$	0.429	0.414	0.420	**0.411**	0.456	0.419	0.428
$\phi_1 = 0.5$	0.443	**0.417**	0.494	***0.419***	0.449	0.421	0.447
AR2	0.458	***0.414***	0.324	**0.409**	0.457	0.441	0.392
Rmse							
$\phi_1 = 0.0$	0.053	0.191	**0.061**	***0.124***	0.219	0.225	0.053
$\phi_1 = 0.2$	0.060	0.192	**0.064**	***0.124***	0.215	0.215	0.060
$\phi_1 = 0.5$	0.091	0.191	**0.115**	***0.125***	0.192	0.201	0.091
AR2	0.090	0.190	0.202	**0.123**	0.183	0.210	0.066
Width							
$\phi_1 = 0.0$	0.204	0.578	**0.226**	***0.389***	0.540	0.557	0.204
$\phi_1 = 0.2$	0.204	0.578	**0.226**	***0.389***	0.530	0.547	0.205
$\phi_1 = 0.5$	0.293	0.578	**0.233**	***0.389***	0.517	0.559	0.284
AR2	0.258	0.578	**0.243**	***0.388***	0.531	0.545	0.249
Coverage							
$\phi_1 = 0.0$	0.950	0.867	**0.940**	***0.883***	0.850	0.820	0.950
l-h	0.027	0.063	0.023	0.053	0.100	0.097	0.027
r-h	0.023	0.070	0.037	0.063	0.050	0.083	0.023
$\phi_1 = 0.2$	0.917	0.857	**0.937**	***0.890***	0.837	0.827	0.917
l-h	0.077	0.070	0.040	0.050	0.107	0.083	0.077
r-h	0.007	0.073	0.023	0.060	0.057	0.090	0.007
$\phi_1 = 0.5$	0.890	***0.860***	0.640	**0.873**	0.843	0.850	0.883
l-h	0.100	0.073	0.357	0.070	0.103	0.093	0.113
r-h	0.010	0.067	0.003	0.057	0.053	0.057	0.003
AR2	0.837	***0.877***	0.550	**0.883**	0.860	0.837	0.953
l-h	0.160	0.057	0.123	0.053	0.087	0.103	0.007
r-h	0.003	0.067	0.327	0.063	0.053	0.060	0.040

Table 5 $\hat{d}_{WLP}$, r = 2, n = 1024

	m_{WLP}^{opt}	$\sqrt{n}$	$\hat{m}_{WLP}^{opt}$	FD: 'seq'	TD: 'blind'	TD: 'local'	MC
Mean							
$\phi_1 = 0.0$	0.408	0.418	**0.408**	***0.414***	0.467	0.438	0.409
$\phi_1 = 0.2$	0.415	0.418	***0.413***	**0.412**	0.458	0.431	0.414
$\phi_1 = 0.5$	0.485	0.419	0.475	**0.415**	0.483	***0.416***	0.445
AR2	0.262	***0.420***	0.515	**0.414**	0.466	0.441	0.401
Rmse							
$\phi_1 = 0.0$	0.066	0.227	**0.068**	***0.159***	0.252	0.244	0.065
$\phi_1 = 0.2$	0.068	0.229	**0.069**	***0.161***	0.257	0.270	0.067
$\phi_1 = 0.5$	0.107	0.229	**0.101**	***0.158***	0.269	0.237	0.089
AR2	0.153	0.225	***0.150***	**0.160**	0.283	0.276	0.068
Width							
$\phi_1 = 0.0$	0.246	0.695	**0.254**	***0.466***	0.639	0.644	0.246
$\phi_1 = 0.2$	0.246	0.695	**0.254**	***0.466***	0.641	0.665	0.249
$\phi_1 = 0.5$	0.246	0.695	**0.255**	***0.466***	0.637	0.668	0.291
AR2	0.246	0.695	**0.276**	***0.466***	0.638	0.655	0.259
Coverage							
$\phi_1 = 0.0$	0.923	***0.870***	**0.930**	0.863	0.847	0.833	0.933
l-h	0.040	0.073	0.030	0.060	0.090	0.093	0.037
r-h	0.037	0.057	0.040	0.077	0.063	0.073	0.030
$\phi_1 = 0.2$	0.937	0.867	**0.940**	***0.870***	0.867	0.847	0.920
l-h	0.040	0.073	0.033	0.057	0.087	0.087	0.047
r-h	0.023	0.060	0.027	0.073	0.047	0.067	0.033
$\phi_1 = 0.5$	0.700	**0.867**	0.750	***0.863***	0.817	**0.867**	0.893
l-h	0.300	0.073	0.250	0.067	0.130	0.063	0.093
r-h	0.000	0.060	0.000	0.070	0.053	0.070	0.013
AR2	0.437	**0.880**	0.543	***0.837***	0.833	0.833	0.940
l-h	0.000	0.073	0.440	0.080	0.110	0.087	0.023
r-h	0.563	0.047	0.017	0.083	0.057	0.080	0.037

6. Concluding remarks

A number of semiparametric methods have been suggested for the estimation of the long-memory parameter based on the log-periodogram regression (LPR), all of which have in common the need to truncate the range of frequencies included in the LPR. Bootstrap selection methods have become popular of late based on obtaining the bootstrap analogue of the optimal number of included frequencies. Such methods can be based on a frequency domain (FD) approach or alternately a time domain (TD) approach. In this paper we have evaluated the three most popular LP-based methods combined with a range of possible methods to choose the number of included frequencies; specifically the popular $\sqrt{n}$ rule, the estimated optimum and FD and TD bootstraps in 'blind', 'local' and, where appropriate, 'sequential' forms. The latter refers to WLP estimator, which is based on increasing the number of included frequencies, where care has to be taken in designing the bootstrap so as to respect the sequential structure of the component LP regressions.

The title of this paper poses a question, namely is it better to weight the LP estimators, as in the WLP estimator, than use the unweighted LP estimator? On the basis of the simulation evidence, the answer is yes using r = 0, with notable shorter confidence interval widths and noting that higher values of r do not generally improve the performance of the WLP estimator or, indeed, the LP estimator, as in the AG case. The 'sequenced' frequency domain version of the WLP works well in this context, but the computationally simpler WLP estimator using $\hat{m}_{WLP}^{opt}$ is a 'cheap' alternative, although it loses out on confidence interval fidelity as the serial correlation (ϕ_1) in the noise increases.

Acknowledgements

The research was partially supported by a financial grant from the British Academy. The authors are grateful to Yixiao Sun for comments, access to unpublished work and a number of MATLAB programs. We are grateful to the editor and two referees for constructive comments.

References

Agiakloglou, C., Newbold, P., Wohar, M., (1993), Bias in an estimator of the fractional difference parameter, Journal of Time Series Analysis 14, 235-246.

Andrews, D.W.K., Guggenberger, P., (2003), A bias-reduced log-periodogram regression estimator for the long-memory parameter, Econometrica, 71, 675-712.

Arteche, J., Orbe, J., (2005), Bootstrapping the log-periodogram regression, Economics Letters, 86, 79-85.

Arteche, J., Orbe, J., (2009a), Small sample bootstrap confidence intervals for long-memory parameter, International Journal of Human and Social Sciences, 4, 1113-1115.

Arteche, J., Orbe, J., (2009b), Bootstrap-based bandwidth choice for log-periodgram regression, Journal of Time Series Analysis, 30, 591-617.

Bisaglia, L., Procidano, I., (2003), Improving the power of unit root tests against fractional alternatives using Bootstrap, Statistica, 13(3), 589-602.

Bisaglia, L., Bordignon, S., Cecchinato, N., (2008), Bootstrap approaches for estimation and confidence intervals of long memory processes, Working Paper Series, Department of Statistical Scinces, University of Padua.

Buhlmann, P., (2002), Bootstrap for Time Series, Statistical Science, 17(1), 52–72

Davidson, J., Hashimzade, N., (2009), Type I and type II fractional Brownian motions: a reconsideration, Computational Statistics and Data Analysis, 53, 2089-2106.

Franco, G.C., Reiesen, V.A., (2004), Bootstrap techniques in semiparametric estimation methods for ARFIMA models: A comparison study, Computational Statistics, 19(2), pp 244-259.

Franco, G.C., Reiesen, V.A., (2007), Bootstrap Approaches and Confidence Intervals for Stationary and Nonstationary Long-Range Dependence Processes, Physica A, 375(2), 546-562.

Geweke, J., Porter-Hudak, S., (1983), The estimation and application of long memory time series models, Journal of Time Series Analysis, 4, 221-238.

Guggenberger, P., Sun, Y., (2006), Bias-reduced log-periodogram and Whittle estimation of the long-memory parameter without variance inflation, Econometric Theory, 22, 863-912.

Hurvich, C.M., 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, Journal of Time Series Analysis, 20, 331-341.

Kapetanios, G., Papailias, F., (2011), Block Bootstrap and Long Memory, Queen Mary, University of London, Working paper No. 679.

Kapetanios, G., (2010), Generalization of a Sieve Bootstrap Invariance Principle to Long Memory Processes, Quantitative and Qualitative Analysis in Social Sciences, 4(1), 19-40.

Murphy, A., Izzeldin, M., (2009), Boostapping long memory tests: some Monte Carlo Results. Computational Statistics and Data Analysis, 53(6), 2335-2369

Paparoditis, E., Politis, D.N., (1999), The local bootstrap for periodogram statistics, Journal of Time Series Analysis, 20(2), 193-222.

Paparoditis, E.W., Politis, D.N., (2001), Tapered Block Bootstrap, Biometrica, 88(4), 1105-1119.

Paparoditis, E.W., Politis, D.N., (2003), Residual-based Block Bootstrap for Unit Root Testing, Econometrica, 71(3), 813-855.

Polotis, D.N., White, H. (2004), Automatic Block – Length Selection for the Dependent Bootstrap., Econometric Reviews, 23(1), 53-70

Poskitt, D., (2008), Properties of the Sieve Bootstrap for Fractionally Integrated and Non-Invertible Processes, Journal of Time Series Analysis, 29(2), 224-250.

Priestley, M.B., (1981), Spectral analysis and time series, Academic Press.

Robinson, P.M., (1995), Log-periodogram regression of time series with long range dependence, The Annals of Statistics 23, 1048-72.

Silva, E., Franco, G.C., Reisen, V.A., and Cruz, F.R.B., (2006), Local Bootstrap Approaches for Fractional Differential Parameter Estimation in ARFIMA Models, Computational Statistics and Data Analysis, 51(2), 1002-1011.

---

[1]The percentile-t intervals were superior to the corresponding percentile confidence intervals. Further, once the rmse minimizing value of m was used, there was no systematic advantage to using the percentiles from the distribution of the bootstrap 't' statistic: $t^{b,m}_{LP} = (\hat{d}^{b,m}_{LP} - \hat{d}^{(1)}_{LP}) / \hat{\sigma}(\hat{d}^{b,m}_{LP})$.

[2]The simulations used the MATLAB FARIMACV routine, which generates data using the autocovariance structure of the ARFIMA process corresponding to Type I fBM, see Davidson and Hashimzade (2009).

[3] Intensive computation was required to carry out the estimations in this simulation analysis and it would have taken approximately 6 months to complete and not practical. To deal with this problem of lengthy simulations, especially those involving simulating bootstrap methods, we used parallel processing across multiple computers and submitted the optimized MATLAB codes for different models and parameters simultaneously.

Kerry Patterson is Emeritus Professor of Econometrics at the University of Reading, England. His career spans over 40 years and he has published widely in econometrics, computational econometrics, macroeconomics, macroeconomic modelling, including articles in leading journals such as The Review of Economics and Statistics, The Economic Journal and Economica. One of his leading contributions to economics, econometrics and macroeconomic model building was to construct an analytical and empirical framework to integrate the data generation and data measurement processes into a unified framework for model building. His recent works include a trilogy of books on the subject of nonstationarity and unit roots in econometrics, published by Palgrave-MacMillan; his textbook, An Introduction To Applied Econometrics (MacMillan, 2000), has become a standard in the area of applied econometrics and is widely cited. He inaugurated the Palgrave-MacMillan Handbook of Econometrics in two volumes, which contains articles by leading econometricians, and is widely referenced. He was also the Founding Editor of the journal entitled Econometrics.

Saeed Heravi is Professor of Quantitative Methods at Cardiff Business School, Cardiff University, Wales. He obtained his Ph.D. in 1985 from the University of Manchester Institute of Science and Technology (UMIST). His research interests are in time series analysis, forecast evaluation, Price indices and applied econometrics. He has widely published in International Journals, including articles in the journal of time series analysis, Journal of Econometrics, Journal of Business and Economic Statistics and International Journal of Forecasting.