

VOLUME LXXIV – N. 1

GENNAIO – MARZO 2020

# RIVISTA ITALIANA DI ECONOMIA DEMOGRAFIA E STATISTICA



**DIRETTORE**

CHIARA GIGLIARANO

**COMITATO SCIENTIFICO**

GIORGIO ALLEVA, GIAN CARLO BLANGIARDO, LUIGI DI COMITE, MAURO GALLEGATI, GIOVANNI MARIA GIORGI, ALBERTO QUADRIO CURZIO, CLAUDIO QUINTANO, SILVANA SCHIFINI D'ANDREA

**COMITATO DI DIREZIONE**

CHIARA GIGLIARANO, CLAUDIO CECCARELLI, PIERPAOLO D'URSO, SALVATORE STROZZA, ROBERTO ZELLI

**REDAZIONE**

LIVIA CELARDO, MARIATERESA CIOMMI, ANDREA CUTILLO, GIUSEPPE GABRIELLI, ALESSIO GUANDALINI, SIMONA PACE, GIUSEPPE RICCIARDO LAMONICA, ANDREA SPIZZICHINO

*Sede Legale:* C/O Studio Associato Cadoni, Via Ravenna n. 34 – 00161 ROMA.  
info@sieds.it, rivista@sieds.it

## INDICE

Margherita Gerolimetto, Stefano Magrini <i>Bootstrap methods for long-range dependence: Monte Carlo evidence</i> .....	5
Giuseppe Ricciardo Lamonica, Luca Salvati, Margherita Carlucci <i>Vertical specialization across the world: evidence from the world input-output table</i> .....	17
Marco Lattanzio <i>The new methodology for the STS indicator on the “number of persons employed”</i> .....	29
Sandro Stancampiano <i>Management of cultural heritage using Big Data Data</i> .....	41
Andrea Fasulo, Marco D. Terribili <i>International cooperation for redesigning Moldovan social surveys</i> .....	51
Claudia De Vitiis, Francesca Inglese <i>La valutazione dell’effetto tecnica nelle indagini sociali a tecnica mista</i> .....	65
Eugenia De Rosa, Claudia De Vitiis, Francesca Inglese, Alberto Vitalini <i>Il web-Respondent Driven Sampling per lo studio della popolazione LGBT+</i> .....	73
Antonella Bernardini, Daniela Bonardo, Alessandra Dentini, Maura Giacummo, Matteo Mazziotta, Alessandra Preti <i>The Italian continuous censuses: managing the complexity among registers, statistics and territories</i> .....	85
Patrizia Collesi, Alessandro Valentini <i>Le attività dell’Istat per la scuola e sul territorio</i> .....	97

## **BOOTSTRAP METHODS FOR LONG-RANGE DEPENDENCE: MONTE CARLO EVIDENCE**

Margherita Gerolimetto, Stefano Magrini

### **1. Introduction**

Bootstrap methods (Efron, 1979) are a computer intensive approach, based on resampling, to statistical inference issues without any statistical assumption on the underlying data generating process. The bootstrap is often adopted because it has much better performance than the conventional approaches (first order asymptotics) and it provides empirical and efficient inference for complicated problems. However, bootstrap methods cannot be equally applicable to all random processes.

In general, for data that are not independent identically distributed (iid), modifications of the original bootstrap set-up have been put forward in order to be able to resample without ignoring the dependence structure. For time series data, among the most common proposals are the block-bootstrap by Künsch (1989) that follows a “data-block” mechanism and the sieve bootstrap by Bühlman (1997) that shares the principle of data transformation. Both methods are time domain-based and have been developed for weakly dependent time series data (e.g. from short memory ARMA models) and. In a parallel line of research, there are also proposals in the frequency domain; in particular, Dahlhaus and Janas (1996) in their seminal paper established a bootstrap method for ratio statistics in case of weakly dependent data, where the Fourier transform is adopted in order to weaken the dependence structure.

In the last decades there has been an increasing interest in strongly dependent (or long-range dependent) time series processes, that can be in general categorized as those processes whose autocovariance decays slowly, in contrast to weakly dependent (or short-range dependent) processes whose covariance decay is fast. Statistical issues between strongly and weakly dependent data change dramatically and, unsurprisingly given the high persistence implied by a slowly decaying autocovariance, the use of resampling methods is more complicated for highly persistent data. The degree of complexity increases even more in case of the dependence structure is not just highly persistent but actually nonstationary. As we

will see in the following sections, this will require a specific additional step of first order differencing of the time series.

The aim of this work is two-fold. On the one hand, we study via Monte Carlo simulations the performance of block bootstrap and sieve bootstrap methods, originally developed for weakly dependent time series, in case of strongly dependent time series, both stationary and nonstationary. On the other hand, we propose an approach to improve the performance of those methods in case of nonstationarity and we will show its finite sample performance via Monte Carlo.

The strongly dependent data generating processes we consider is long memory processes, in particular FARIMA( $p, d, q$ ). For these processes, in their stationary version ( $d < 1/2$ ), in literature there are contributes where it has been extended the validity of the block bootstrap (Kim and Nordman, 2011) and sieve bootstrap (Poskitt, 2008). To the best of our knowledge, what happens in the nonstationarity case ( $d > 1/2$ ) is not clear yet and in this paper we intend to make a move in this direction by investigating the effective performance of these methods for data generating processes that are not just strongly dependent, but non-stationary. Results show that the performance of these bootstrap methods worsens with the increase of the persistence, however it improves if the bootstrap algorithms are augmented with the additional step depicted in this paper.

The structure of the paper is as follows. In the second section, we provide details on the long memory. In the third section, we describe the bootstrap methods under examination. In the fourth section, we present our proposal of a modification of the bootstrap algorithms to handle nonstationarity. In the fifth section, we present our Monte Carlo experiments and some conclusions.

## 2. Long-range dependence

A linear stationary process  $\{X_t\}$  with mean  $E(X_t) = \mu$  is defined as

$$X_t = \mu + \sum_{j \in \mathbb{Z}} b_j \epsilon_{t-j} \quad (1)$$

where  $\{\epsilon_t\}$  are i.i.d. innovations with  $E(\epsilon_t) = 0$  and  $E(\epsilon_t^2) < \infty$ ,  $\sum_{j \in \mathbb{Z}} b_j^2 < \infty$ . The process  $\{X_t\}$  is characterized by long-range dependence (LRD) if the autocovariance function  $r(k) = Cov(X_0, X_k)$  satisfies the following slow decay condition:

$$r(k) \sim \sigma^2 k^{-\theta}, k \rightarrow \infty \quad (2)$$

for some  $\theta \in (0,1)$  and  $\sigma^2 > 0$ . If  $\theta = 1$ , the process is characterized by short-range dependence (SRD). Another condition for the process  $\{X_t\}$  to be LRD is that the partial sums of the autocovariances  $\sum_{k=1}^{\infty} |r(k)| = O(n^{1-\theta})$  diverges as  $n \rightarrow \infty$  (Robinson, 1995). On the other hand, for short-range dependent time series the autocovariance decay rapidly to 0 as  $k \rightarrow \infty$ , so that  $\sum_{k=1}^{\infty} |r(k)| < \infty$ .

Since in this paper we consider inference about the sample mean  $\bar{X}_n = n^{-1} \sum_{t=1}^n X_t$ , we need to characterize the LRD property in terms of the behavior of the sample variance of the sample mean. If we denote  $\sigma_{n,\theta}^2 = n^\theta \text{var}(\bar{X}_n)$ , the slow decay condition implies that

$$\lim_{n \rightarrow \infty} \sigma_{n,\theta}^2 = \sigma_{n,\infty}^2 > 0 \quad (3)$$

holds for a constant  $\sigma_{n,\infty}^2$ , depending on  $\theta \in (0,1)$ . The implication behind (3) is important as it means that the sample variance estimator of the sample mean,  $\text{var}(\bar{X}_n)$ , decays under LRD at a slower rate  $O(n^{-\theta})$  as  $n \rightarrow \infty$  than the usual  $O(n^{-1})$  under SRD (that indeed it is obtained if  $\theta = 1$ ).

So, for a linear process with LRD, if condition (3) holds, we have the normal limit for the scaled sample mean (Davidov, 1970)

$$n^{\theta/2}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma_{n,\infty}^2) \quad \text{as } n \rightarrow \infty \quad (4)$$

In this paper, we will focus in particular on the class of long memory FARIMA( $p,d,q$ ) models

$$\phi(B)(1-B)^d X_t = \mu + \psi(B)\epsilon_t \quad (5)$$

where  $\{\epsilon_t\}$  are white noise, as above,  $B$  is the lag operator and the long memory parameter  $d$  is  $d = \frac{1-\theta}{2}$ . So, if  $\theta = 1$ , we have that  $d = 0$ , *i.e.* ARMA( $p,q$ ) models. Stationarity holds if  $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ . The long memory parameter  $d$  can be estimated in a variety of ways, here we will adopt two well-known methods: the GHP method (Geweke and Porter-Hudack, 1983) and Whittle method (Fox and Taqqu, 1986; Dahlhaus, 1989).

### 3. Bootstrap for time series: an overview

The bootstrap is a method to estimate the distribution of an estimator or a test statistic by resampling one's data. Under iid conditions, the bootstrap provides

approximations that are at least as accurate first order asymptotics. Often the bootstrap provides approximations that are even more accurate, especially when the sample size is not large

When data are not iid, as in time series case, the bootstrap should be revisited in order to preserve the dependence structure in the data generating process (DGP). In the next subsections, we revise two methods developed with that purpose.

### 3.1. Block Bootstrap

The block bootstrap (Künsch, 1989) is based on the principle of capturing or preserving the dependence structure of the original time series. In practice, it means resampling data blocks, consisting of consecutive groups of data points.

Two types of block bootstrap have been proposed and for both validity has been originally proved under short-range dependence. One is the moving block bootstrap (MBB), when resampling is done with respect to overlapping blocks (Künsch, 1989) the other one (Carlstein, 1986) is when resampling is done with respect to non-overlapping blocks (NBB). The following is the algorithm:

- 1) Given a time series  $X_t, t=1, \dots, n$ , define block size  $l < n$  and compute the number of blocks for the resampling procedure  $b = \lfloor \frac{n}{l} \rfloor$
- 2) Construct data blocks (either overlapping, MBB, or not overlapping, NBB)
- 3) Generate bootstrap replicates from the data-block set as:
  - (a) for a MBB series  $X_1^*, \dots, X_N^*$ , where  $N = bl$ , we generate  $I_1^*, \dots, I_b^*$  from iid uniform random variables  $\{I_1^*, \dots, I_{n-l+1}^*\}$
  - (b) for a NBB series  $X_1^*, \dots, X_N^*$ , where  $N = bl$ , we generate  $I_1^*, \dots, I_b^*$  from iid uniform random variables  $\{I_1^*, \dots, I_{n+l(b-1)}^*\}$
- 4) Make a MBB/NBB series  $X_1^*, \dots, X_N^*$ , where  $N = bl$

Approximating the distribution of  $n^{\theta/2}(\bar{X}_n - \mu)$  with the bootstrap counterpart  $N^{\theta/2}(\bar{X}_n - E_*\bar{X}_n^*)$ , that is valid in case of short-range dependence, appears to be a natural choice also in case of LRD. However, in this latter case, it is actually wrong as pointed out by Künsch (1989) and Lahiri (1993). More recently, Kim and Nordman (2011) proved that for long memory time series, the sample mean should be inflated by adjustment factor of  $b^{1-\theta/2}$  and the distribution of  $n^{\theta/2}(\bar{X}_n - \mu)$  will be approximated as follows:

$$\sup_{x \in \mathbb{R}} \left| P_* \left( b^{\frac{1}{2}} l^{\frac{\theta}{2}} (\bar{X}_n^* - E_* \bar{X}_n^*) \leq x \right) - P \left( n^{\frac{\theta}{2}} (\bar{X}_n - \mu) \leq x \right) \right| \xrightarrow{p} 0 \quad (6)$$

In other words, the inflating factors ensures that the MBB version  $b^{\frac{1}{2}}l^{\frac{\theta}{2}}\bar{X}_n^*$  has the “right” variance for approximating the distribution of  $n^{\theta/2}\bar{X}_n^*$  under LRD (the same hold for NBB).<sup>1</sup>

Note that the performance of the BB depends on a block choice and optimal blocks for variance estimation are shown to decrease as the strength of the underlying process increases. Blocks of size  $O(n^{1/2})$  may be a compromise for use in practice.

### 3.2. Sieve Bootstrap

The sieve bootstrap (Bühlmann, 1997) approximates a general linear invertible process by a finite autoregressive model of order  $p = p(n)$ , where  $p(n) \rightarrow \infty$ , then it resamples from the approximated autoregression. This method takes up the older idea of fitting parametric models first and then resampling from the residuals, but instead of considering a fixed finite-dimensional model, an infinite-dimensional, non-parametric model is approximated by a sequence of finite-dimensional models. This method can be considered non parametric because it is model-free in the class of the linear invertible processes.

The properties of the sieve bootstrap have been rigorously investigated, among other, by Kreiss (1992), Paparoditis (1996), Bühlmann (1997), Bickel and Bühlmann (1999) who established its asymptotic validity for several statistics assuming that the data generating process is an infinite order autoregressive. Kapetanios and Psaradakis (2006) and Poskitt (2008) proved that under regularity conditions (satisfied by stationary long memory processes) the sieve bootstrap provides an asymptotically valid approximation to the distributions of several statistics. The following is the algorithm:

- 1) Given a time series  $X_t$ ,  $t = 1, \dots, n$ , fit an AR( $h$ ) model  $h > 0$ . Obtain the residuals of the AR( $h$ ) model and standardize them, denote by  $\hat{\zeta}_t$
- 2) Create a new randomly resampled residuals set, denoted by  $\hat{\zeta}_t^*$
- 3) Generate the bootstrap time series  $X_t^*$  as

$$X_t^* = \hat{\alpha}_1 X_{t-1}^* + \dots + \hat{\alpha}_h X_{t-h}^* + \hat{\zeta}_t^* \quad (7)$$

- 4) By repeating the above procedure a number of times  $B$  we obtain a bootstrap approximation to the distribution of the desired statistic

---

<sup>1</sup> We emphasize that (6) holds for stationary long memory, *i.e.*  $|d| < 1/2$ , the nonstationary area ( $d > 1/2$ ) has not been investigated yet.

For stationary long memory time series Poskitt (2008) proved that the distribution of  $n^{\theta/2}(\bar{X}_n - \mu)$  will be approximated as follows:

$$\sup_{x \in \mathbb{R}} \left| P_* \left( n^{\frac{\theta}{2}} (\bar{X}_n^* - \bar{X}_n) \leq x \right) - P \left( n^{\frac{\theta}{2}} (\bar{X}_n - \mu) \leq x \right) \right| \xrightarrow{p} 0 \quad (8)$$

Note that the SB performance depends on the order  $h$  of the AR model fitted to the data. A reasonable choice is to adopt the Akaike criterion as suggested in Bühlmann (1997).

#### 4. Bootstrapping nonstationary long memory time series

When,  $X_t$  is nonstationary, *i.e.*  $d > 1/2$ , bootstrapping is an even more complex issue. Inspired by the literature of bootstrap for unit root-test, the idea we propose here is to preliminarily first difference the nonstationary long memory time series (Palm *et al.* 2008), in order to bring it back to the stationary area; then we apply the chosen bootstrap to  $Z_t = \Delta X_t$ . This is also in line with Psaradakis (2001) and Chang and Park (2003) who proved that applying the sieve bootstrap to first difference is a valid bootstrap approach to nonstationary I(1) time series.<sup>2</sup>

So, the chosen bootstrap algorithm is augmented by two additional steps, before and after the implementation of the bootstrap itself:

[Add 1]: Take the first difference  $Z_t = \Delta X_t$

[Boot]: Obtain a bootstrap sample  $Z_t^*$  with the chosen bootstrap

[Add 2]: Obtain  $X_t^* = X_{t-1}^* + Z_t^*$

By preliminarily differentiate the time series, the conditions required to guarantee the validity of the bootstrap are respected. Indeed, the series which is effectively bootstrapped is  $Z_t = \Delta X_t$ , whose long memory parameter is  $d_z = d - 1$ , will be  $|d_z| < 1/2$ , so stationary long memory.

---

<sup>2</sup> Palm *et al.* (2008) also showed that for some data generating processes (not our case), residuals from a first order autoregression can lead to an even better performance of the sieve bootstrap in terms of asymptotic validity, compared to first difference.



## 5. Monte Carlo experiment

The aim of the experiment is to investigate the performance of block bootstrap and sieve bootstrap to obtain confidence intervals for  $\mu$  the mean of the process for a variety of long memory DPGs, both stationary and nonstationary. The number of MC replications is  $S=5000$ . Both stationary ( $d=0.25, 0.45$ ) and nonstationary ( $d=0.65, 0.80$ ) FARIMA models have been considered, in four versions (that include or not the short-range components):

DGP1: FARIMA(1,d,1),  $\phi=0.3, \psi=0.4$

DGP2: FARIMA(0,d,1),  $\psi=0.4$

DGP3: FARIMA(1,d,0),  $\phi=0.3$

DGP4: FARIMA(0,d,0)

A further DGP has been included for benchmarking with the short memory case.

DGP5: ARMA(1,0,1),  $\phi=0.3, \psi=0.4$

The sample size is  $n = 100, 250, 500$ , innovations are  $\epsilon_t \sim N(0,1)$ . The considered methods are block bootstrap (BB) and sieve bootstrap (SB) both in the stationary and revised version for nonstationarity.

The BB is implemented in the overlapping version, the block length is  $l = \lceil \sqrt{n} \rceil$ , in line with, for instance Kim and Nordman (2011). The SB is implemented using Yule walker estimates, the AR order  $h$  is set adopting the Akaike criterion as in Bühlmann (1997).  $B=500$  is the number of bootstrap samples.

The performance is expressed in terms of empirical coverage of (symmetric) bootstrap intervals at 90%:  $\bar{X}_n^* \pm n^{-\theta/2} q_{0.90}^*$ , where  $q_{0.90}^*$  is defined as

$$P_* \left( \frac{1}{b^2 l^2} |\bar{X}_n^* - E_* \bar{X}_n^*| \leq q_{0.90}^* \right) = 0.9 \text{ for the block bootstrap}$$

$$P_* \left( \frac{\theta}{n^2} |\bar{X}_n^* - \bar{X}_n| \leq q_{0.90}^* \right) = 0.9 \text{ for the sieve bootstrap}$$

Note that  $\theta = 1 - 2d$ . Bootstrap interval are computed without knowledge on  $d$  and it is adopted the Whittle or GPH estimate, so that  $\hat{\theta} = 1 - 2\hat{d}$ .

Results are presented in the following Tables 1, 2, 3. The performance, *i.e.* the empirical coverage, should be read in the sense that the closer it is to 0.90, the more satisfactory it the behavior of the bootstrap.

**Table 1** – Empirical coverage of bootstrap confidence intervals at 90% - stationarity.

DGP ( $\phi, d, \psi$ )	Bootstrap	N=100		N=250		N=500	
		$\hat{d}_{wh}$	$\hat{d}_{gph}$	$\hat{d}_{wh}$	$\hat{d}_{gph}$	$\hat{d}_{wh}$	$\hat{d}_{gph}$
DGP 1, $d=0.25$	BB	0.648	0.773	0.779	0.756	0.810	0.820
(0.3, $d, 0.4$ )	SB	0.635	0.607	0.788	0.848	0.813	0.884
DGP 2, $d=0.25$	BB	0.621	0.735	0.746	0.790	0.800	0.842
(0.0, $d, 0.4$ )	SB	0.632	0.687	0.722	0.854	0.823	0.862
DGP 3, $d=0.25$	BB	0.638	0.766	0.745	0.784	0.815	0.822
(0.3, $d, 0.0$ )	SB	0.612	0.683	0.732	0.871	0.821	0.893
DGP 4, $d=0.25$	BB	0.786	0.758	0.826	0.789	0.892	0.839
(0.0, $d, 0.0$ )	SB	0.808	0.689	0.851	0.845	0.901	0.883
DGP 1, $d=0.45$	BB	0.489	0.594	0.530	0.601	0.569	0.619
(0.3, $d, 0.4$ )	SB	0.565	0.615	0.540	0.628	0.581	0.763
DGP 2, $d=0.45$	BB	0.501	0.565	0.527	0.584	0.551	0.604
(0.0, $d, 0.4$ )	SB	0.544	0.581	0.550	0.799	0.561	0.861
DGP 3, $d=0.45$	BB	0.523	0.576	0.539	0.598	0.581	0.611
(0.3, $d, 0.0$ )	SB	0.534	0.505	0.547	0.823	0.590	0.857
DGP 4, $d=0.45$	BB	0.588	0.566	0.632	0.570	0.670	0.606
(0.0, $d, 0.0$ )	SB	0.649	0.532	0.700	0.815	0.748	0.821
DPG 5, short mem.	BB	0.833		0.843		0.874	
(0.3, 0, 0.4)	SB	0.801		0.822		0.855	

Table 1, shows the performance of BB and SB in case of stationary long memory. We see that, as expected, the performance improves with the increase of  $n$ . Not surprisingly it tends to worsen when  $d=0.45$ , as a consequence of the closeness to the nonstationarity region. There is no large difference between the SB and the BB method, however, especially for  $d=0.45$  it seems that the SB performs slightly better. As for the type of DGP, the presence of short memory components in DGP 1-3 slightly affects the performance of the method. Indeed, in case of fractional noise, DGP 4, BB and SB perform better, especially if the estimate of  $d$  is obtained via Whittle method. Nevertheless, the GPH method is still a very good option, especially in those case where there is no exact knowledge about the short memory components. The short memory DGP (DGP 5) has been included as a benchmark, to show how effectively this methods work in case of weak dependence, i.e. what they have been created for. The performance is on average better, but encouragingly not too far from that of the long memory DGP.

Table 2 and 3 focus on nonstationary versions of DGPs 1-4. In the table the original version of the SB and BB has been coupled with the revised versions (Rev-BB, Rev-SB) proposed in the previous section.

**Table 2** – Empirical coverage of bootstrap confidence intervals at 90% - nonstationarity  $d=0.65$ 

DGP ( $\phi, d, \psi$ )	Bootstrap	N=100		N=250		N=500	
		$\hat{d}_{wh}$	$\hat{d}_{gph}$	$\hat{d}_{wh}$	$\hat{d}_{gph}$	$\hat{d}_{wh}$	$\hat{d}_{gph}$
DGP 1, $d=0.65$ (0.3, $d, 0.4$ )	BB	0.499	0.537	0.600	0.622	0.679	0.744
	Rev-BB	0.548	0.638	0.676	0.682	0.700	0.781
	SB	0.500	0.589	0.619	0.711	0.758	0.806
	Rev-SB	0.589	0.658	0.700	0.803	0.850	0.886
DGP 2, $d=0.65$ (0.0, $d, 0.4$ )	BB	0.500	0.578	0.600	0.634	0.659	0.754
	Rev-BB	0.623	0.703	0.639	0.689	0.700	0.980
	SB	0.526	0.586	0.623	0.708	0.717	0.808
	Rev-SB	0.684	0.754	0.678	0.779	0.779	0.891
DGP 3, $d=0.65$ (0.3, $d, 0.0$ )	BB	0.528	0.588	0.598	0.681	0.676	0.671
	Rev-BB	0.589	0.685	0.739	0.709	0.719	0.773
	SB	0.550	0.595	0.628	0.739	0.709	0.821
	Rev-SB	0.723	0.644	0.747	0.777	0.801	0.847
DGP 4, $d=0.65$ (0.0, $d, 0.0$ )	BB	0.686	0.593	0.706	0.601	0.714	0.694
	Rev-BB	0.729	0.637	0.829	0.717	0.841	0.780
	SB	0.734	0.622	0.801	0.770	0.856	0.807
	Rev-SB	0.778	0.648	0.852	0.810	0.890	0.880

**Table 3** – Empirical coverage of bootstrap confidence intervals at 90% - nonstationarity  $d=0.80$ 

DGP ( $\phi, d, \psi$ )	Bootstrap	N=100		N=250		N=500	
		$\hat{d}_{wh}$	$\hat{d}_{gph}$	$\hat{d}_{wh}$	$\hat{d}_{gph}$	$\hat{d}_{wh}$	$\hat{d}_{gph}$
DGP 1, $d=0.80$ (0.3, $d, 0.4$ )	BB	0.469	0.515	0.584	0.628	0.688	0.744
	Rev-BB	0.539	0.600	0.581	0.678	0.676	0.774
	SB	0.500	0.523	0.504	0.572	0.605	0.706
	Rev-SB	0.564	0.617	0.629	0.713	0.717	0.805
DGP 2, $d=0.80$ (0.0, $d, 0.4$ )	BB	0.477	0.538	0.536	0.604	0.588	0.699
	Rev-BB	0.615	0.674	0.585	0.683	0.677	0.789
	SB	0.518	0.529	0.512	0.573	0.590	0.620
	Rev-SB	0.674	0.701	0.658	0.746	0.731	0.818
DGP 3, $d=0.80$ (0.3, $d, 0.0$ )	BB	0.500	0.515	0.533	0.623	0.595	0.683
	Rev-BB	0.564	0.633	0.617	0.705	0.686	0.747
	SB	0.529	0.584	0.544	0.632	0.660	0.700
	Rev-SB	0.660	0.717	0.698	0.761	0.773	0.821
DGP 4, $d=0.80$ (0.0, $d, 0.0$ )	BB	0.678	0.569	0.663	0.602	0.729	0.701
	Rev-BB	0.724	0.629	0.729	0.667	0.834	0.786
	SB	0.708	0.607	0.710	0.654	0.800	0.720
	Rev-SB	0.756	0.676	0.784	0.758	0.876	0.835

By reading Table 2 and 3, also in comparison with Table 1, we see that the performance of the BB and SB worsens, the more deeply we enter into the nonstationary areas. However, the revised versions (Rev-BB, Rev-SB) improve with respect to the unrevised ones, as their performance resembles in magnitude order the performance in case of stationary long memory. The improvement is more evident for the SB, that was better performing also in the original version.

Also in this nonstationary case, both Whittle and GPH are very reasonable tool to estimate  $d$ .

Future research lines are to consider the same issues with respect to resampling the periodogram ordinates for long memory, also outside the case when the analytical expression of the spectral density function is known.

## References

- BERAN, J. 1994. *Statistics for long memory time processes*. New York: Chapman and Hall.
- BICKEL P.J, BÜHLMANN P. 1999. A new mixing notion and functional central limit theorems for a sieve bootstrap in time series, *Bernoulli*, Vol. 5, pp. 413-446.
- BÜHLMANN, P. 1997. Sieve bootstrap for time series, *Bernoulli*, Vol. 3, pp. 123-148.
- CARLSTEIN, E. 1986. The use of subseries methods for estimating the variance of a general statistic for a stationary time series, *Annals of Statistics*, Vol. 14, pp. 1171-1179.
- CHANG Y., PARK J.Y. 2003. A Sieve Bootstrap For The Test Of A Unit Root, *Journal of Time Series Analysis*, Vol. 24, pp. 379-400.
- DAHLHAUS, R. 1989. Efficient parameter estimation for self-similar processes, *The Annals of Statistics*, Vol. 17, pp. 1749-1766.
- DAHLHAUS R., JANAS D. 1996. A frequency domain bootstrap for ratio statistics in time series analysis, *The Annals of Statistics*, Vol. 24, pp. 1934-1963.
- DAVIDOV, Y.A. 1970. The invariance principle for stationary processes, *Theory of Probability & Its Applications*, Vol. 15, pp. 487-498.
- EFRON, B. 1979. Bootstrap methods: another look at the jackknife, *Annals of Statistics*, Vol. 7, pp. 1-26.
- FOX R., TAQQU M.S. 1986. Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *The Annals of Statistics*, Vol. 14, pp. 517-532.
- GEWEKE J., PORTER-HUDACK S. 1983. The Estimation and Application of Long-Memory Time Series Models, *Journal of Time Series Analysis*, Vol. 4, pp. 221-237.
- KAPETANIOS G., PSARADAKIS Z. 2006. Sieve Bootstrap for Strongly Dependent Stationary Processes. *Working Papers 552*, Queen Mary University of London, School of Economics and Finance.
- KIM Y.M., NORDMAN D.J. 2011. Properties of block bootstrap under long range dependence, *Sankhya: The Indian Journal of Statistics, series*, Vol. 73, pp. 79-109.

- KREISS, J.P. 1992. Bootstrap procedures for  $AR(\infty)$ -processes. In *Lecture Notes in Economics and Mathematical Systems, Bootstrapping and Related Techniques*, Vol. 376, pp. 107-113.
- KÜNSCH, H.R. 1987. Statistical aspects of self-similar processes. In *Proceedings of the 1<sup>st</sup> World Congress of Bernoulli Society*, Vol. 1, Utrecht: VNU Science Press, pp.67-74.
- KÜNSCH, H.R. 1989. The jackknife and the bootstrap for general stationary observations, *The Annals of Statistics*, Vol. 17, pp. 1217-1241.
- LAHIRI, S.N. 1993. On the moving block bootstrap under long-range dependence, *Statistics and Probability Letters*, Vol. 18, pp. 405-413.
- PALM F.C., SMEEKES S., URBAIN J.P. 2008. Bootstrap Unit Root test: Comparison and Extensions, *Journal of time series analysis*, Vol. 29, pp. 371-400.
- PAPARODITIS, E. 1996. Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes, *Journal of Multivariate Analysis*, Vol. 57, pp. 277-296.
- POSKITT, D.S. 2008. Properties of the Sieve Bootstrap for Fractionally Integrated and Non-Invertible Processes, *Journal of time series analysis*, Vol. 29, pp. 224-250.
- PSARADAKIS, Z. 2001. Bootstrap tests for an autoregressive unit root in the presence of weakly dependent errors, *Journal of Time Series Analysis*, Vol. 22, pp-577-594.

## SUMMARY

### **Bootstrap methods for long-range dependence: Monte Carlo evidence**

In this paper we present a review of some well-known bootstrap methods for time series data. We concentrate on block bootstrap and sieve bootstrap, whose validity has been proved to be extended to stationary long memory time series.

We will start by reviewing briefly the peculiar features of the bootstrap methods and the issues raised in case of long range dependent data; then we present a Monte Carlo experiment to compare the performance of the methods for a variety of ARFIMA processes. Comments about the finite sample performance of the methods will be provided also in light of the established theoretical properties of the methods

---

MARGHERITA GEROLIMETTO, Università Ca' Foscari Venezia, Dipartimento di Economia. E-mail: [margherita.gerolimetto@unive.it](mailto:margherita.gerolimetto@unive.it)

STEFANO MAGRINI, Università Ca' Foscari Venezia, Dipartimento di Economia. E-mail: [stefano.magrini@unive.it](mailto:stefano.magrini@unive.it)