

ON BOOTSTRAP IN NON-NEGATIVE AR(1) PROCESS

JIŘÍ HÁJEK

ABSTRACT. Various methods of the choice of the bootstrap sample size in the m out of n bootstrap for non-negative AR(1) process are presented as well as some numerical results for these methods. Then a smoothed bootstrap is proposed as a solution of problems that appear in the m out of n bootstrap and some numerical results show its potential.

Резюме. В данной статье изучается проблема оценки параметров в неотрицательных процессах авторегрессии методом "m-n" бутстрэп и показаны различные методы выбора объёма выборки. На числовых результатах видны проблемы этих методов. Как решение предложено гладкий бутстрэп и при помощи симуляции показаны его возможности.

1. INTRODUCTION

The bootstrap received quite a lot attention recently, because availability of fast computers makes it possible to use this technique in many applications. It have been used in the context of time series too, although it requires slightly different approach. There exists a number of methods that consistently bootstrap time series. In this paper we are interested in non-negative time series, which means that also innovations need to be non-negative. This brings serious problems and standard bootstrap methods cannot be used for consistent estimation of distribution of parameter estimates.

2. NON-NEGATIVE AR(1) PROCESS

Non-negative AR(1) process is defined as

$$(1) \quad X_t = \phi X_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, n,$$

where $0 < \phi < 1$ and $\varepsilon_t \geq 0$ is a strict white noise with a continuous distribution function F . A natural estimator of ϕ for this process is

$$(2) \quad \hat{\phi}_n = \min \left(\frac{X_1}{X_0}, \frac{X_2}{X_1}, \dots, \frac{X_n}{X_{n-1}} \right),$$

which was for the first time studied in [1]. This estimator is also the maximum likelihood estimator when innovations are exponentially distributed (with the density $f(y) = \lambda^{-1} \exp(-y/\lambda) \mathbb{I}[y > 0]$, where $\mathbb{I}[A]$ denotes the indicator of the event A).

We will focus on (2) in the rest of the article, because our main target will be to study possibilities of bootstrapping this estimator.

2000 *Mathematics Subject Classification.* 62M10.

Key words and phrases. Bootstrap, smoothed bootstrap, autoregressive processes, nonnegative time series.

This work was partially supported by Grants MSM 113200008 and GAČR 201/00/0769.

3. MOON BOOTSTRAP

As we know, bootstrap yields consistent estimates of distribution in very general class of problems. However there are reported problems in the literature, where the common bootstrap scheme is unable to give satisfactory results. An example of this problem is the estimator (2). An usual solution is to change the common bootstrap scheme to a bootstrap, where the bootstrap sample size (usually denoted m) is smaller than original sample size n . To simplify the notation we can refer to this bootstrap as *moon bootstrap* (where “moon” is an abbreviation of “**m** out of **n**”), as the authors of [4] do.

The first paper that discussed use of the moon bootstrap for non-negative autoregressive processes was [2], where authors presented a description of the method in case of non-negative AR(1) process together with the proof of consistency of this method. Later Feigin and Resnick in [3] extended the method to higher order non-negative autoregressive processes, again with the proof of consistency. In case of higher orders the simple estimator (2) must be replaced by linear programming estimators.

The whole method in the case of non-negative AR(1) process can be described as follows.

- (1) Estimate ϕ using the formula (2) for $\hat{\phi}_n$.
- (2) Calculate residuals by

$$\hat{\varepsilon}_i = X_i - \hat{\phi}_n X_{i-1}, \quad 1 \leq i \leq n.$$

- (3) Generate ε_i^* , $1 \leq i \leq m$ by a simple random sampling with replacement from $\{\hat{\varepsilon}_i, 1 \leq i \leq n\}$, where m is the bootstrap sample size which must satisfy $m = m(n) \rightarrow \infty$ and $m = o(n^\theta)$ for some θ depending on the distribution of innovations. Exact conditions can be found for example in [2] or [5].
- (4) Construct the bootstrap process X^* as

$$X_i^* = \hat{\phi}_n X_{i-1}^* + \varepsilon_i^*, \quad i = 1, 2, \dots, m,$$

where X_0^* can be chosen as 0, which was proposed in [2]. Although the effect of this starting value is asymptotically negligible, better results can be achieved with a value better approximating the stationary distribution of X_i . The easiest way to manage it is to set X_0^* as a randomly chosen value from $\{X_i : 1 \leq i \leq n\}$.

- (5) Calculate the estimate of ϕ for the bootstrap process as

$$\hat{\phi}_m^* = \min_{1 \leq i \leq m} \{X_i^*/X_{i-1}^*\}.$$

- (6) Repeat steps 3-5 B times.

Using this method we can approximate the distribution $T_n = \hat{\phi}_n - \phi$ by

$$(3) \quad T_n^* = \frac{a_n}{a_m} (\hat{\phi}_m^* - \hat{\phi}_n),$$

where a_n is defined as

$$(4) \quad a_n = F^{-1}(n^{-1}) = \inf\{x : F(x) \geq n^{-1}\}.$$

This result can be then used for example to remove the bias of $\hat{\phi}_n$, as is shown in [5] together with other results.

4. THE CHOICE OF BOOTSTRAP SAMPLE SIZE

The method described in the previous section needs for its application a selection of bootstrap sample size m . We have some asymptotical constrains for m , but they are of little use in practical applications.

One method for the choice of m was proposed in [2] exactly for non-negative AR(1) process. This is an ad-hoc method without any theoretical background and is based on the jackknife. In this method we repeat the bootstrap for various bootstrap sample sizes m . Then for each $1 \leq i \leq n$ the bootstrap as described above is made, but ε_k^* , $1 \leq k \leq n$ are resampled from $\{\hat{\varepsilon}_j, 1 \leq j \leq n, j \neq i\}$. Each of these bootstraps gives us a delete-one estimate $\hat{q}_{m,-i}$ of some quantity q , which can be for example the mean or a percentile. We denote by \hat{q}_m the ordinary moon bootstrap estimate of this quantity and define

$$(5) \quad L(m) = \sum_{i=1}^n (\hat{q}_{m,-i} - \hat{q}_m)^2,$$

which estimates a quantity proportional to the risk of the estimator \hat{q}_m . We select such m which minimizes $L(m)$ as the best bootstrap sample size and corresponding \hat{q}_m as an estimate of q .

Another possibility is to fix m approximately as \sqrt{n} , which is based on the empirical observations in [5] and seems to work well in a wide class of models.

Completely different approach of choosing bootstrap sample size was proposed in [4]. This method works generally for all moon bootstrap problems. The main idea consist in the fact that the distance between $L_n(P)$ and $\hat{L}_m(\hat{P}_n)$ (denoted Δ_m) is *stochastically equivalent* to the distance between $\hat{L}_m(\hat{P}_n)$ and $\hat{L}_{\frac{m}{2}}(\hat{P}_n)$ (denoted $\hat{\Delta}_m$). Here $L_n(P)$ is the true and unknown distribution while $L_m(\hat{P}_n)$ denotes a distribution estimated using m samples and the empirical distribution \hat{P}_n . As an optimal bootstrap sample size is then chosen such m which minimizes $\hat{\Delta}_m$.

In order to find out which of these proposed methods is the most suitable for choosing bootstrap sample size we made a numerical simulation of the problem which compares Kolmogorov's distance¹ of true and estimated distributions. The results presented in Tables 1 and 2 show that neither of the adaptive methods is better than the "fixed" method that always uses bootstrap sample size m approximately as \sqrt{n} . Although the adaptive methods are getting better with higher length of original time series n , it seems that the problem of choosing bootstrap sample size (not only in case of non-negative time series) is still an open issue and should be subject of further investigation. The jackknife method is also very time consuming and thus we can not suggest using it. By contrast the other adaptive method is quite fast and the results for longer time series are not so bad, so it could be useful in some cases.

Method	average distance	std.dev.	95% conf.interval
Jackknife	0.181	(0.088)	(0.171;0.191)
Adaptive	0.166	(0.079)	(0.157;0.175)
Fixed	0.145	(0.078)	(0.137;0.154)

Tab. 1 Average Kolmogorov's distance between bootstrap estimated and true distributions of $\hat{\phi}_n - \phi$ for exponential distribution of innovations and $n = 50$ (300 simulations).

¹Kolmogorov's distance of two distribution functions F and G is defined as $\sup_{x \in \mathbb{R}} |F(x) - G(x)|$

Method	average distance	std.dev.	95% conf.interval
Jackknife	0.134	(0.063)	(0.127;0.141)
Adaptive	0.130	(0.058)	(0.123;0.137)
Fixed	0.126	(0.063)	(0.119;0.133)

Tab. 2 Average Kolmogorov's distance between bootstrap estimated and true distributions of $\hat{\phi}_n - \phi$ for exponential distribution of innovations and $n = 100$ (300 simulations).

5. SMOOTHED BOOTSTRAP

We have seen in previous sections that a consistent method exists for bootstrapping non-negative autoregressive process. However we can go further and ask if the method is the best what we can achieve or if there is a better way. There are some obvious problems in the moon bootstrap method described earlier. The first is that we need to choose a proper bootstrap sample size m , which can be quite challenging problem as we saw in the previous section. Then also the distribution of innovations is needed because of normalizing constant in (3) and that limits the method a lot. Overall performance of the method can be discussed as well.

In order to get further we need to study what actually makes problems in the bootstrap. Let us take the ordered estimated innovations of the process and denote them $\hat{\varepsilon}_{[1]}, \hat{\varepsilon}_{[2]}, \dots, \hat{\varepsilon}_{[n]}$. As we can easily see, the smallest estimated innovation $\hat{\varepsilon}_{[1]}$ is always equal to 0. Then, when we set bootstrap sample size m equal to n , use $\hat{\varepsilon}_{[1]}$ for resampling and define the distribution $T_n^* = \hat{\phi}_n - \hat{\phi}_n^*$, we can easily prove that

$$(6) \quad P^*(T_n^* = 0) = 1 - \left(1 - \frac{1}{n}\right)^n \rightarrow 1 - e^{-1}.$$

However that means that this method is inconsistent, because as in a continuous distribution $P(T_n^* = 0)$ must equal 0. So we must leave out $\hat{\varepsilon}_{[1]}$ (and possibly others that equal zero) from the resampling scheme, but then our results are biased. These problems with consistency can be solved by employing the moon bootstrap, but it means that other problems described earlier appear.

This analysis suggests that our problem is in the discrete behaviour of estimated innovations. Therefore we would like to replace the empirical distribution function, that is usually used in the context of bootstrap for resampling purposes, by a distribution function of a continuous distribution. That is the reason we get to the smoothed bootstrap where exactly a continuous distribution is used for resampling instead of a discrete one. So the task now is to make a non-parametric (or semi-parametric) estimation of the distribution of innovations ε_t . Estimated density function (or corresponding quantile function) of such distribution is used for generation of ε_i^* , which can be then used for construction of the bootstrap process X^* and calculation of corresponding estimate $\hat{\phi}_n^*$ as is described in the moon bootstrap procedure.

We have a wide selection of possible methods which can make non-parametric estimation of distribution function F (or a density f). Usually kernel estimates are used, but in our case there is a problem with density cut off at 0, so a different method which can handle this problem would be useful. A promising family of method seems to be the methods based on splines, particularly a logspline method.

The logspline method we use is described in [6] and [7] and its main idea consists in that logarithm of density function is modeled by cubic splines. This objective is achieved in the following way. Let t_1, \dots, t_K be a sequence of real numbers such that $-\infty < t_1 < \dots < t_K < \infty$, where K is an integer satisfying $K \geq 4$. Let S_0 denote a group of twice continuously differentiable functions s on \mathbb{R} such that the

restriction of s to each of the intervals $(-\infty, t_1), (t_1, t_2), \dots, (t_{K-1}, t_K), (t_K, \infty)$ is a cubic polynomial. Functions s from S_0 are called cubic splines with knots t_1, \dots, t_K , where the third derivative of s is not continuous. Let S denote a subspace of S_0 such that $s \in S$ is linear on $(-\infty, t_1)$ and on (t_K, ∞) . S has a basis $1, B_1, \dots, B_{K-1}$ and we further choose B_1 as a linear function with negative slope on $(-\infty, t_1)$, B_{K-1} as a linear function with positive slope on (t_K, ∞) , B_2, \dots, B_{K-1} as constants on $(-\infty, t_1)$ and B_1, \dots, B_{K-2} as constants on (t_K, ∞) . Then the density function is defined as

$$(7) \quad f(y, \theta) = \exp(\theta_1 B_1(y) + \dots + \theta_{K-1} B_{K-1}(y) - c(\theta)), \quad y \in \mathbb{R},$$

where $c(\theta)$ is a normalizing constant defined as

$$(8) \quad c(\theta) = \log \left(\int_{\mathbb{R}} \exp(\theta_1 B_1(y) + \dots + \theta_{K-1} B_{K-1}(y)) dy \right),$$

and $\theta = (\theta_1, \dots, \theta_{K-1})' \in \Theta$ is a vector of real parameters with $\theta_1 < 0$ and $\theta_{K-1} < 0$. Then the condition $\int_{\mathbb{R}} f(y; \theta) dy = 1$ is clearly fulfilled. $f(\cdot; \theta), \theta \in \Theta$ creates a logspline family of density functions.

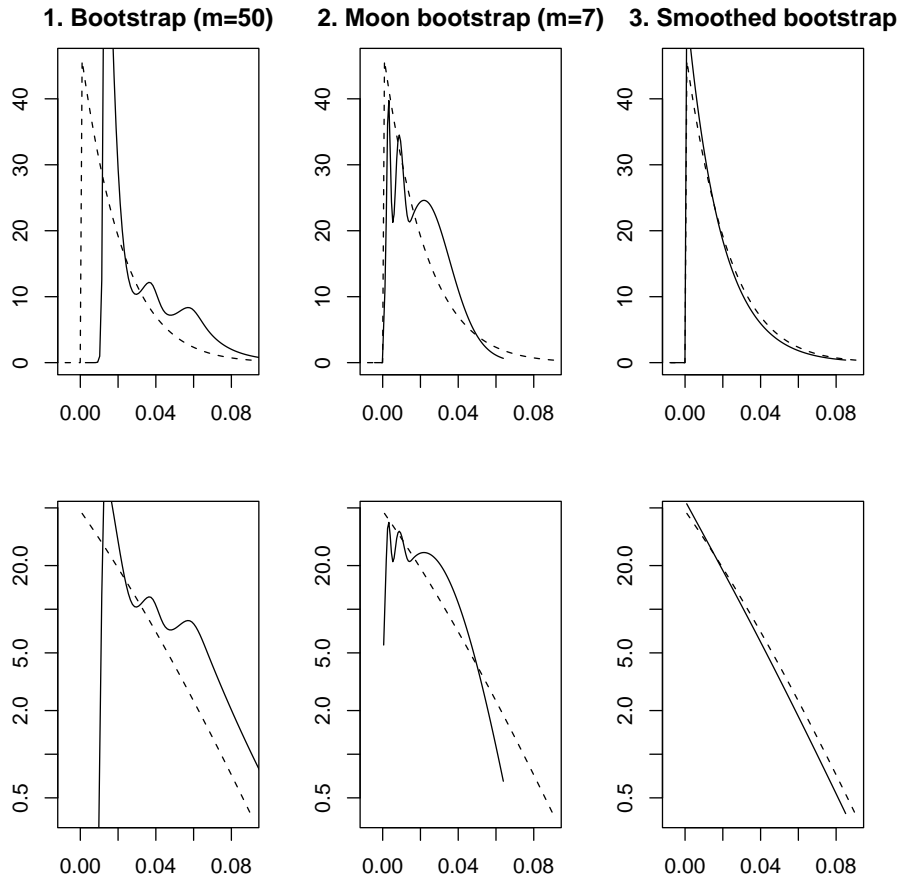


Fig. 1 Comparison of three different bootstrap methods for exponentially distributed innovations.

Maximum-likelihood estimate $\hat{\theta}$ of θ is obtained by maximizing the log-likelihood function using modified Newton–Raphson algorithm. Other details of the method, like handling tails of distributions and selecting knots count and position, can be found in [6].

In order to find out potential of the smoothed bootstrap for non-negative autoregressive models, we made a numerical simulation that compares it with the moon bootstrap. There are results presented in Figures 1 and 2. All the simulations are made with a series of length 50 and with ϕ chosen to be 0.5. The exponential distribution of innovations with $\lambda = 1$ is used in Figure 1 while uniform $R(0, 2)$ is used in Figure 2. The dashed line represents theoretical distribution $T_{50} = \hat{\phi}_{50} - \phi$ simulated by Monte-Carlo method and solid lines represent results of each bootstrap method used for estimation of T_{50} . In the first column the theoretical result is compared with ordinary bootstrap where bootstrap sample size is chosen the same as original sample size, i.e. $m = 50$. In the second column is the result of moon bootstrap with $m = 7$ and in the third column is the result of smoothed bootstrap with logspline semi-parametric density estimation used as implemented in R language. There are estimated density plots in the upper panel and to emphasize details logarithm of density is used in the lower panel.

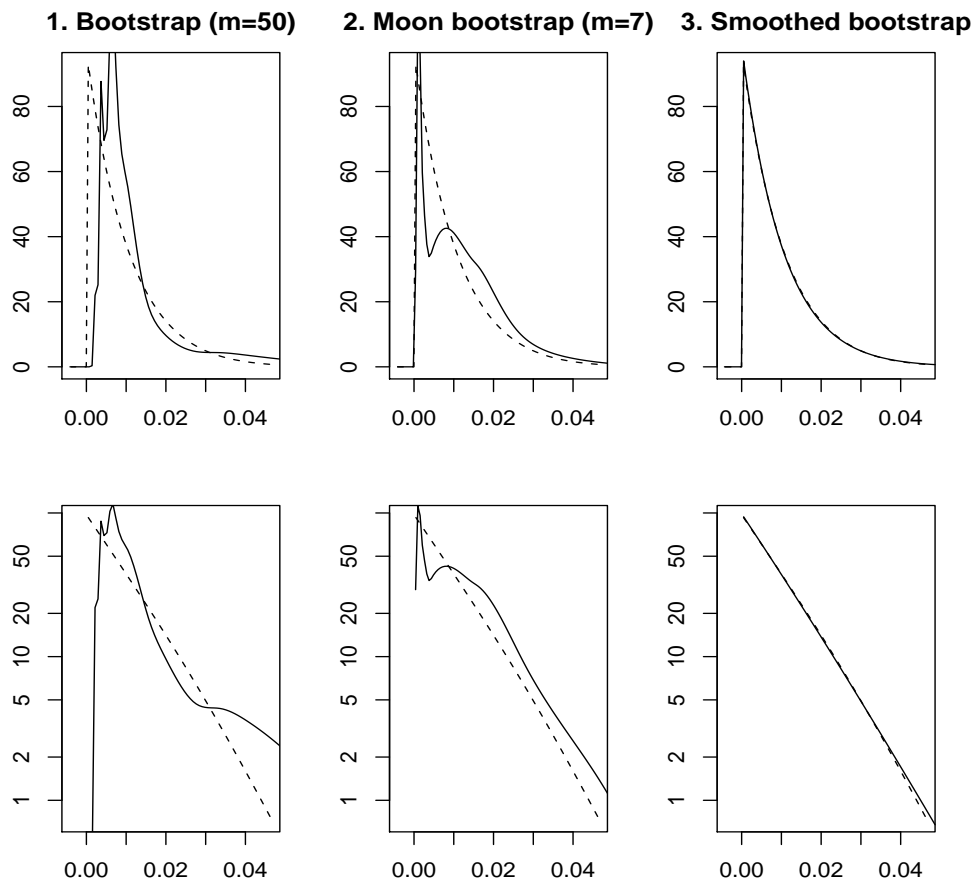


Fig. 2 Comparison of three different bootstrap methods for uniformly distributed innovations.

The results in Figures 1 and 2 are quite nice. However these results are based just on one simulation which can be misleading. Therefore results of another simulation are presented in Table 3, where 1000 repetitions make the results more informative.

	Moon bootstrap ($m = 7$)	Smoothed bootstrap
Bias corrected $\hat{\phi}$	0.49921	0.50045
95% conf.interval for bias corrected $\hat{\phi}$	(0.49857; 0.49987)	(0.49979; 0.5011)
Average Kolmogorov's distance from the true distr. of $\hat{\phi} - \phi$	0.1384	0.0896

Tab. 3 Results of 1000 bootstrap simulations for the exponential distribution of innovations and $n = 50$.

Results of all these simulations are very promising, because the smoothed bootstrap approximates the theoretical distribution far better than the moon bootstrap does. Certainly more theoretical investigation is necessary in order to state properties of the method. Some more numerical studies are needed as well to prove its abilities to work in practical applications.

6. CONCLUSION

We have seen that an adaptive choice of bootstrap sample size for the m out of n bootstrap is still a challenging and unsolved issue. It was shown by a numerical study that none of the methods performs really well and so none of them can be recommended for practical use. The smoothed bootstrap method, that was presented later, gives very promising results, which need to be further studied and particularly a deeper theoretical background should be created.

REFERENCE

- [1] C.B. Bell and E.P. Smith. Inference for non-negative autoregressive schemes. *Comm. Statist. Theory Methods*, 15:2267–93, 1986.
- [2] S. Datta and W. McCormick. Bootstrap inference for a first-order autoregression with positive innovations. *Journal of the American Statistical Association*, 90(432):1289–1300, 1995.
- [3] P.D. Feigin and S.I. Resnick. Linear programming estimators and bootstrapping for heavy tailed phenomena. 1995.
- [4] F. Götze and A. Račkauskas. Adaptive choice of bootstrap sample size. Working paper, 2001.
- [5] J. Hájek. Non-negative time series. *WDS'01*, pages 62–67, 2001.
- [6] C. Kooperberg and C. J. Stone. A study of logspline density estimation. *Computational Statistics & Data Analysis*, 12:327–347, 1991.
- [7] C. Kooperberg and C. J. Stone. Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*, 1:301–328, 1992.

CHARLES UNIVERSITY IN PRAGUE, DEPARTMENT OF STATISTICS, SOKOLOVSKÁ 83, CZ–186 75 PRAGUE 8–KARLÍN, CZECH REPUBLIC