Chap 2 : Semiparametric Bootstrap for time series

(Running Title: **Semiparametric Bootstrap(s) for time series.**)

by

Patrice Bertail[*]

CREST, Laboratoire de Statistique

Key Words and Phrases: Bootstrap, semiparametric, time series model, sieve approximation, second order theory.

[*]email : Patrice.Bertail@ensae.fr

1 Introduction

The naive bootstrap is essentially designed for i.i.d. r.v.'s. Of course in most case of dependent data, it fails to respect the structure of the dependence and thus fails to be asymptotically valid. Adaptations to take into account the dependence structure are however possible in many situations. One more time, the adequate solution highly depends on the hypotheses and the model, considered for the data generating process. In this chapter, we will see that if the data can be modeled (in a semiparametric way) by some structural model, with some underlying i.i.d. structure (just like in example 3 of chap. 1, the linear model) then it is easy to propose a semiparametric version of the bootstrap, which will somehow enjoy the same properties as in the i.i.d. case. We then focus essentially on the autoregressive case to give some very basic ideas and point out the problems, which may appear in this setting. We will in particular discuss the problems encountered in the non-stationary case. We will also briefly discuss how these kind of ideas may be used in more general models via the notion of sieve approximations.

2 Failure of the naive bootstrap for 1-dependent data?

We illustrate the failure of the naive bootstrap for the classical moving average. P will now denote the distribution of the whole process.

Consider the stationary moving average process $MA(1)\ :\ X_t = \varepsilon_t + \theta\varepsilon_{t-1}$, where the ε_t are i.i.d. $Q\ \in \mathbb{P}^0$, then we have $V(X_t) = 1+\theta^2$ and $Cov(X_t, X_t+1) = \theta$. Assume that we observe a stretch of observations from this process say $(X_1, ..., X_n)$,a CLT for $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$, estimating $E_P X_t = 0$ (by stationarity, this is an unbiased estimator) may be easily obtained. We have in that case

$$\sqrt{n}(\overline{X}_n - E_P X_t) \underset{n\to\infty}{\to} N(0, \sigma_\infty^2),$$

with $\sigma_\infty^2 = (1+\theta^2+2\theta) = (1+\theta)^2$. Denote by $K_n(., P)$ the distribution of $\sqrt{n}(\overline{X}_n - E_P X_t)$. Notice that for $\theta = -1$, the normal distribution degenerates since we have $\overline{X}_n = \frac{1}{n}(\varepsilon_n - \varepsilon_0)$: in that case, the right standardization is rather n, we thus have in this simple context an equicontinuity problem at $\theta = -1$.

The naive bootstrap would consist in generating $X_i^*\ i.i.d.\ P_n^1,\ \ P_n^1 = \frac{1}{n}\sum_{j=1}^n \delta_{X_j}\ ,\ i = 1, ...n$. Notice that P_n^1 is an estimator of the marginal distribution $P^1 = P_{X_1}$ but certainly not of P. It is easy to see that the variance of these variables are given by

$$Var^*(X_1^*) = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \underset{n\to\infty}{\longrightarrow} Var_P(X_t) = 1+\theta^2$$

The Bootstrap distribution $K_n(., P_n^1)$ estimates actually $K_n(., P^1)$ that is the distribution of the mean of i.i.d. variables with mean 0 and variance $(1+\theta^2)$ and we have thus (using for instance the standard Berry-Esséen theorem used in chap. 1)

$$||K_n(., P_n^1) - \Phi(./(1+\theta^2))|| \to 0 \ \ a.\ s.$$

It follows that

$$||K_n(., P_n^1) - K_n(., P)||_\infty \nrightarrow 0,$$

as soon as $\theta \neq 0$. This is not astonishing since we are using an estimator $K_n(., P_n^1)$, which has nothing to do with the original problem... Notice however that if we work with correctly studendized statistics, then we have

$$||L_n(., P_n^1) - \Phi(.)|| \to 0 \quad a.\ s.$$

Here $L_n(., P_n^1)$ is the studentized bootstrap distribution (using an i.i.d. bootstrap)

$$L_n(., P_n^1) = \Pr{}^*\{n^{1/2}(\overline{X}_n^* - \overline{X}_n)/S_n^* \leq x | (X_1, ..., X_n)\},$$

$$S_n^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i^* - \overline{X}_n^*)^2.$$

We may as well standardize the original empirical mean and put

$$L_n(x, P) = \Pr\{n^{1/2}(\overline{X}_n^* - \overline{X}_n)/\widehat{\sigma}_{\infty,n} \leq x\},$$

with, for instance,

$$\widehat{\sigma}_{\infty,n}^2 = 1 + \widehat{\theta}_n^2 + 2\widehat{\theta}_n.$$

It follows that

$$||L_n(., P_n^1) - L_n(., P)||_\infty \to_{n\to\infty} 0\ ,\ a.s.,$$

that is the correctly studentized naive bootstrap asymptotically works for dependent data! Actually this result is not very interesting in that we completely loose here the automaticity of the bootstrap procedure : for more general functionals and for more general data generating process, it will not be so easy to find an adequate standardization. Standardization will actually be a key point for obtaining good results for dependent data. Moreover, this procedure is somehow totally artificial, even if we apply the i.i.d bootstrap to r.v.'s, which have nothing to do with the original data, for instance the weekly production of strawberries, we will get the same results. It is thus expected that the finite sample properties of this studentized bootstrap will not be very good and will have the same properties (or be even worse) than the asymptotic distribution.

To solve this problem, we recast it in a semiparametric framework. The problem is to find here an estimator of the distribution P of the whole process which makes sense in the model of interest. The original model may be seen as a simple semiparametric model indexed by θ and the distribution $Q \in \mathbb{P}^0$ say

$$P_{\Theta,\mathbb{P}_O} = \{P_{\theta,Q}^{MA(1)},\ \theta \in \Theta,\ Q \in \mathbb{P}_0\},$$

where $P_{\theta,Q}^{MA(1)}$ is the full distribution of a stationary MA(1) process, which has parameter θ and residuals i.i.d Q.

Just like in the parametric case, we want an estimator of $P_{\theta,Q}^{MA(1)}$. Take for instance, $\widehat{\theta}_n = \frac{1}{n}\sum_{t=1}^{n-1} X_t X_{t+1}$, we can as well estimate the residuals for instance by considering a truncated version of $(1-\theta B)^{-1}$ (at least if $\theta \neq -1$) and standardize these residuals so that they belong to $\mathbb{P}_0$, this leads to an estimator Q_n in $\mathbb{P}_0$. The bootstrap rule is thus now to study all the quantity of interest under $P_{\widehat{\theta}_n,Q_n}^{MA(1)}$. From a practical point of view, this means that we are now considering pseudo-arrays of $MA(1)$ process of the form

$$X_{t,n}^* = \varepsilon_{t,n}^* + \widehat{\theta}_n \varepsilon_{t-1,n}^*, \text{ with } \varepsilon_{t,n}^* \text{ i.i.d. } Q_n.$$

It can be easily seen that this semiparametric bootstrap method is going to work. For instance if we choose the DM_2 metric (Wasserstein or Mallows metric) in the space of distributions $K_m(.,P_{\theta,Q}^{MA(1)})$ then we have for any i.i.d couples $(\varepsilon_t^*, \varepsilon_t)$ with respective marginal distributions Q and Q^*,

$$\begin{aligned}
& DM_2(K_m(.,P_{\theta,Q}^{MA(1)}), K_m(.,P_{\theta^*,Q^*}^{MA(1)}))^2 \\
& \leq nE\left(n^{-1}\sum_{t=1}^{n}(\varepsilon_t^* + \theta^*\varepsilon_{t-1}^*) - n^{-1}\sum_{t=1}^{n}(\varepsilon_t + \theta\varepsilon_{t-1})\right)^2 \\
& \leq 2\{E\left(\varepsilon_t^* - \varepsilon_t\right)^2 + E\left(\theta^*\varepsilon_{t-1}^* - \theta\varepsilon_{t-1}\right)^2\} \\
& \leq 2\{E\left(\varepsilon_t^* - \varepsilon_t\right)^2 + E\left(\theta^*\varepsilon_{t-1}^* - \theta\varepsilon_{t-1}^* + \theta\varepsilon_{t-1}^* - \theta\varepsilon_{t-1}\right)^2\} \\
& \leq 4\{E\left(\varepsilon_t^* - \varepsilon_t\right)^2 + \theta^2 E(\varepsilon_{t-1}^* - \varepsilon_{t-1})^2 + Var(\varepsilon_t^*)|\theta_* - \theta|^2\}.
\end{aligned}$$

That is by taking the infimum over all r.v.$'s$ $(\varepsilon_t^*, \varepsilon_t)$, we get

$$DM_2(K_m(.,P_{\theta,Q}^{MA(1)}), K_m(.,P_{\theta^*,Q^*}^{MA(1)}))^2 \leq 4(1+\theta^2)DM_2(Q^*,Q) + 4|\theta_* - \theta|^2$$

In particular, it is easy to prove that, for any θ,

$$4(1+\theta^2)DM_2(Q_n,Q) + 4|\widehat{\theta}_n - \theta|^2 \ \to 0\ ,\ (a.s.\ or\ pr.),$$

provided that we can find estimators such that $\widehat{\theta}_n \to \theta,\ Q_n \to_w Q$ and $\int x^2 dQ_n \to \int x^2 dQ$ (a.s. or pr.) as $n \to \infty$. Notice as well, that we only need a convergent estimator of θ (without any specific rate) to get the asymptotic validity of the bootstrap. It can also be shown that this specific form of bootstrap is second order correct under additional moment and Cramer conditions on the ε_t.

One should however be careful with this result. Indeed if $\theta \neq 1$, then the limiting distribution of $\sqrt{n}(\overline{X}_n - E_P X_t)$ is non degenerated, thus we can deduce as well that the bootstrap distribution will be asymptotically gaussian. But in the case $\theta = -1$, the distribution of $\sqrt{n}(\overline{X}_n - E_P X_t)$ will be degenerated, the asymptotic result that we obtain means that the bootstrap distribution will be degenerated as well, not that the bootstrap distribution is asymptotically correct (with another normalization). Actually it can be shown that the bootstrap fails to give a correct approximation at the point of hyperefficiency $\theta = -1$.

3 Model based Bootstrap for autoregressive processes.

All the ideas described in the preceding example may be applied to much more general models, provided that we have a semiparametric description of the model. Typically, if the process is characterized by some semiparametric model $P_{\Theta,\mathbb{P}_O} = \{P^M_{\theta,Q},\ \theta \in \Theta,\ Q \in \mathbb{P}_0\}$, then once again, we may try to find estimators $(\widehat{\theta}_n, Q_n)$ of the parameters θ and of the underlying probability Q. The model based bootstrap for time series will consist in studying the distribution of the statistics of interest under $P^M_{\widehat{\theta}_n,Q_n}$. This approach was iniated by Efron (1981) and developped in Freedman (1984) in some stationary models.

We will give only a few results concerning this kind of bootstrap in the case of AR models, with a given order. In this particular case, the procedure is very easy to implement, since the value of the AR model are defined recursively.

Let $\{X_t,\ t \in \mathbb{Z}\}$ be an AR(p) model of the form

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \varepsilon_t,$$

with ε_t i.i.d Q with mean $E_Q \varepsilon_t = 0$ and variance $V_Q \varepsilon_t = \sigma^2$. We denote $\phi = (\phi_1, ..., \phi_p)$. We assume that we observe the sequence $(X_1, ..., X_n)$ from this process.

3.1 The stationary case

We now assume that all the roots of the polynomial

$$P(x) = 1 - (\phi_1 x + \phi_2 x^2 + ... + \phi_p x^p)$$

are outside the unit circle, then $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \varepsilon_t$ admits a unique stationary solution with distribution $P_{\phi,Q}$

The least-square estimator $\widehat{\phi}_n$ of ϕ is an asymptotically convergent estimator of ϕ. We also obtain the estimated residuals

$$\widehat{\varepsilon_t} = X_t - (\widehat{\phi}_1 X_{t-1} + \widehat{\phi}_2 X_{t-2} + ... + \widehat{\phi}_p X_{t-p}),\ t = p+1, ..., n,$$

and the centered residuals

$$\widetilde{\varepsilon_t} = \widehat{\varepsilon_t} - (n-p)^{-1} \sum_{i=p+1}^{n} \widehat{\varepsilon_t}.$$

Define the empirical distribution of the recentered residuals

$$Q_n = (n-p)^{-1} \sum_{i=1}^{n} \delta_{\widetilde{\varepsilon}_t}.$$

Let Σ_n be the usual l.s.e. estimator of the asymptotic variance Σ of $\widehat{\phi}_n$. We are mainly interesting in this framework in bootstrapping the statistics $\widehat{\phi}_n$. Denote

$$K_n(x, P_{\phi,Q}) = \Pr{}_{P_{\phi,Q}}\{n^{1/2}\Sigma^{-1/2}(\widehat{\phi}_n - \phi) \le x\},$$

$$L_n(x, P_{\phi,Q}) = \Pr{}_{P_{\phi,Q}}\{n^{1/2}\Sigma_n^{-1/2}(\widehat{\phi}_n - \phi) \le x\}.$$

Then we have the following results due to Bose (1988).

Theorem 1 *Bose(1988). Stationary case*

Assume that ε_t is such that $E_P \varepsilon_t^8 < \infty$, and that the following Cramer condition holds

$$\overline{\lim}_{t=(t_1,t_2),||t||->\infty} |E_Q \exp(it'(\varepsilon_t, \varepsilon_t^2))| < 1,$$

then

$$||K_n(x, P_{\phi,Q}) - K_n(x, P_{\widehat{\phi}_n,Q_n})||_\infty = O_P(n^{-1/2}),\ as\, n \to \infty. \tag{1}$$

If, in addition $E_P \varepsilon_t^{16} < \infty$, then

$$||L_n(x, P_{\phi,Q}) - L_n(x, P_{\widehat{\phi}_n,Q_n})||_\infty = o_P(n^{-1/2}),\ as\, n \to \infty. \tag{2}$$

Proof. The proof of this kind of result relies heavily on the Edgeworth expansion for strong mixing time series with exponential decreasing rate proved by Götze and Hipp (1983), which will be given in the next chapter. Bose (1988) actually proved essentially 1 as well a version for a version of K standardized by the true variance. A close look at its proof shows that the same arguments also hold for the studentized case provided that one has additional moment conditions on the ε_t. The moment conditions given here are probably not optimal. Some additional refinements show (at least in the easier case 1) that the rate of convergence may be improved to $O_P(n^{-1})$ as in the i.i.d. case (see Choi and Hall, 2000). ■

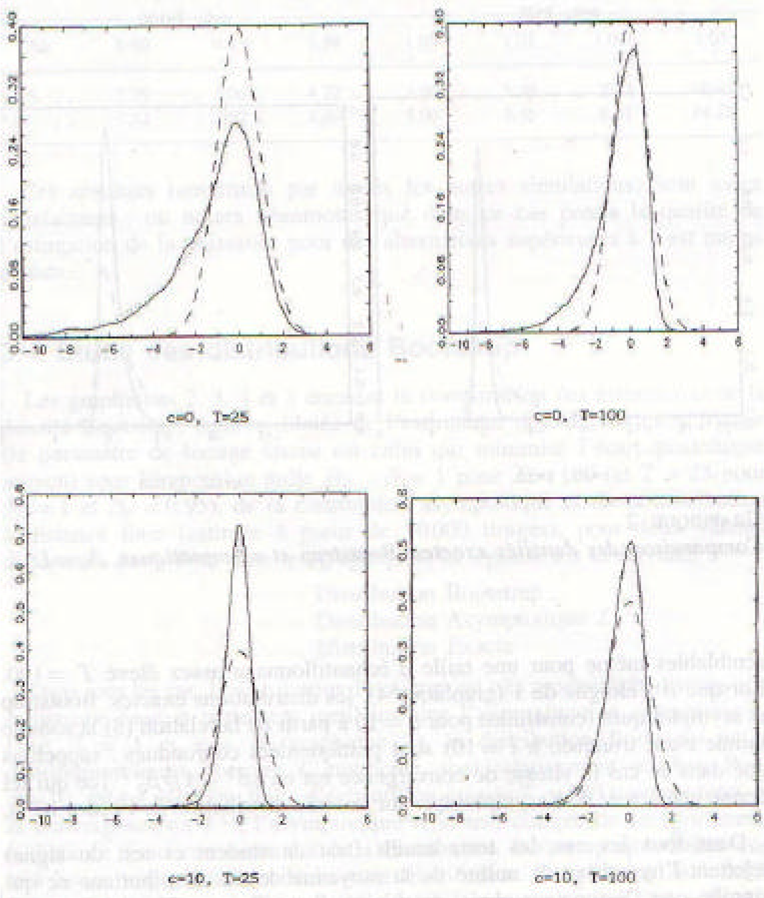
The following graphic taken from Bertail (1994) shows that in the case of an AR(1) model with $\phi = 0.95,\ X_0 = c$ and $\sigma = 1$, the improvement brought by the bootstrap distribution may be quite important for small and moderate sample sizes. The true distribution (denoted by ...) , the bootstrap distribution (resp. _) are almost the same, whereas the asymptotic distribution (resp. - - -) is very far from the true one. This is clearly due to the important asymmetry of the distributions especially when $X_0 = c$ is close to 0.

Remark 1 : This result shows that this semi-parametric bootstrap has somehow the same properties as the usual bootstrap in the i.i.d. case. This is of course due to the particular structure of the model. From a practical point of view, there is however a difficulty in generating a stationary time series from the process with distribution $P_{\widehat{\phi}_n,Q_n}$. Indeed the result is valid for a **stationary** process X_t^* having distribution $P_{\widehat{\phi}_n,Q_n}$. If one uses simply the recursion,

$$X_t^* = \phi_1 X_{t-1}^* + \phi_2 X_{t-2}^* + ... + \phi_p X_{t-p}^* + \varepsilon_t^*,$$

with ε_t^* i.i.d. Q_n and initial values $X_1^* = X_1,\ X_2^* = X_2,\ ...X_p^* = X_p$ (or $X_1^* = 0,\ X_2^* = 0,\ ..., X_p^* = 0$), there might be some stationary problem at the beginning of the time-series. This non-stationary effect vanishes if we wait a long time before considering a stretch of length n."How long?" is another question... We will come back later to this problem, when studying the Bootstrap for general Harris recurrent Markov chains. This problem also highlights the fact that the hypothesis of strong stationarity (or homogeneity for random fields) is a strong hypothesis. We can in many practical situations, assume that the underlying process is almost stationary and tends to a stationary process. We will see in the last chapter that if a process is asymptotically stationary with a bit of non stationarity (for instance due to the initial values), the non-stationarity may have a dramatic influence on the second order terms.

Remark 2 : Asymptotic results may also be generalized to other non-linear models. See for instance Franke and al (2002), for results for non-linear models.



GRAPHIQUE
Comparaison des densités exactes, Bootstrap et asymptotiques, $\beta_0 = 0.95$.

3.2 The non stationary and explosive case

When the AR model is such that the polynomial has one or several unit roots then we typically have an equicontinuity problem. Indeed it can be shown that in that case the process is non stationary and that the distribution of the least square (at least for one of its component) is non standard. One more time the bootstrap (here in its semiparametric form) will fail and some adaptation are necessary. To explain this phenomenon, we will focus on the simpler AR(1) model and also consider the explosive case.

Consider $(X_1, ..., X_n)$ a realization of the process $\{X_t\}_{t \in \mathbb{Z}}$,

$$X_t = \phi X_{t-1} + \varepsilon_t, \text{ with } \varepsilon_t \text{ i.i.d. } Q \in \mathbb{P}^0,$$

with

$$X_0 = c.$$

The distribution of $\{X_t\}_{t \in \mathbb{Z}}$ will be denoted $P_{\phi,c,Q}$ (here we need to specify the dependence in c, because the limiting distribution in the explosive case will depend on c). The asymptotic behaviour of the least square estimator $\widehat{\phi}_n$ is very different whether $|\phi| < 1$, $|\phi| = 1$ $|\phi| > 1$, both in term of rate of convergence and limit distribution. It can be easily shown that as $n \to \infty$,

$$\text{if } |\phi| < 1 \text{ then, with } \tau_n(\phi) = \frac{n^{1/2}}{1-\phi^2}, \; \tau_n(\phi)(\widehat{\phi}_n - \phi) \to N(0,1)$$

$$\text{if } |\phi| = 1 \text{ then, with } \tau_n(\phi) = n, \quad \tau_n(\phi)(\widehat{\phi}_n - \phi) \to sign(\beta) K_1$$

$$K_1 = \frac{1}{2}(W^2(1) - 1) / \int_0^1 W^2(t),$$

$$\text{if } |\phi| > 1 \text{ then, with } \tau_n(\phi) = \frac{\phi^T}{(\phi^2 - 1)}, \; \tau_n(\phi)(\widehat{\phi}_n - \phi) \to K_{\phi,c}.$$

See Dickey and Fuller (1979) Phillips (1987), for the unit root case and White (1958, 1959) for the explosive case. The limiting distribution K_1 has a distribution $K_1(x)$ which is highly asymmetric. In the explosive case, the rate of convergence is geometric and depends on the true value of the parameter. The limiting distribution may have very fat tail. For instance when the residuals are i.i.d. $N(0,1)$ the density of $K_{\beta,c}$. is given by

$$f_{\beta,c} = \frac{\exp(-q)}{\pi(1+x^2)} \left(1 + \sum_{i=1}^{\infty} \frac{(4q)^i (i!)}{(1+x^2)^i (2i!)} \right),$$

with $q = c^2(\phi^2 - 1)/2$. If $c = 0$ then the limiting distribution is Cauchy. Moreover, it should also be noticed that in the explosive case the limiting distribution is very sensitive to the first value $X_0 = c$ of the time series. In the following we will be essentially interested in estimating the distribution of the l.s.e correctly standardized by the rate $\tau_n(\phi)$

$$K_n(x, P_{\phi,c,Q}) = \Pr{}_{P_{\phi,c,Q}}\{\tau_n(\phi)(\widehat{\phi}_n - \phi) \le x\}.$$

It should be noticed that the rate of convergence $\tau_n(\phi)$ may be replaced by the scaling factor $\Sigma_n^{-1/2}$ $=(\sum_{i=1}^n X_t^2)^{1/2}$ which is asymptotically equivalent to $\tau_n(\phi)$, so that we may also be interested in

$$\widetilde{K}_n(x, P_{\phi,c,Q}) = \Pr{}_{P_{\phi,c,Q}}\{\Sigma_n^{-1/2}(\widehat{\phi}_n - \phi) \le x\}.$$

or if $\sigma_{\widehat{\phi}_n n}$ is the l.s.e estimator of the variance of $\widehat{\phi}_n$

$$L_n(x, P_{\phi,c,Q}) = \Pr{}_{P_{\phi,c,Q}}\{\sigma^{-1}_{\widehat{\phi}_n n}(\widehat{\phi}_n - \phi) \le x\}.$$

One more time the semiparametric bootstrap rule is to try to estimate $K_n(x, P_{\phi,c,Q})$ by $K_n(x, P_{\widehat{\phi}_n,c,Q_n})$. Because of the importance of the first value at least in the explosive case, it is preferable to simulate the bootstrap process with the same starting value c.

As shown by Basawa et al. (1989, 1991), this bootstrap fails for $|\phi| = 1$ but still works in the explosive case under certain conditions. We refer to Bertail (1994) for an unified approach. Subsampling with $\frac{m}{n} \to 0$ that is using in that case $K_m(x, P_{\widehat{\phi}_n,c,Q_n})$ as an estimator of $K_n(x, P_{\phi,c,Q})$ will yield convergent estimators (the proof is left to the reader and simply follows from the fact that the estimator may be reduce to a functional of some i.i.d. residuals). However as suggested by Bertail (1994), the bootstrap may be adapted by using directly P_{1,c,Q_n}, when $\phi = 1$, P_{-1,c,Q_n} when $\phi = -1$. This may be useful in a testing framework since under the null hypothesis $H_0: \ \phi = 1$ one may retrieve a good approximation of the l.s.e. This idea may be practically implemented in the general case by using a threshold version of the l.s.e. (see chap. 1). For this define, for instance,

$$\begin{aligned} \widetilde{\phi}_n = {} & \widehat{\phi}_n 1_{\{|\widehat{\phi}_n| \le 1 - n^{-1/2}\ln n\}} + \operatorname{sign}(\widehat{\phi}_n) 1_{\{1 - n^{-1/2}\ln n < |\widehat{\phi}_n| \le 1 + n^{-1}\ln n\}} \\ & + \widehat{\phi}_n 1_{\{|\widehat{\phi}_n| > 1 + n^{-1}\ln n\}} \end{aligned}$$

That is, we force the estimator to be exactly 1 if the l.s.e. is sufficiently close to 1. Better choice for the threshold estimator are also possible (see Datta et Sriram (1997) for variants). We have however the following theorem :

Theorem 2 *Assume that $E_Q(\varepsilon_t^3) < \infty$, then the bootstrap of the l.s.e. with threshold estimator $\widetilde{\phi}_n$ is asymptotically correct for any value of the parameter ϕ,*

$$||K_n(x, P_{\widetilde{\phi}_n,c,Q_n}) - K_n(., P_{\phi,c,Q})||_\infty \to 0 \ pr.,$$

as well as

$$||\widetilde{K}_n(x, P_{\widetilde{\phi}_n,c,Q_n}) - \widetilde{K}_n(., P_{\phi,c,Q})||_\infty \to 0 \ pr.$$

Moreover if $\frac{m}{n} \to 0$, the subsampling distribution is universally valid (whether we use the threshold estimator or the l. s.e.)

$$||K_m^U(x, P_{\widetilde{\phi}_n,c,Q_n}) - K_n(., P_{\phi,c,Q})||_\infty \to 0 \ pr.$$

The following graphic also taken from Bertail (1994) shows the improvements of the bootstrap distribution over the asymptotic distributions in the non-stationary and explosive case. The important point here is that even if the rate is exponential in the explosive case, the asymptotic distribution may be very difficult to obtain practically. The Bootstrap distribution gives in that case a very accurate approximation (even if we do not know anything on the second order properties).

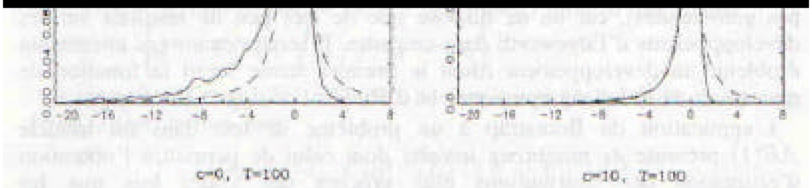
These kind of results may be easily extended to AR(p) model, p>1. It should be however noticed that we have not dealt here with the studentized version of the l.s.e. so that there is no real reason in the stationary (or even non-stationary case, the asymptotic distribution being tabulated) to prefer the asymptotic distribution over the bootstrap distribution. In particular, for $|\phi| < 1$, the result of Bose (1988) shows that neither $K_n(., P_{\widetilde{\phi}_n,c,Q_n})$ nor $\widetilde{K}_n(., P_{\widetilde{\phi}_n,c,Q_n})$ will be second order correct, if $E\varepsilon_t^3 \neq 0$. We have indeed in that case

$$||K_n(., P_{\widetilde{\phi}_n,c,Q_n}) - K_n(., P_{\phi,c,Q})||_\infty = O_P(n^{-1/2}).$$

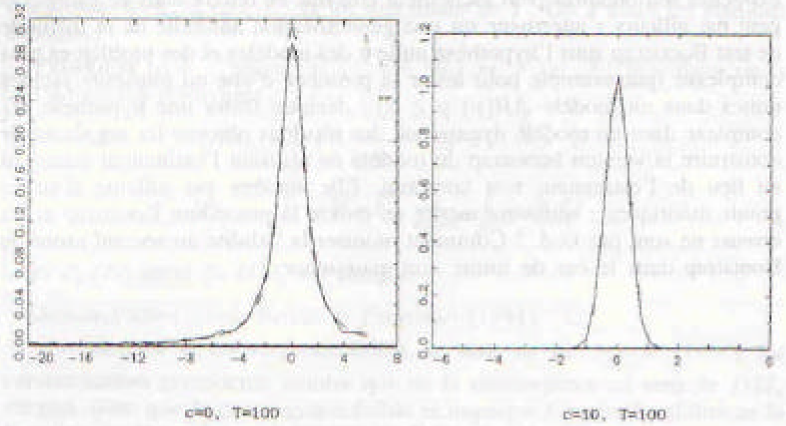
Of course, this is not true anymore, if we standardize by the true variance (but from a practical point of view such results have absolutely no interest). The second order theory in the true studentized case is much more difficult to deal with in the non-stationary case and even more difficult in the explosive case (notice that the limiting distribution has in the $N(0,1)$ case no finite moments). We will see briefly in the last chapter how Edgeworth expansion of the these quantities may be obtained in the non stationary case (random walk case). Some recent results on the bootstrap of studentized version in the non-stationary case have recently been obtained by Park (2004). He shows that when working under the null hypothesis $\phi = 1$,

$$||L_n(x, P_{1,c,Q_n}) - L_n(., P_{1,c,Q})||_\infty = o(n^{-1/2}),$$

[redacted]



GRAPHIQUE 4
Comparaison des densités exactes, Bootstrap et asymptotiques, $\beta_0 = 1.01$.



GRAPHIQUE
Comparaison des densités exactes, Bootstrap et asymptotiques, $\beta_0 = 1.1$.

this result also holds for more general AR(p) model with one unit root. This is an interesting and astonishing result since it is expected with some Markov Chain consideration (in the null recurrent case) that Edgeworth expansion are rather here on power of $n^{1/4}$ (that is the square root of the number of regenerations of the chain for a random walk).

3.3 Autoregressive Sieve Bootstrap

The idea of the Sieve Bootstrap introduced by Bühlmann (1997) (2002) is to use the infinite autoregressive representation of many linear times series models. Assume that $\{X_t\}_{t\in\mathbb{Z}}$ is stationary and admits the representation

$$X_t = \sum_{j=1}^{\infty} \phi_j X_{t-j} + \varepsilon_t,$$

with $\sum_{j=1}^{\infty} \phi_j^2 < \infty$ and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ *i.i.d.* Q (to be relax later), with $E_Q X = 0$. Let $P^{\infty}_{\phi,Q}$ be the corresponding distribution of the process.

It is sometimes difficult to obtain estimators $\phi_n = \{\widehat{\phi}_{j,n}\}_{j=1}^{\infty}$of the infinite sequence of the $\{\phi_j\}_{j=1}^{\infty}$. Even if we have some , as well an estimator of Q say Q_n, generating a stationary realization from $P^{\infty}_{\phi_n,Q_n}$ is practically infeasible. The idea of sieve approximations is simply to try in a first step to replace the model $P^{\infty}_{\phi,Q}$ by a tractable approximation $P^{(l)}_{\phi,Q}$ such that $P^{(l)}_{\phi,Q} \to P^{\infty}_{\phi,Q}$ for some adequate metric. For instance, we may think to approximate $\{X_t\}_{t\in\mathbb{Z}}$ by some sequence of processes $\{X_t^{(l)}\}_{t\in\mathbb{Z}}$ admitting the representation

$$X_t^{(l)} = \sum_{j=1}^{l} \phi_j X_{t-j}^{(l)} + \varepsilon_t,$$

where l is some sequence converging to ∞. We denote by $P^{(l)}_{\phi,Q}$ the law of this process.

When one observes $(X_1, ..., X_n)$, the sieve problem is essentially to choose $l = l(n)$ such that it is big enough for $P^{(l(n))}_{\phi_{l(n)},Q}$, with $\phi_{l(n)} = (\phi_1, ..., \phi_{l(n)})$ to accurately approximate $P^{\infty}_{\phi,Q}$, but small enough to allow the estimation of the $l(n)$ parameters $\phi_{l(n)}$ in the sieve model. However, the optimal choice (according to some criteria) of the length $l(n)$ is not so easy to find and depends on the rate, at which the ϕ_j decreases to 0 as $j \to \infty$. If we have a sieve approximation of the model, we then can get estimators $\widehat{\phi}_{l(n),n} = (\widehat{\phi}_{1,n}, \widehat{\phi}_{2,n}, ..., \widehat{\phi}_{l(n),n})$ of $\phi_{l(n)}$, as well as a convergent estimator $Q_n^{(l(n))}$, by considering the empirical distribution of the centered estimated residual in the sieve model.

In the following, we will assume that $\widehat{\phi}_{l(n),n}$ are the l.s.e satisfying the Yule-Walker equation. The idea of the bootstrap is then to study all the estimators of interest under

$$P^{(l(n))}_{\widehat{\phi}_{n,l(n)},Q_n}.$$

Practically, a stationary stretch from such process will be obtained by generating arrays of processes (as $n->\infty$),

$$X_t^{(l(n))*} = \sum_{j=1}^{l(n)} \widehat{\phi}_{j,n} X_{t-j}^{(l(n))*} \;\; + \varepsilon_t^* \text{ with } \varepsilon_t^* \text{ i.i.d. } Q_n,$$

until stationarity is reached... In the following we are essentially interested in the behaviour of the mean

$$K_n(x, P_{\phi,Q}) = \Pr{}_{P_{\phi,Q}}\{n^{1/2}(\overline{X}_n - E_{P_{\phi,Q}}X) \leq x\}.$$

The sieve Bootstrap distribution is given by

$$K_n(., P^{(l(n))}_{\widehat{\phi}_{n,l(n)},Q_n}) = \Pr{}_{P_{\phi,Q}}\{n^{1/2}(\overline{X}^*_n - \overline{X}_n) \leq x |\; (X_1, ..., X_n)\}.$$

Bühlmann (1997) has shown the asymptotic validity of the method under natural conditions on the underlying process. It is also remarkable to notice that, practically, this method can lead to very accurate approximations (see Choi and Hall, 2000). The second order properties of the method in the studentized case has not been studied until now to our knowledge. It would be also of interest to study the properties of the method in the presence of unit-root by combining threshold ideas and the sieve bootstrap.

Assume the following conditions :

A1 : The process $\{X_t\}_{t\in\mathbb{Z}}$ is linear stationary and admits an $MA(\infty)$ representation

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},\ \psi_0 = 1,$$

with $\sum_{j=1}^{\infty}\psi_j^2 < \infty$.The polynomial $\Psi(x) = \sum_{j=0}^{\infty}\psi_j x^j$ is bounded away from 0 for $|x| \leq 1$.

A2 $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a stationary ergodic sequence with $E(\varepsilon_t|F_{t-1}) = 0,\ V(\varepsilon_t|F_{t-1}) = \sigma^2$, where $F_t = \sigma(\{\varepsilon_u, u \leq t\})$ and $E_Q\varepsilon_t^4 < \infty$,

A3(r) : There exists r such that $\sum_{j=0}^{\infty} j^r\psi_j < \infty$ then

Theorem 3 *Assume A1,A2 and A3(1), and*

$$n^{1/2}(X_n - E_P X) \underset{n\to\infty}{\xrightarrow{L}} N(0, \Sigma_{k=-\infty}^{\infty} cov(X_t, X_{t+k})),$$

then if $l(n) = o((n/log(n))^{1/4})$, $l(n) \to \infty$, *we have uniformly*

$$K_n(.,P^{(l(n))}_{\widehat{\phi}_{n,l(n)},Q_n}) - K_n(.,P_{\phi,Q}) \underset{n\to\infty}{\to} 0, \text{ Pr}.$$

If in addition A3(r) holds with $r \geq 1$ *and*

$$\sum_{t_1,t_2,t_3=1}^{\infty} E_{P_{\phi,Q}}((X_0 - EX)(X_{t_1} - EX)(X_{t_2} - EX)(X_{t_3} - EX)) < \infty$$

then for $l(n) = o((n/log(n))^{1/(2r+2)})$, *we have*

$$\begin{aligned}\Delta V_n &= Var_{P^{(l(n))}_{\widehat{\phi}_{n,l(n)},Q_n}}(n^{1/2}(\overline{X}^*_n - \overline{X}_n)) - Var_{P_{\phi,Q}}(n^{1/2}(\overline{X}_n - E_{P_{\phi,Q}}X) \\ &= O_P\left(\left(\frac{l(n)}{n}\right)^{1/2}\right) + O_P\left(\frac{1}{l(n)^r}\right)\end{aligned}$$

The last result means that the bootstrap variance estimator may be an accurate estimator of the variance. In particular, in the case of exponentially decreasing coefficient ψ_j, then for any $1/2 > k > 0$, we have

$$\Delta V_n = O_P(n^{-1/2+k}),$$

by choosing $r > 1/(2k) - 1$, and $l(n) = n^{1/(2r+2)} \log(n)^{-1/(2r+2)-1}$.

Such a result is actually very important when one considers studentized statistics. Actually we will see in the next chapter that the choice of the variance is one of the main problem in getting good asymptotic properties in a non i.i.d. framework.

An interesting discussion of the avantages of this approach may be find in Bühlmann (2002).

References

Basawa, I.V., Mallik, A.K., McCormick, W.P., Taylor, R.L.(1989). Bootstrapping in Explosive Autoregressive Processes, *Ann. Statist.*, 17, 1479-1486.

Basawa, I.V., Mallik, A.K., McCormick, W.P., Reeves, J.H., Taylor, R.L.(1991): Bootstrapping Unstable Autoregressive Processes, *Ann. Statist.*, 19, 1098-1101.

Bertail, P. (1994). Un test bootstrap dans un modèle AR(1), *Annales d'Economie et de Statistiques*, 36, 57-81.

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Ann. Statist.,* **16**, 1709-1722.

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

Bühlmann, P. (2002). Bootstrap for time series. *Statistical Science*, **17**, 52-72.

Choi, E., Hall, P. (2000). Bootstrap confidence regions computed from autoregression of abitrary order, *Journal of the Statistical Royal Society*, ser. B, **62**, 461-477.

Datta, S. , Sriram, T.N. (1997). A modified bootstrap for autoregression without stationarity, *Journal of Statistical Planning and Inference*, 59, 19-30.

Dickey, D.A., Fuller, W.F. (1979). Distribution of The Estimators for Autoregressive Time Series with Unit Roots, *J.A.S.A.*, 74, 427-431.

Efron, B. (1982) : *The Jackknife, the Bootstrap, and Other Resampling Plans*, CBMS−NF 38, S.I.A.M., Philadelphia.

Franke, J., Kreiss, J.P., Mammen, E. (2002), Bootstrap of kernel smoothing in nonlinear time series. Bernoulli 8(1), 2002, 1–37

Freedman, D. (1984). On Bootstrapping 2-Stage-Least-Squares Estimates In Stationary Models, *Ann. Statist.*, 12, 827-842.

Götze, F. and Hipp, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrsch. Verw. Gebiete*, **64**, 211-239.

Park, J.Y. (2000?)(2004). Bootstrap unit-root tests, Rice University, preprint.

Phillips, P.C.B. (1987): Towards a Unified Asymptotic Theory of Autoregression, *Biometrika*, 74, 535-547.

White, J.S. (1958). The Limiting Distribution of the Serial Correlation Coefficient in the Explosive, Case I, *Ann. Math. Statist.*, 29, 1188-1197.

White, J.S. (1959). The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case II, *Ann. Math. Statist.*, 30, 831-834.