Working Paper Series, N. 6, 2016

Estimation of INAR(p) models using bootstrap

Luisa Bisaglia
Department of Statistical Sciences
University of Padua
Italy

Margherita Gerolimetto
Department of Economics
University Ca' Foscari, Venice
Italy

Abstract: In this paper we investigate bootstrap techniques applied to the estimation of the thinning parameters in INAR(p) models. We propose a new bootstrap approach based on sieve bootstrap. The approach is then applied to the Yule-Walker estimator of the thinning parameters. Monte Carlo experiments are carried out to valuate the performance of bootstrap estimator and the superiority of our proposal is stated in terms of low bias and Mean Square Error (MSE). Finally, a real time series is analyzed.

Keywords: INAR(p) models, estimation, bootstrap

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Luisa Bisaglia
tel: +39 049 827 4168
luisa.bisaglia@unipd.it

Estimation of INAR(p) models using bootstrap

Luisa Bisaglia
Department of Statistical Sciences
University of Padua
Italy

Margherita Gerolimetto
Department of Economics
University Ca' Foscari, Venice
Italy

Abstract: In this paper we investigate bootstrap techniques applied to the estimation of the thinning parameters in INAR(p) models. We propose a new bootstrap approach based on sieve bootstrap. The approach is then applied to the Yule-Walker estimator of the thinning parameters. Monte Carlo experiments are carried out to valuate the performance of bootstrap estimator and the superiority of our proposal is stated in terms of low bias and Mean Square Error (MSE). Finally, a real time series is analyzed.

Keywords: INAR(p) models, estimation, bootstrap

1 Introduction

Recently, there has been a growing interest in studying nonnegative integer-valued time series and, in particular, time series of counts. Examples are the number of road accidents, number of traded stocks in a firm, number of visitors to a website, incidence of a disease, number of absent workers in a firm, number of guests in a hotel and so on. In some cases, the discrete values of the time series are large numbers and may be analyzed by using continuous-valued models such as the ARMA ones with Gaussian errors. However, a good model for time series should be consistent with the properties of the data and be unable to predict values which violate known constraints. Thus, when the values are small, as in the case of counting processes, the usual linear ARMA processes are of limited use for modeling and forecasting purposes in that they would invariably produce non-integer forecast values. The most common approach to build an integer-valued autoregressive (INAR) process is based on a probabilistic operation called binomial thinning, as reported in Al-Osh and Alzaid (1987) and McKenzie (1988) who first introduced INAR processes. While theoretical properties of INAR models with Poisson innovations have been extensively studied in the literature (see, for instance, Freeland and McCabe (2004), Bu et al. (2008), and the references therein), relatively few contributions discuss the development of estimation methods when the distribution of the error terms is different from the Poisson. The Poisson distribution, however, has the disadvantage of allowing only for equi-dispersion. Thus, unlike the usual applications where the error terms are Poisson, in order to investigate over and under-dispersion cases, we have to assume that the error terms follow different integer distributions.[1] When the distribution of the error term is not known, it is not possible to calculate the likelihood and so to estimate the parameters of the model. There exist other methods, like for instance, the Yule-Walker (YW), but in this case we cannot construct confidence intervals nor perform hypothesis testing on the thinning parameters. In this work we contribute to cater this problem by means of bootstrap approaches. In particular, we consider the block bootstrap (BB) completely nonparametric, a version of the sieve bootstrap (SB) used by Cardinal et al. (1999) and Kim and Park (2008) and a new proposal based on the sieve bootstrap (SB-INAR) that take into account the integer nature of data. Some extensive Monte Carlo experiments are carried out to evaluate the performance of bootstrap estimators. These experiments show evidence in favor of the approach we propose.

The paper is outlined as follows. In section 2, INAR(p) models are described. Section 3 details the bootstrap methods we used and the one we propose. Section 4 is devoted to present the results of the Monte Carlo experiments. Section 5 presents an application to a real data set. Section 6 concludes.

2 INAR(p) models

In spite of the central role of the Box-Jenkins ARMA, there is no such a leading technique for count time series. A proposal is the integer-valued autoregressive process (INAR) (McKenzie (1985), Al-Osh and Alzaid (1987)):

$$X_t = \alpha_1 \circ X_{t-1} + \ldots + \alpha_p \circ X_{t-p} + \epsilon_t$$

where '$\circ$'is the thinning operator defined to satisfy $\alpha \circ X = \sum_{i=1}^{X} Y_i$ where $X \in \mathbf{N}$, $\alpha \in [0,1]$ and Y_i is a sequence of iid count random variables (typically $Ber(\alpha)$), independent of X, with common mean α. While INAR(1) model is defined univocally, for INAR(p) model there are added complexities not present in the INAR(1) case, and, depending on (different) binomial thinning mechanism, we can distinguish different types of INAR(p) processes. Here we recall that of Alzaid and Al-Osh (1990) which is a direct extension of INAR(1) and that due to Du and Li (1991) which is closer to the linear Gaussian AR(p). In this work we refer to the last specification. For this specification the stationarity of the process is guaranteed if $0 \leq \sum_{j=1}^{p} \alpha_j < 1$, the correlation properties are identical to the linear Gaussian AR(p) model and the conditional mean (regression) function is linear and given by:

$$E(X_t \mid \mathcal{F}_{t-1}) = \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + \mu_\epsilon$$

where $\mathcal{F}_{t-1} = X_{t-1}, X_{t-2}, \ldots$ and $\mu_\epsilon = E[\epsilon_t]$. INAR($p$) models strongly depends on the parametric assumption for the error term. Usually it is assumed that ϵ_t is

[1] One interesting paper is that of Sun and McCabe (2013), where the authors propose the use of the Katz family or the generalized Poisson as distributions for the innovation processes. These families of distributions take into account under- and over- dispersion.

distributed as a Poisson (in this case the model is also called PoINAR), but with count data it may be desirable to model also under- or over- dispersion.

The estimation problem connected with the INAR(p) process is more complicated than that of the AR(p) process. The complication arises from the fact that the conditional distribution of X_t given $(X_{t-1},\dots,X_{t-p})$ is the convolution of the distribution of the innovation process, ϵ_t, and that of p binomial distributions with parameters X_{t-i} and $\alpha_i,\ i=1,\dots,p$.

The main approaches to estimate the INAR's parameters include the Yule-Walker estimator (Al-Osh and Alzaid (1987) for $p=1$, Du and Li (1991) and Latour (1998) for $p\geq 2$) and the Conditional Least Squares (Al-Osh and Alzaid (1987) for $p=1$, Du and Li (1991) and Latour (1998) for $p\geq 2$). The implementation of both approaches is simple and they are asymptotically equivalent.

Al-Osh and Alzaid (1987) and Freeland and McCabe (2004) showed how Maximum Likelihood (ML) can be implemented for estimating the parameters of INAR(1) model. Simulation studies compare the finite sample properties of the three methods and conclude that ML has the best performance in terms of bias and mean square error. Bu et al. (2008) develop a general framework for likelihood analysis of INAR(p) processes with general thinning operators and innovation distributions. Then, they specialize to the situation to the INAR(2) specification, where the thinning processes are binomial and the innovation sequence is Poisson, and provide specific formulae for computational use in this case.

Finally, the general case where $p>1$ and the innovation terms are whatever is, to the best of our knowledge, not considered. This justifies this work, where we attach the problem of estimation of thinning parameters in a general case by means of bootstrap approaches.

3 Bootstrap method

Bootstrap methods, initially proposed by Efron (1979) for independent observations, has revealed inefficient when data are dependent, as in the case of time series data. In this case the use of bootstrap for the estimation of population characteristics must be judicious since the time series structure may be lost in a careless resampling. Thus, time series data must be resampled indirectly. A very recent and good review about bootstrap for time series is, for example, that of Kreiss and Lahiri (2012). In the context of INAR processes, to the best of our knowledge, we found only few papers about bootstrap and INAR(p) model. That of Cardinal et al. (1999) and Kim and Park (2008) which propose a bootstrap approach for deriving forecasts and confidence intervals and that of Kim and Park (2010) that apply bootstrap to INAR(p) models to obtain estimated standard errors for the estimated parameters of the model. In this section we describe the bootstrap methods we apply to INAR(p) models.

3.1 Block bootstrap

The first bootstrap method we consider is the block bootstrap introduced by Künsch (1989) and Liu and Singh (1992) for time series that are not assumed to have a specific structural form. Their idea is to resample blocks of observations at time. By retaining the neighboring observations together within the blocks, the dependence structure of the random variables as short lag distances is preserved. As a result, resampling blocks allows one to carry this information over to the bootstrap variables. The BB can be summarized as follow.

Let be $x_t, \ \ t = 1, \ldots, n$ a stationary time series. Let l be an integer satisfying $1 \leq l \leq n$. Define the overlapping blocks $\mathcal{B}_1, \ldots, \mathcal{B}_N$ of length l as $\mathcal{B}_i = (x_i, \ldots, x_{i+l-1})$ starting with $x_i, \ \ 1 \leq i \leq N$ where $N = n - l + 1$. For simplicity, suppose that l divides n and let $b = n/l$. The BB sample is obtained by selecting b blocks at random with replacement from the collection $\mathcal{B}_1, \ldots, \mathcal{B}_N$. Since each resampled block has l elements, concatenating the elements of the b resampled blocks serially yields $b \cdot l$ bootstrap observations $x_1^*, \ldots, x_n^*$.

Performance of the BB method crucially depends on the choice of the block size and on the dependent structure of the process. In this work we choose $l = \sqrt{n}$.

For the other several variants of block bootstrap and further details, see Kreiss and Lahiri (2012) and the reference therein.

3.2 Sieve bootstrap

The sieve bootstrap was first introduced by Kreiss (1992) and then developed by Bühlmann (1997). This method is based on the idea of sieve approximation: it approximates a general linear, invertible process by a finite autoregressive model with order increasing with the sample size, and resampling from the approximated autoregressions. By viewing such autoregressive approximations as a sieve for the underlying infinite-order process, the bootstrap procedure may still be regarded as a non parametric one. Cardinal et al. (1999) and Kim and Park (2008) employ this approach after some modifications to incorporate the nature of integer-valued time series as the following steps.

1. Estimate the thinning parameters $(\alpha_1, \ldots, \alpha_p)$ with, for example, the Yule-Walker estimator.

2. Compute the residuals $\hat{\epsilon}_t = x_t - \sum_{i=1}^{p} \hat{\alpha}_i x_{t-i}$ for $t = p+1, \ldots, n$.

3. Construct the empirical distribution for modified residuals defined by $\tilde{\epsilon}_t = [\hat{\epsilon}_t]$ where $[\cdot]$ represents the value rounded to the nearest integer.

4. For $b = 1, \ldots, B$, define the bootstrapped series x_t^b by

$$x_t^b = \hat{\alpha}_1 \circ x_{t-1} + \ldots + \hat{\alpha}_p \circ x_{t-p} + \epsilon_t^b$$

   where ϵ_t^b for $t = 1, 2, \ldots, n$ is an i.i.d. sample from the residuals computed previously.

5. Based on x_t^b, compute the estimation of the thinning parameters $(\hat{\alpha}_{1,b}, \ldots, \hat{\alpha}_{p,b},)$ as in step 1).

6. Estimates of α_i can be obtained considering the sample mean: $\hat{\alpha}_i^* = \sum_{i=1}^{B} \hat{\alpha}_{i,b}/B$ for $i = 1, \ldots, p$.

3.3 Sieve bootstrap for INAR(p) models

Now we propose a parametric bootstrap algorithm, based on SB, to estimate the thinning parameters of INAR(p) models. Our approach differs from the previous approach in the computation of the residuals. Let be x_t a time series, our proposal is as follows.

1. Estimate the thinning parameters $(\alpha_1, \ldots, \alpha_p)$ with, for example, the Yule-Walker estimator.

2. Compute the residuals $\hat{\epsilon}_t = x_t - (\hat{\alpha}_1 \circ x_{t-1} + \ldots + \hat{\alpha}_p \circ x_{t-p})$. Observe that $\hat{\alpha}_i \circ x_{t-i}$ are realizations of $Bi(x_{t-i}, \hat{\alpha}_i)$, for $i = 1, \ldots, p$.

3. Since computed residuals could be negative, if $p = 1$ we propose to use the modified residuals
$$\tilde{\epsilon}_t = \begin{cases} \hat{\epsilon}_t & \text{if } \hat{\epsilon}_t \geq 0 \\ 0 & \text{if } \hat{\epsilon}_t < 0 \end{cases}$$
If $p > 1$, modified residuals will be $\tilde{\epsilon}_t = \hat{\epsilon}_t$ if $\hat{\epsilon}_t \geq 0$, but if $\hat{\epsilon}_t < 0$, one computational solution is to recalculate $\tilde{\epsilon}_t$ until it is greater than zero.

4. For $b = 1, \ldots, B$, define the bootstrapped series x_t^b by
$$x_t^b = \hat{\alpha}_1 \circ x_{t-1} + \ldots + \hat{\alpha}_p \circ x_{t-p} + \epsilon_t^b$$
where ϵ_t^b for $t = 1, 2, \ldots, n$ is an i.i.d. sample from the residuals computed previously.

5. Based on x_t^b, compute the estimation of the thinning parameters $(\hat{\alpha}_{1,b}, \ldots, \hat{\alpha}_{p,b},)$ as in step 1).

6. Estimates of α_i can be obtained considering the sample mean: $\hat{\alpha}_i^* = \sum_{i=1}^{B} \hat{\alpha}_{i,b}/B$ for $i = 1, \ldots, p$.

4 Some Monte Carlo results

In this section we report some Monte Carlo results in order to attest for the efficiency of the bootstrap method detailed in the previous section. In the simulation study, we generated 1000 different realization from the following DGP: (i) PoINAR(p) with $\lambda = 1, 5$ (ii) INAR(p) with binomial error term (BINAR) with $m = 6, \pi = 0.2, 0.5, 0.8$, (iii) INAR($p$) with negative binomial error term (NBINAR) with $r = 6, \pi = 0.2, 0.5, 0.8$. For $p = 1$ we consider $\alpha = 0.3, 0.6, 0.9$, for $p = 2$ we consider $\alpha_1 = 0.5, \alpha_2 = 0.3$. The number of bootstrap replications is $B = 500$. The sample size is $n = 250, 500$. The thinning parameters are estimated by the Y-W method. The statistics used to evaluate the bootstrap method are the Monte Carlo bias and mean square error (MSE). R Core Team (2015) is the software package employed for simulations. Results are reported in Table 1-6.

Obviously, the bias of bootstrap method is always greater comparing to MC simulations, but the MSE is of the same order of magnitude. Notice that in case

PoINAR(1), $\lambda = 1$		n=250			n=500		
Method		$\alpha = 0.3$	$\alpha = 0.6$	$\alpha = 0.9$	$\alpha = 0.3$	$\alpha = 0.6$	$\alpha = 0.9$
Monte Carlo	Bias	-0.0081	-0.0132	-0.0162	-0.0068	-0.0062	-0.0082
	MSE	0.0040	0.0031	0.0011	0.0021	0.0016	0.0005
SB-INAR	Bias	**-0.0173**	**-0.0273**	**-0.0328**	**-0.0114**	**-0.0133**	**-0.0170**
	MSE	**0.0041**	**0.0035**	**0.0019**	**0.0021**	**0.0016**	**0.0007**
SB	Bias	-0.0210	-0.0274	-0.0333	-0.0114	-0.0136	-0.0171
	MSE	0.0044	0.0036	0.0020	0.0021	0.0017	0.0007
BB	Bias	-0.0332	-0.0597	-0.0814	-0.0230	-0.0380	-0.0537
	MSE	0.0045	0.0061	0.0075	0.0024	0.0028	0.0033
PoINAR(1), $\lambda = 5$		n=250			n=500		
Monte Carlo	Bias	-0.0076	-0.0149	-0.0101	-0.0053	-0.0063	-0.0071
	MSE	0.0040	0.0029	0.0006	0.0018	0.0014	0.0003
SB-INAR	Bias	-0.0165	**-0.0284**	**-0.0218**	**-0.0096**	**-0.0130**	**-0.0140**
	MSE	0.0041	**0.0034**	**0.0009**	**0.0018**	**0.0016**	**0.0005**
SB	Bias	**-0.0163**	-0.0284	-0.0231	0.0097	-0.0131	-0.0141
	MSE	**0.0041**	0.0034	0.0010	0.0018	0.0016	0.0005
BB	Bias	-0.0324	-0.0539	-0.0729	-0.0221	-0.0378	-0.0518
	MSE	0.0046	0.0062	0.0061	0.0022	0.0027	0.0030

Table 1: Bias and MSE for $\hat{\alpha}$. DGP: PoINAR(1). In bold the best performance with respect to MC.

of INAR(1) models the bias is always negative, but as sample size increases bias reduces. Moreover, bias increases with increasing the value of α but the MSE decreases. Finally, bias and MSE do not change with the different DGPs.

Initially, we notice that the SB for INAR models we propose, has almost always the best performance, even if the SB provides similar results. By increasing the sample size to $n = 500$ and for INAR models with Binomial and Negative Binomial innovations, the SB-INAR presents the best bias and MSE properties compared to MC simulations, for all considered cases.

In case of INAR(2) models, the bootstrap method still works well, even when the bootstrap bias and the MSE of parameter α_2 result greater than that of α_1. Moreover, the SB-INAR method presents always the best performance.

Block bootstrap seems to work worse with respect to sieve bootstrap, especially when the value of thinning parameter increases. This can be due to the fact that the performance of this method crucially depends on the choice of the block size and on the dependent structure of the process. To the best of our knowledge, the problem of choosing the block size is still an open question and it might be worth further investigation, but it is beyond the scope of this paper.

5 An empirical application

The time series under study is the monthly number of killed in motorway accidents in Italy, from January, 2001 to December 2014 and consists of 168 observations (source: ISTAT, http://dati.istat.it/Index.aspx?DataSetCode=DCIS_MORTIFERITISTR1#). The series has median 23, mean 26.43 and variance 151, hence the data are strongly over-dispersed. A plot of the series together with its empirical autocorrelation functions is shown in Fig. 1. The empirical autocorrelation functions of the series

BINAR(1), $m=6,\ \pi=0.2$		n=250			n=500		
Method		$\alpha=0.3$	$\alpha=0.6$	$\alpha=0.9$	$\alpha=0.3$	$\alpha=0.6$	$\alpha=0.9$
Monte Carlo	Bias	-0.0095	-0.0163	-0.0177	-0.0041	-0.0076	-0.0095
	MSE	0.0040	0.0031	0.0013	0.0021	0.0014	0.0005
SB-INAR	Bias	**-0.0187**	**-0.0302**	**-0.0337**	**-0.0088**	**-0.0146**	**-0.0181**
	MSE	**0.0042**	**0.0036**	**0.0021**	**0.0021**	**0.0015**	**0.0008**
SB	Bias	-0.0188	-0.0304	-0.0343	-0.0089	-0.0147	-0.0182
	MSE	0.0042	0.0037	0.0021	0.0021	0.0016	0.0008
BB	Bias	-0.0346	-0.0617	-0.0823	-0.0212	-0.0396	-0.0551
	MSE	0.0047	0.0064	0.0076	0.0023	0.0028	0.0035
BINAR(1), $m=6,\ \pi=0.5$		n=250			n=500		
Monte Carlo	Bias	-0.0082	-0.0125	-0.0188	-0.0018	-0.0060	-0.0093
	MSE	0.0037	0.0027	0.0013	0.0018	0.0014	0.0005
SB-INAR	Bias	**-0.0171**	**-0.0259**	**-0.0285**	-0.0063	**-0.0127**	**-0.0147**
	MSE	**0.0038**	**0.0031**	**0.0018**	0.0019	**0.0015**	**0.0007**
SB	Bias	-0.0173	-0.0264	-0.0298	**-0.0062**	-0.0129	-0.0153
	MSE	0.0038	0.0031	0.0019	**0.0018**	0.0015	0.0007
BB	Bias	-0.0332	-0.0581	-0.0838	-0.0189	-0.0377	-0.0548
	MSE	0.0045	0.0057	0.0079	0.0020	0.0026	0.0034
BINAR(1), $m=6,\ \pi=0.8$		n=250			n=500		
Monte Carlo	Bias	-0.0071	-0.0106	-0.0211	-0.0046	-0.0073	-0.0104
	MSE	0.0038	0.0028	0.0015	0.0020	0.0015	0.0006
SB-INAR	Bias	**-0.0159**	**-0.0241**	**-0.0328**	**-0.0090**	**-0.0141**	**-0.0172**
	MSE	**0.0039**	**0.0032**	**0.0021**	**0.0020**	**0.0016**	**0.0008**
SB	Bias	-0.0160	-0.0242	-0.0330	**-0.0090**	-0.0142	-0.0175
	MSE	0.0040	0.0032	0.0021	**0.0020**	0.0016	0.0008
BB	Bias	-0.0318	-0.0566	-0.0857	-0.0212	-0.0390	-0.0558
	MSE	0.0044	0.0055	0.0083	0.0023	0.0028	0.0036

Table 2: Bias and MSE for $\hat{\alpha}$. DGP: BINAR(1). In bold the best performance with respect to MC.

are coherent with that of a seasonal autoregressive model, in particular with a SARIMA$(2,0,0)(1,0,0)$ model which correspond to an AR(14) where all autoregressive coefficients between 3 and 11 are fixed to zero.

Thus, we apply bootstrap approach together with YW estimation method, to estimate the thinning parameters of the model:

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-2} + \ldots + \alpha_{14} \circ X_{t-14} + \epsilon_t.$$

The number of bootstrap replications is B=1000. Table 7 reports the results of the estimation together with the bootstrap $s.e.$

It is possible to see that the SB for INAR and SB give very similar results and the only significative parameters are α_1 and α_12, while for the BB approach only α_1 is significantly different from zero. In any case, the sum of the thinning parameters in lesser than 1, so the stationary condition is respected. Figure 2 reports the empirical autocorrelation functions of residuals and we can see that, even if it remains a little bit of seasonal correlation, SB-INAR and SB residuals can be considered incorrelated. Instead, BB residuals show again an evident seasonal correlation.

NBINAR(1), $r = 6,\ \pi = 0.2$		n=250			n=500		
Method		$\alpha = 0.3$	$\alpha = 0.6$	$\alpha = 0.9$	$\alpha = 0.3$	$\alpha = 0.6$	$\alpha = 0.9$
Monte Carlo	Bias	-0.0086	-0.0132	-0.0180	-0.0031	-0.0071	-0.0093
	MSE	0.0037	0.0028	0.0014	0.0018	0.0014	0.0005
SB-INAR	Bias	**-0.0172**	**-0.0261**	**-0.0220**	**-0.0075**	**-0.0136**	**-0.0117**
	MSE	**0.0038**	**0.0032**	**0.0015**	**0.0018**	**0.0015**	**0.0006**
SB	Bias	-0.0173	-0.0279	-0.0240	**-0.0075**	**-0.0136**	-0.0128
	MSE	0.0038	0.0033	0.0016	**0.0018**	**0.0015**	0.0006
BB	Bias	-0.0328	-0.0599	-0.0828	-0.0195	-0.0383	-0.0548
	MSE	0.0043	0.0059	0.0078	0.0020	0.0027	0.0034
NBINAR(1), $r = 6,\ \pi = 0.5$		n=250			n=500		
Monte Carlo	Bias	-0.0096	-0.0144	-0.0190	-0.0038	-0.0070	-0.0092
	MSE	0.0039	0.0028	0.0014	0.0018	0.0014	0.0005
SB-INAR	Bias	**-0.0183**	**-0.0276**	**-0.0267**	**-0.0082**	-0.0138	**-0.0136**
	MSE	**0.0040**	**0.0033**	**0.0018**	**0.0018**	0.0015	**0.0006**
SB	Bias	-0.0184	-0.0279	-0.0289	**-0.0082**	**-0.0137**	-0.0147
	MSE	0.0041	0.0033	0.0019	**0.0018**	**0.0015**	0.0007
BB	Bias	-0.0349	-0.0599	-0.0839	-0.0205	-0.0383	-0.0547
	MSE	0.0047	0.0059	0.0079	0.0021	0.0027	0.0034
NBINAR(1), $r = 6,\ \pi = 0.8$		n=250			n=500		
Monte Carlo	Bias	-0.0069	-0.0148	-0.0174	-0.0028	-0.0066	-0.0091
	MSE	0.0039	0.0028	0.0013	0.0017	0.0015	0.0005
SB-INAR	Bias	-0.0159	**-0.0283**	**-0.0301**	**-0.0072**	**-0.0132**	**-0.0160**
	MSE	0.0040	**0.0033**	**0.0019**	**0.0018**	**0.0016**	**0.0007**
SB	Bias	**-0.0158**	-0.0286	-0.0316	**-0.0072**	-0.0133	-0.0166
	MSE	**0.0039**	0.0033	0.0020	**0.0018**	0.0016	0.0007
BB	Bias	-0.0321	-0.0604	-0.0824	-0.0280	-0.0380	-0.0547
	MSE	0.0045	0.0059	0.0078	0.0020	0.0028	0.0034

Table 3: Bias and MSE for $\hat{\alpha}$. DGP: NBINAR(1). In bold the best performance with respect to MC.

PoINAR(2), $\lambda = 1$		n=250		n=500	
Method		$\alpha_1 = 0.5$	$\alpha_2 = 0.3$	$\alpha_1 = 0.5$	$\alpha_2 = 0.3$
Monte Carlo	Bias	0.0958	-0.1209	0.0972	-0.1091
	MSE	0.0236	0.0285	0.0217	0.0231
SB-INAR	Bias	**0.0957**	**-0.1429**	**0.0959**	**-0.1207**
	MSE	**0.0236**	**0.0339**	**0.0217**	**0.0256**
SB	Bias	0.1019	-0.1460	0.0993	-0.1226
	MSE	0.0247	0.0346	0.0223	0.0261
BB	Bias	0.0997	-0.1932	0.1019	-0.1630
	MSE	0.0243	0.0497	0.0228	0.0372
PoINAR(2), $\lambda = 5$		n=250		n=500	
Method		$\alpha_1 = 0.5$	$\alpha_2 = 0.3$	$\alpha_1 = 0.5$	$\alpha_2 = 0.3$
Monte Carlo	Bias	0.0917	-0.1197	0.0958	-0.1073
	MSE	0.0220	0.0280	0.0208	0.0236
SB-INAR	Bias	**0.1040**	**-0.1481**	**0.1032**	**-0.1230**
	MSE	**0.0245**	**0.0351**	**0.0223**	**0.0270**
SB	Bias	0.1418	-0.1711	0.1263	-0.1382
	MSE	0.0335	0.0416	0.0274	0.0306
BB	Bias	0.0947	-0.1910	0.1035	-0.1617
	MSE	0.0227	0.0489	0.0223	0.0377

Table 4: Bias and MSE for $\hat{\alpha}_1$ $\hat{\alpha}_2$. DGP: PoINAR(2). In bold the best performance with respect to MC.

BINAR(2), $m = 6,\ \pi = 0.2$					
Monte Carlo	Bias	0.0923	-0.1178	0.0981	-0.1129
	MSE	0.0227	0.0276	0.0213	0.0255
SB-INAR	Bias	**0.0933**	**-0.1403**	**0.0987**	**-0.1246**
	MSE	**0.0229**	**0.0329**	**0.0214**	**0.0282**
SB	Bias	0.1015	-0.1447	0.1031	-0.1272
	MSE	0.0243	0.0339	0.0223	0.0287
BB	Bias	0.0964	-0.1904	0.1045	-0.1667
	MSE	0.0235	0.0485	0.0226	0.0399
BINAR(2), $m = 6,\ \pi = 0.5$					
Monte Carlo	Bias	0.0963	-0.1216	0.0972	-0.1104
	MSE	0.0234	0.0283	0.0213	0.0242
SB-INAR	Bias	**0.1036**	**-0.1470**	**0.1017**	**-0.1242**
	MSE	**0.0248**	**0.0347**	**0.0222**	**0.0273**
SB	Bias	0.1276	-0.1611	0.1158	-0.1331
	MSE	0.0299	0.0384	0.0251	0.0294
BB	Bias	**0.0994**	-0.1928	0.1044	-0.1642
	MSE	**0.0238**	0.0495	0.0227	0.0385
BINAR(2), $m = 6,\ \pi = 0.8$					
Monte Carlo	Bias	0.0924	-0.1195	0.0978	-0.1100
	MSE	0.0228	0.0271	0.0229	0.0238
SB-INAR	Bias	**0.1053**	**-0.1480**	**0.1053**	**-0.1258**
	MSE	**0.0253**	**0.0343**	**0.0229**	**0.0274**
SB	Bias	0.1453	-0.1727	0.1298	-0.1420
	MSE	0.0348	0.0414	0.0285	0.0314
BB	Bias	0.0961	-0.1912	0.1045	-0.1635
	MSE	0.0234	0.0484	0.0229	0.0379

Table 5: Bias and MSE for $\hat{\alpha}_1$ $\hat{\alpha}_2$. DGP: BINAR(2). In bold the best performance with respect to MC.

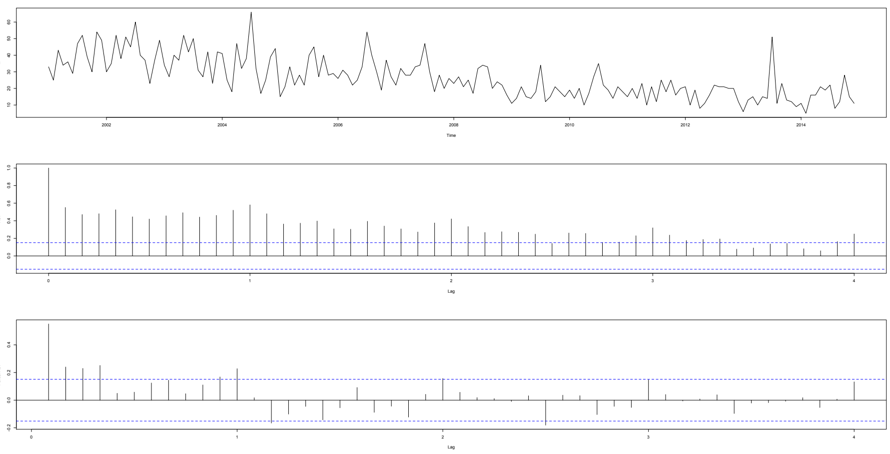

Figure 1: Plot of the number of killed in motorway accidents in Italy and its ACF and PACF.

NBINAR(2), $r = 6,\ \pi = 0.2$					
Monte Carlo	Bias	0.0941	-0.1170	0.0953	-0.1088
	MSE	0.0227	0.0278	0.0213	0.0233
SB-INAR	Bias	**0.1282**	**-0.1587**	**0.1144**	**-0.1320**
	MSE	**0.0302**	**0.0383**	**0.0253**	**0.0286**
SB	Bias	0.2179	-0.2215	0.1789	-0.1782
	MSE	0.0613	0.0603	0.0442	0.0423
BB	Bias	**0.0993**	-0.1899	0.1023	-0.1625
	MSE	**0.0238**	0.0488	0.0226	0.0380
NBINAR(2), $r = 6,\ \pi = 0.5$					
Monte Carlo	Bias	0.0948	-0.1176	0.0971	-0.1094
	MSE	0.0228	0.0275	0.0214	0.0239
SB-INAR	Bias	**0.1085**	**-0.1468**	**0.1047**	**-0.1252**
	MSE	**0.0256**	**0.0347**	**0.0229**	**0.0274**
SB	Bias	0.1490	-0.1722	0.1291	-0.1413
	MSE	0.0356	0.0418	0.0286	0.0313
BB	Bias	**0.0994**	-0.1906	0.1043	-0.1633
	MSE	**0.0238**	0.0487	0.0230	0.0380
NBINAR(2), $r = 6,\ \pi = 0.8$					
Monte Carlo	Bias	0.0917	-0.1172	0.0967	-0.1093
	MSE	0.0228	0.0275	0.0212	0.0239
SB-INAR	Bias	**0.9374**	**-0.1402**	**0.0981**	**-0.1214**
	MSE	**0.0231**	**0.0329**	**0.0214**	**0.0265**
SB	Bias	0.1036	-0.1454	0.1034	-0.1246
	MSE	0.0248	0.0341	0.0224	0.0272
BB	Bias	0.0951	-0.1896	0.1042	-0.1636
	MSE	0.0232	0.0483	0.0226	0.0382

Table 6: Bias and MSE for $\hat{\alpha}_1$ $\hat{\alpha}_2$. DGP: NBINAR(2). In bold the best performance with respect to MC.

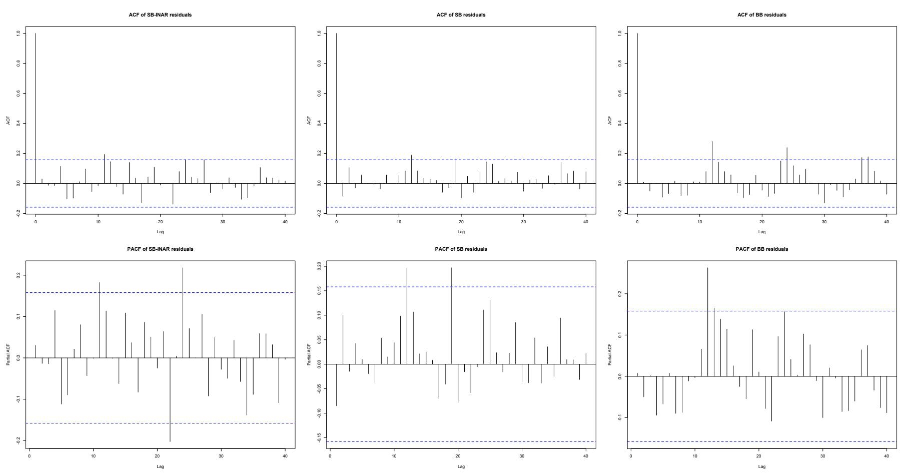

Figure 2: Empirical ACF and PACF of bootstrap residuals.

Table 7: Bootstrap estimation results *(standard error)*. In bold the estimates significantly different from zero.

Parameter	SB-INAR	SB	BB
α_1	**0.217**	**0.222**	**0.357**
	(0.079)	*(0.081)*	*(0.063)*
α_2	0.091	0.094	0.112
	(0.072)	*(0.073)*	*(0.079)*
α_3	0.039	0.038	0.087
	(0.052)	(0.051)	(0.069)
α_4	0.0.33	0.034	0.122
	(0.049)	*(0.048)*	*(0.071)*
α_5	0.037	0.0.038	0.034
	(0.050)	*(0.051)*	*(0.047)*
α_6	0.029	0.032	0.030
	(0.043)	*(0.047)*	*(0.050)*
α_7	0.032	0.033	0.048
	(0.046)	*(0.046)*	*(0.055)*
α_8	0.029	0.027	0.049
	(0.046)	*(0.043)*	*(0.056)*
α_9	0.032	0.034	0.027
	(0.046)	*(0.047)*	*(0.043)*
α_{10}	0.026	0.026	0.025
	(0.043)	*(0.042)*	*(0.041)*
α_{11}	0.028	0.032	0.034
	(0.042)	*(0.046)*	*(0.049)*
α_{12}	**0.187**	**0.188**	0.020
	(0.073)	*(0.075)*	*(0.039)*
α_{13}	0.047	0.046	0.014
	(0.053)	*(0.053)*	*(0.032)*
α_{14}	0.022	0.022	0.024
	(0.036)	*(0.039)*	*(0.038)*

6 Conclusion

In this work, we propose a new bootstrap approach to take into account the integer nature of count time series. In particular, our proposal is based on a modification of the well known sieve bootstrap for time series. With an extensive Monte Carlo experiment we show that our method works very well in all considered cases. Moreover, the bootstrap technique we propose is useful also to obtain coherent predictions for INAR(p) models once the thinning parameters have been estimated, but this is the topic of another paper.

References

Al-Osh, M. A. and A. A. Alzaid (1987). First order integer-valued autoregressive INAR(1) process. *Journal of Time Series Analysis 8*(3), 261–275.

Alzaid, A. A. and M. A. Al-Osh (1990). An integer-valued ph-order autoregressive structure INAR(p) process. *Journal of Applied Probability 27*, 314–323.

Bu, R., B. McCabe, and K. Hadri (2008). Maximum likelihood estimation of higher-order integer-valued autoregressive process. *Journal of Time Series Analysis 29*, 973–994.

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli 3*, 123–148.

Cardinal, M., R. Roy, and J. Lambert (1999). On the application of integer-valued time series models for the analysis of disease incidence. *Statistics in Medicine 18*, 2025–2039.

Du, J. G. and Y. Li (1991). The integer-valued autoregressive INAR(p) model. *Journal of Time Series Analysis 12*, 129–142.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics 7*, 1–26.

Freeland, R. K. and B. P. M. McCabe (2004). Analysis of low count time series data by poisson autoregression. *Journal of Time Series Analysis 25*, 701–722.

Kim, H. Y. and Y. Park (2008). A non-stationary integer-valued autoregressive model. *Statistical Papers 49*, 485–502.

Kim, H. Y. and Y. Park (2010). Coherent forecasting in binomial AR(p) model. *Communications for Statistical Applications and Methods 17*, 27–37.

Kreiss, J. (1992). Bootstrap procedures for ar(∞)-process. In K. J¨ckel, G. Rothe, and W. Sender (Eds.), *Bootstrapping and related techniques*. Springer, Heidelberg.

Kreiss, J. P. and S. N. Lahiri (2012). Bootstrap methods for time series. *Time Series Analysis: Methods and Applications 30*, 3–26.

Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics 17*, 1217–1241.

Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis 19*, 439–455.

Liu, R. Y. and K. Singh (1992). Moving blocks jackknife and bootstrap capture weak dependence. In R. LePage and L. Billard (Eds.), *Exploring the Limits of Bootstrap*. Wiley, New York.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin 21*, 645–650.

McKenzie, E. (1988). Some ARMA models for dependent sequences of poisson counts. *Advances in Applied Probability 20*, 822–835.

R Core Team (2015). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Sun, J. and B. McCabe (2013). Score statistics for testing serial dependence in count data. *Journals of Time Series Analysis 34*, 315–329.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it
Most of the working papers can also be found at the following url: http://wp.stat.unipd.it