Approximate regenerative-block bootstrap for Markov chains: some simulation studies

Patrice Bertail, Stéphan Clémençon

Approximate regenerative-block bootstrap for Markov chains: some simulation studies

Patrice Bertail [a] & Stéphan Clémençon [b,c,d]

[a] *Laboratoire de Statistique - CREST - INSEE*

[b] *MODALX - Université Paris X Nanterre*

[c] *Unité Metarisk - Institut National de le Recherche Agronomique*

[d] *LPMA - UMR CNRS 7599 - Universités Paris 6 et Paris 7*

Abstract

In [7], a novel methodology for bootstrapping general Harris Markov chains has been developed, the (approximate) regenerative block-bootstrap. It is built on the renewal properties of the chain (or of a Nummelin extension of the latter) and has theoretical properties that surpass other existing methods within the Markovian framework. This paper is devoted to discuss practical issues related to the implementation of this specific resampling method and to present various simulation studies for investigating its performance and comparing it to other bootstrap resampling schemes, standing as natural candidates in the Markov setting.

Key words: Markov chains, Bootstrap, Regeneration, Small set, Nummelin splitting technique, Simulation.
PACS: 62E20, 62D05, G2N01

1 Introduction

Adapting the naive bootstrap method introduced by [19] in the i.i.d. setting, so as to produce asymptotically valid procedures for dependent data, and time series in particular, constitutes an intense field of research, still developing. The principle underlying such generalizations generally consists in resampling whole blocks of observations instead of single observed values in order to mimic the data dependence (see [29] and the references therein). The *moving-block bootstrap* (MBB) illustrates this idea, it lies in resampling (overlapping or disjoint) data blocks of fixed length to capture the dependence structure of the observations (refer to [14] and [36] for recent surveys). Although this approach may yield consistent procedures in many weakly dependent settings,

it has several important drawbacks. First, stationarity of the observations is usually required by the validity framework of the MBB approach. Furthermore, implementing the MBB method calls for a preliminary estimation of the bias and of the asymptotic variance of the statistic of interest. This makes its application difficult in practice (see [22]). And from a theoretical viewpoint, the rate of convergence of the MBB distribution is slower than the one of the i.i.d. bootstrap: at best it is of order $O_{\mathbb{P}}(n^{-3/4})$ under restrictive conditions (stipulating that all moments are finite and that strong mixing coefficients decrease exponentially fast). Eventually, the results obtained highly depend on the choice of the block size. Except in very particular situations (see [22]), no general method for determining the adequate block size has yet been developed. In [24], [15] and [37], data-driven methods for selecting the block size are proposed, aiming at approximating the variance, rather than at estimating directly the studentized distribution in a precise fashion (see [29]).

Various approaches for bootstrapping particular types of stationary Markov chains have recently been proposed in the statistical literature. If a parametric Markovian model is *a priori* specified (an ARMA model for instance), the problem simply reduces then to sampling randomly estimated centered residuals (see [10]). Following this idea, [13] introduced a *sieve bootstrap method* based on approximating the time series by some AR model with large lag-order. This method presents both promising theoretical results and good practical performance at the same time, but is well suited to linear stationary time series rather than to general Markov chains. Another approach to bootstrapping Markov chains follows the proposal of [3] (see also [38]) in the finite case, which uses a nonparametric estimate of the transition probability for reproducing the markovian structure of the data series (see also the *local Markov bootstrap* in [35] based on similar ideas).

Following the approach proposed by [18], it has been shown in [7], that a specific resampling technique for bootstrapping some statistics of *regenerative Markov chains* offers attracting advantages both regarding to asymptotic second order properties and from a practical viewpoint. This method, the *Regenerative Block Bootstrap* (RBB), consists in resampling (a random number of) data blocks corresponding to 'cycles' of the observed path (*i.e.* data segment between consecutive regeneration times) until the length of the reconstructed trajectory is larger than the initial one. In the general (non-regenerative) setting, [7] proved that this principle can still be successfully applied in numerous situations, by adding a preliminary stage consisting in estimating the distribution of a regenerative extension of the chain. Due to the approximation step, we call this method *Approximate Regenerative Block Bootstrap* (ARBB).

The purpose of the paper is to describe the mechanic of the (A)RBB method from a practical angle and evaluate empirically its performance in comparison with major competitors. It is organized as follows. In section 2, basics concern-

ing the regenerative method in connection with the Markov chain theory and the Nummelin splitting technique are briefly recalled. Section 3 is devoted to the description of the (A)RBB algorithm. Practical issues related to its implementation are discussed as well. The performance of the (A)RBB methodology is then investigated through several applications: simulation studies are carried out in section 4, which show on some (regenerative and pseudo-regenerative) examples of Markov chains, arising from operational research or standard time series analysis, performs, when compared to natural bootstrap method competitors. In section 5, some concluding remarks are collected, together with several lines of further research.

2 Theoretical Background

In what follows $X = (X_n)_{n\in\mathbb{N}}$ denotes an aperiodic Harris Markov chain on a countably generated state space $(E, \mathcal{E})$, with transition probability Π, and initial probability distribution ν (see [39] for an account of the Markov chain theory). We also denote by $\mathbb{P}_\nu$ (respectively by $\mathbb{P}_x$ for x in E) the probability measure on the underlying space such that $X_0 \sim \nu$ (resp. $X_0 = x$), by $\mathbb{E}_\nu[.]$ the $\mathbb{P}_\nu$-expectation (resp. by $\mathbb{E}_x[.]$ the $\mathbb{P}_x$-expectation) and by $\mathbb{I}\{\mathcal{A}\}$ the indicator function of the event $\mathcal{A}$.

We now recall key notions, concerning the regenerative method and its application to the analysis of the behavior of general Harris chains via the Nummelin splitting technique (refer to [34], for further detail).

Regenerative Markov chains Here we assume that the chain X possesses an *accessible atom*, *i.e.* a Harris set A such that for all $(x,\, y) \in A^2$, $\Pi(x, .) = \Pi(y, .)$. Denote by $\tau_A = \tau_A(1) = \inf\{n \geq 1, X_n \in A\}$ the hitting time on A, by $\tau_A(j) = \inf\{n > \tau_A(j-1), X_n \in A\}$ for $j \geq 2$, the successive return times to A, and by $\mathbb{E}_A[.]$ the expectation conditioned on $X_0 \in A$.

Regeneration blocks From the *strong Markov property* it is immediate that, for any initial distribution ν, the sample paths of the chain may be divided into i.i.d. blocks of random length corresponding to consecutive visits to the atom A

$$\mathcal{B}_1 = \left(X_{\tau_A(1)+1}, ...,\ X_{\tau_A(2)}\right), ...,\ \mathcal{B}_j = \left(X_{\tau_A(j)+1}, ...,\ X_{\tau_A(j+1)}\right), ... \tag{1}$$

taking their values in the torus $\mathbb{T} = \cup_{n=1}^{\infty} E^n$. The $\tau_A(j)$'s are thus successive random times at which the chain forgets its past, namely *regeneration times.* In this regenerative setting, the *stochastic stability* properties of the chain amount to properties concerning the speed of return time to the atom only. For instance, X is positive recurrent iff $\mathbb{E}_A[\tau_A] < \infty$ (see Kac's theorem in

[31]). In such a case the unique invariant probability distribution μ is the Pitman's occupation measure given by:

$$\forall B \in \mathcal{E},\ \mu(B) = \frac{1}{\mathbb{E}_A[\tau_A]}\mathbb{E}_A[\sum_{i=1}^{\tau_A} \mathbb{I}\{X_i \in B\}]. \tag{2}$$

In particular, parameters of interest for a positive Harris Markov chain may be expressed in many cases in terms of regeneration cycles only (mainly those related to the long term behaviour of the process, as indicated by (2)). Numerous examples are given in [8]).

The regenerative method We point out that first order limit results such as the Law of Large Numbers, the Central Limit Theorem or the Law of Iterated Logarithm for an additive functional $\sum_i f(X_i)$ of a regenerative Harris positive chain X may be easily derived by applying the corresponding i.i.d. results to functionals of the i.i.d. regeneration blocks $(\mathcal{B}_j)_{j\geqslant 1}$ (see [31] for such illustrations of the *regenerative method* introduced by [42]). However, when the matter is to establish higher order limit results (see [9], [30] or [4] for refinements of the CLT in the Markovian setting), the fact that the data blocks $\mathcal{B}_0 = (X_1, ..., X_{\tau_A(1)})$, $\mathcal{B}_1, ..., \mathcal{B}_{l_n-1}$, $\mathcal{B}_{l_n}^{(n)} = (X_{\tau_A(l_n)+1}, ..., X_n)$ defined by the $l_n = \sum_{1\leqslant k\leqslant n} \mathbb{I}\{X_k \in A\}$ regeneration times over a trajectory of finite length n are *not* independent (the sum of the block lengths is of course n) is essential. The randomness of the number of blocks plays a crucial role in the distribution of any statistic based on a finite sample. This observation lies at the heart of the RBB procedure (see section 3, details are in [7]).

Regeneration-based statistics In the time series framework, inference is generally based on a single trajectory $X_1, ..., X_n$ of the Harris chain X. Therefore, in a nonstationary setting the distribution of the non-regenerative blocks, $\mathcal{B}_0$ and $\mathcal{B}_{l_n}^{(n)}$, cannot be estimated from a single realization of the chain only (due to their dramatical dependence upon the unknown initial distribution). Furthermore, their contribution to the value of a functional $T(X_1, ..., X_n)$ of interest is generally significant, leading to first order bias terms in particular (see the discussion in [4]). Hence, statistics involving $\mathcal{B}_0$ and $\mathcal{B}_{l_n}^{(n)}$ must be avoided in practice, when estimating sampling distributions is essential (for building confidence intervals for instance, see [8]).

As an illustration, when X is a positive recurrent chain with limiting probability distribution μ, consider $f : E \to \mathbb{R}$ a μ-integrable function. In the nonstationary case, when the matter is to recover the asymptotic mean $\mu(f) = \int f d\mu$ from data $X_1, ..., X_n$ (notice that $\mu(f) = \mathbb{E}_A[\tau_A]^{-1}\mathbb{E}_A[\sum_{1\leqslant i\leqslant \tau_A} f(X_i)]$), together with assessing the accuracy of the estimation, rather than the standard sample mean $\overline{\mu}_n(f) = n^{-1}\sum_{1\leqslant k\leqslant n} f(X_k)$, it is preferable to use the truncated mean computed using data collected between the first and last regeneration

times, namely

$$\widehat{\mu}_n(f) = \frac{\sum_{k=1+\tau_A}^{\tau_A(l_n)} f(X_k)}{\tau_A(l_n) - \tau_A} = \frac{\sum_{j=1}^{l_n-1} f(\mathcal{B}_j)}{\sum_{j=1}^{l_n-1} L(\mathcal{B}_j)}, \tag{3}$$

where $f(\mathcal{B}_j) = \sum_{k=1+\tau_A(j)}^{\tau_A(j+1)} f(X_k)$ and $L(\mathcal{B}_j) = \tau_A(j+1) - \tau_A(j)$ for $j \geqslant 1$, with the convention that $\widehat{\mu}_n(f) = 0$ when $l_n \leqslant 2$. Furthermore, under suitable moment conditions (see the paragraph below) [4] have shown that the estimator $\widehat{\mu}_n(f)$ is asymptotically normal with asymptotic mean $\mu(f)$ and variance 1, when standardized by the following sequence, based on regeneration data blocks as well (see [26])

$$\widehat{\sigma}_n^2(f) = \frac{\sum_{j=1}^{l_n-1} \left(f(\mathcal{B}_j) - \widehat{\mu}_n(f) L(\mathcal{B}_j)\right)^2}{\sum_{j=1}^{l_n-1} L(\mathcal{B}_j)}. \tag{4}$$

Precisely, they proved that $\widehat{\sigma}_n^2(f)$ is a strongly consistent and asymptotically normal estimator of the limiting variance $\sigma_f^2 = \mu(A)\mathbb{E}_A[\sum_{1 \leqslant i \leqslant \tau_A} f(X_i)]$, with a bias of order $O(1/n)$ as $n \to \infty$.

When implementing the MBB, the choice of the standardization, the bias it induces and the definition of its Bootstrap counterpart are key points to obtain the second order validity of the method. As shown in [7], the standardization (4), which is specifically tailored for the regenerative setting, does not weaken the performance of the RBB (see also section 3 below), while the standardization of the MBB distribution in the strong mixing case is the main barrier to achieve good performance (as shown by [22]). In most practical situations (except for the very special case of m -dependence), positive moving-block based estimates of the asymptotic variance with such good properties are not available.

Technical assumptions The assumptions required by the validity framework of the RBB method (see [7] for further details) are of the following type.

- Regularity conditions: there exists $\kappa \geq 1$ such that $\mathbb{E}_A[\tau_A^\kappa] < \infty$ and $\mathbb{E}_\nu[\tau_A^\kappa] < \infty$.
- Block-moment conditions: there exists $\kappa \geq 1$ such that $\mathbb{E}_A[(\sum_{i=1}^{\tau_S} |f(X_i)|)^\kappa] < \infty$ and $\mathbb{E}_\nu[(\sum_{i=1}^{\tau_S} |f(X_i)|)^\kappa] < \infty$.

Using well-known results (see Chapt. 11 in [31] and the references therein), such assumptions may be checked in practice by establishing drift criteria of Lyapounov's type for the chain. When considering Edgeworth expansion or second order results, the "block-Cramer conditions" of the following type are additionally required.

- Block-Cramer condition:

$$\limsup_{t\to\infty} \mid \mathbb{E}_A[\exp(it\sum_{i=1}^{\tau_A} f(X_i))] \mid < 1.$$

Such conditions are checked on various models in section 4.3 of [8] (see also [26]), including the examples considered in section 4.

2.1 Regenerative extension

Now we recall the *splitting technique* introduced in [33]. This theoretical construction aims at extending in some sense the probabilistic structure of a general Harris chain, so as to artificially build a regeneration set. It is based on the following notion. A set $S \in \mathcal{E}$ is *small* for X if there exist $m \in \mathbb{N}^*$, a probability measure Φ supported by S, and $\delta > 0$ such that

$$\forall x \in S, \forall A \in \mathcal{E}, \quad \Pi^m(x, A) \geq \delta\Phi(A), \tag{5}$$

where Π^m denotes the m-th iterate of Π. Roughly speaking, the small sets are the ones on which an iterate of the transition probability is uniformly bounded below. When (5) holds, we shall say that X satisfies the *minorization condition* $\mathcal{M}(m, S, \delta, \Phi)$. Small sets do exist for irreducible chains [1] , *a fortiori* for Harris chains (any accessible set actually contains small sets, see [27]). Suppose that X fulfills $\mathcal{M} = \mathcal{M}(m, S, \delta, \Phi)$ for some accessible set S. Take $m = 1$, even if it entails to replace X by the chain $((X_{nm}, ..., X_{n(m+1)-1}))_{n\in\mathbb{N}}$. The regenerative chain onto which the initial chain X is embedded is constructed by expanding the sample space, so as to define a specific sequence $(Y_n)_{n\in\mathbb{N}}$ of independent Bernoulli r.v.'s with parameter δ. This joint distribution $\mathbb{P}_{\nu,\mathcal{M}}$ is obtained by randomizing the transition Π each time the chain X hits S (this happens a.s. since X is Harris). If $X_n \in S$ and

- if $Y_n = 1$ (which happens with probability $\delta \in]0, 1[$), then draw X_{n+1} according to Φ,
- if $Y_n = 0$, (which happens with probability $1-\delta$), then draw X_{n+1} according to $(1 - \delta)^{-1}(\Pi(X_n, .) - \delta\Phi(.))$.

For obtaining an insight into this construction, observe simply that, if condition (5) holds with $m = 1$, when $X_n \in S$, one may write the distribution of X_{n+1} conditioned on X_n as the following mixture

$$\Pi(X_n, .) = (1 - \delta)\{(1 - \delta)^{-1}(\Pi(X_n, .) - \delta\Phi(.))\} + \delta\Phi(.), \tag{6}$$

[1] Recall that a Markov chain X with state space $(E, \mathcal{E})$ and transition Π is irreducible if there exists a positive measure that dominates $\sum_{n\geq 1} \Pi^n(x.)$ for all $x \in E$.

which second component is independent from X_n. The bivariate Markov chain $X^{\mathcal{M}} = ((X_n, Y_n))_{n\in\mathbb{N}}$ constructed this way is called the *split chain*. The key point lies in the fact that $S \times \{1\}$ is then an atom for the split chain $X^{\mathcal{M}}$, the latter inheriting all the communication and stochastic stability properties from X. In particular the blocks constructed from the consecutive times when $X^{\mathcal{M}}$ visits $S \times \{1\}$ are independent (if X satisfies $\mathcal{M} = \mathcal{M}(m, S, \delta, \Phi)$ for $m > 1$, the resulting blocks are 1-dependent only, a form of dependence that can also be easily handled). Using this construction, one may enlarge the range of applications of the regenerative method so as to extend all of the results established for atomic chains to general Harris chains. We omit the subscript $\mathcal{M}$ in what follows and abusively denote by $\mathbb{P}_\nu$ the extensions of the underlying probability we consider.

2.2 On approximating the regenerative extension

Here we assume further that the conditional distributions $\{\Pi(x, dy)\}_{x\in E}$ and the initial distribution ν are dominated by a σ-finite measure λ of reference, so that $\nu(dy) = f(y)\lambda(dy)$ and $\Pi(x, dy) = p(x, y)\lambda(dy)$ for all $x \in E$. For simplicity's sake, we suppose that condition $\mathcal{M}$ is fulfilled with $m = 1$. This entails that Φ is absolutely continuous with respect to λ too, and that

$$p(x, y) \geq \delta\phi(y),\ \lambda(dy)a.s. \tag{7}$$

for any $x \in S$, with $\Phi(dy) = \phi(y)dy$.

If we were able to generate practically binary random variables $Y_1, ..., Y_n$, so that $X^{\mathcal{M}\,(n)} = ((X_1, Y_1), ..., (X_n, Y_n))$ be a realization of the split chain $X^{\mathcal{M}}$ described above, then we could divide the sample path $X^{(n)} = (X_1, ..., X_n)$ into regeneration blocks, as in §2.1. Therefore, knowledge of Π is required to draw $Y_1, ..., Y_n$ this way. As a matter of fact, the distribution $\mathcal{L}^{(n)}(p, S, \delta, \phi, x^{(n+1)})$ of $Y^{(n)} = (Y_1, ..., Y_n)$ conditioned on $X^{(n+1)} = (x_1, ..., x_{n+1})$ is the tensor product of Bernoulli distributions given by: $\forall \beta^{(n)} = (\beta_1, ..., \beta_n) \in \{0, 1\}^n$, $\forall x^{(n+1)} = (x_1, ..., x_{n+1}) \in E^{n+1}$,

$$\mathbb{P}_\nu(Y^{(n)} = \beta^{(n)} \mid X^{(n+1)} = x^{(n+1)}) = \prod_{i=1}^{n} \mathbb{P}_\nu(Y_i = \beta_i \mid X_i = x_i,\ X_{i+1} = x_{i+1}) \tag{8}$$

with for $1 \leqslant i \leqslant n$: if $x_i \notin S$,

$$\mathbb{P}_\nu(Y_i = \beta_i \mid X_i = x_i,\ X_{i+1} = x_{i+1}) = Ber_\delta\,(\beta_i)\,, \tag{9}$$

and if $x_i \in S$,

$$\mathbb{P}_\nu(Y_i = 1 \mid X_i = x_i,\ X_{i+1} = x_{i+1}) = \delta\phi(x_{i+1})/p(x_i, x_{i+1}),$$
$$\mathbb{P}_\nu(Y_i = 0 \mid X_i = x_i,\ X_{i+1} = x_{i+1}) = 1 - \delta\phi(x_{i+1})/p(x_i, x_{i+1}). \tag{10}$$

In short, given $X^{(n+1)}$, the Y_i's are Bernoulli r.v.'s with parameter δ, unless X has hit the small set S at time i: in this case Y_i is drawn from the Bernoulli distribution with parameter $\delta\phi(X_{i+1})/p(X_i, X_{i+1})$. Our proposition for constructing data blocks relies in approximating this construction by computing first an estimate $p_n(x, y)$ of the transition density $p(x, y)$ from data $X_1, ..., X_{n+1}$, and then drawing a random vector $(\widehat{Y}_1, ..., \widehat{Y}_n)$ from the distribution $\mathcal{L}^{(n)}(p_n, S, \delta, \phi, X^{(n+1)})$, obtained by simply plugging p_n in (10) and (10) (the estimate $p_n(x, y)$ may be picked such that $p_n(x, y) \geq \delta\phi(y)$, $\lambda(dy)$ a.s., and $p_n(X_i, X_{i+1}) > 0$, $1 \leqslant i \leqslant n$).

From a practical viewpoint, it actually suffices to draw the $\widehat{Y}_i$'s only at times i when the chain hits the small set S, $\widehat{Y}_i$ indicating then whether the trajectory should be divided at time point i or not (see Fig. 1 for instance). This way, on gets the *approximate regeneration blocks* $\widehat{\mathcal{B}}_1, ..., \widehat{\mathcal{B}}_{\widehat{l}_n - 1}$ with $\widehat{l}_n = \sum_{1 \leqslant k \leqslant n} \mathbb{I}\{X_k \in S,\ Y_k = 1\}$. Of course, knowledge of parameters (S, δ, ϕ) of condition (7) is required for this construction. In § 3.2, we shall discuss a practical method for selecting those parameters.

The question of accuracy of this approximation has been addressed in [7]. Precisely, they established a bound for the deviation between the distribution of $((X_i, Y_i))_{1 \leqslant i \leqslant n}$ and the one of the $((X_i, \widehat{Y}_i))_{1 \leqslant i \leqslant n}$ in the sense of the Mallows distance, which essentially depends on the rate of the uniform convergence of $p_n(x, y)$ to $p(x, y)$ over $S \times S$.

3 The (A)RBB methodology

Now that necessary background material has been reviewed, we turn to describe the following block-resampling method and discuss practical issues encountered for implementing the latter.

3.1 The (A)RBB algorithm

Suppose that the finite sample path has been divided into true or approximate regeneration blocks $\mathcal{B}_1, ..., \mathcal{B}_{l_n - 1}$. The *(approximate) regenerative block-bootstrap* algorithm for estimating the sample distribution of some statistic $T_n = T(\mathcal{B}_1, ..., \mathcal{B}_{l_n - 1})$ estimating some parameter θ with standardization

$\sigma_n = \sigma(\mathcal{B}_1, ..., \mathcal{B}_{l_n-1})$, namely

$$H(x) = \mathbb{P}(\sigma_n^{-1}(T_n - \theta) \leqslant x), \tag{11}$$

is performed in three steps as follows.

(1) Draw sequentially bootstrap data blocks $\mathcal{B}_1^*, ..., \mathcal{B}_k^*$ independently from the empirical distribution $F_n = (l_n - 1)^{-1} \sum_{j=1}^{l_n-1} \delta_{\mathcal{B}_j}$ of the blocks $\mathcal{B}_1, ..., \mathcal{B}_{l_n-1}$, conditioned on $X^{(n)}$ until the length of the bootstrap data series $l^*(k) = \sum_{j=1}^{k} l(\mathcal{B}_j^*)$ is larger than n. Let $l_n^* = \inf\{k \geqslant 1,\, l^*(k) > n\}$.

(2) From the bootstrap data blocks generated at step 1, reconstruct a pseudo-trajectory by binding the blocks together, getting the reconstructed *(A)RBB sample path*

$$X^{*(n)} = (\mathcal{B}_1^*, ..., \mathcal{B}_{l_n^*-1}^*). \tag{12}$$

Then compute the *(A)RBB statistic* and the *(A)RBB standardization*

$$T_n^* = T(X^{*(n)}) and \sigma_n^{*(n)} = \sigma(X^{*(n)}). \tag{13}$$

(3) The *(A)RBB distribution* is then given by

$$H_{(A)RBB}(x) = \mathbb{P}^*(\sigma_n^{*-1}(T_n^* - T_n) \leqslant x \mid X^{(n+1)}), \tag{14}$$

denoting by $\mathbb{P}^*(. \mid X^{(n+1)})$ the conditional probability given $X^{(n+1)}$.

A Monte-Carlo approximation to $H_{ARBB}(x)$ may be straightforwardly computed by repeating independently N times the procedure above. Based on Edgeworth expansions proved in [4], one may show that in the regenerative positive recurrent case, the RBB method inherits the accuracy of the standard i.i.d. bootstrap (see [23]) up to $O_{\mathbb{P}_\nu}(n^{-1})$ for additive functionals of type $n^{-1} \sum_{1 \leqslant k \leqslant n} f(X_k)$ under weak conditions (see Theorem 3.3 in [7] for further details). In [8] asymptotic validity of the RBB has also been established for more general functionals, including U or V statistics based on regeneration blocks. In the general Harris recurrent case, the ARBB method for bootstrapping Markov chains simply relies in applying the RBB procedure to the data $((X_1, \widehat{Y}_1), ..., (X_n, \widehat{Y}_n))$ as if they were exactly drawn from the atomic chain $X^{\mathcal{M}}$. However, as shown in [7], even if it requires to use a consistent estimate of the "nuisance parameter" p and the corresponding approximate blocks it induces, this bootstrap method still remains asymptotically valid.

In [4] (see Prop. 3.1) it is shown that in the nonstationary case (*i.e.* when the initial law ν differs from μ), the first data block $\mathcal{B}_0$ induces a significant bias, of order $O(n^{-1})$, which cannot be estimated from a single realization $X^{(n)}$ of the chain starting from ν. Practitioners are thus recommended not to

use estimators based on the whole trajectory. This fact is known as the *burn-in* (time) problem in the bayesian literature on *MCMC* algorithms, related to the time from which the 1-dimensional marginal of a (simulated) chain is close enough to the limit distribution μ. When the statistic is built using regeneration blocks only, the first (non-regenerative) block has no impact on the second order properties of the ARBB estimate . However, from a practical viewpoint, it may happen that the size of the first block is large compared to the size n of the whole trajectory (for instance in the case where the expected return time to the (pseudo-)atom when starting with ν is large), the effective sample size for constructing the data blocks and the corresponding statistic is then dramatically reduced. In such a case, for mimicking the distribution of the original statistic, it is preferable, heuristically speaking, to draw sequentially the bootstrap blocks $\mathcal{B}_1^*, ..., \mathcal{B}_k^*$ independently from the empirical distribution F_n, until $l(\mathcal{B}_0) + \sum_{j=1}^{k} l(\mathcal{B}_j^*)$ is larger than n, taking practically into account the size $l(\mathcal{B}_0)$ this way (although it does not play any role in the asymptotic behavior, since $l(\mathcal{B}_0)/n = O_{\mathbb{P}_\nu}(n^{-1})$ as $n \to \infty$).

3.2 Tuning parameters

In the general (non-regenerative) case, the procedure above may be very sensitive to the choice of the minorization condition parameters (S, δ, Φ). It is essential to pick the latter in a data-driven fashion, so that enough blocks may be obtained for computing meaningful statistics, their accuracy increasing as the mean number of pseudo-regenerative blocks, that is

$$N_n(S) = \mathbb{E}_\nu[\sum_{i=1}^{n} \mathbb{I}\{X_i \in S, Y_i = 1\} \mid X^{(n+1)}], \tag{15}$$

for a given realization of the trajectory. Therefore, this is somehow determined by the size of the small set chosen. More precisely, it depends on how often the chain visits the latter in a finite length path) and how sharp is the lower bound in the minorization condition. The trade-off is as follows: as the size of the small set S used for the data blocks construction increases, the number of points of the trajectory that are candidates for determining a 'cut' in the trajectory naturally increases, but, since the uniform lower bound for $p(x, y)$ over S^2 then decreases, the probability of drawing $Y_i = 1$ also decreases (see expression (10)). Thus one may heuristically expect better numerical results for the ARBB, when one implements it by choosing S so as to maximize the expected number of data blocks given the trajectory, namely $N_n(S) - 1$.

In the case when the chain takes real values and in lack of any prior information about its structure, a possible data-driven method for selecting the tuning parameters could be as follows. Let $\mathcal{S}$ be a collection of borelian sets

S (typically compact intervals) and let $\mathcal{U}_S(dy) = \phi_S(y).\lambda(dy)$ denote the uniform distribution on S, where $\phi_S(y) = \mathbb{I}\{y \in S\}/\lambda(S)$ and λ is the Lebesgue measure on $\mathbb{R}$. For any $S \in \mathcal{S}$, we clearly have $p(x,y) \geq \delta(S)\phi_S(y)$ for all x, y in S, with $\delta(S) = \lambda(S).\inf_{(x,y)\in S^2} p(x,y)$. When $\delta(S) > 0$, the theoretical criterion (15), that one would ideally seek to maximize over $\mathcal{S}$, can be written as follows

$$N_n(S) = \inf_{(x,y)\in S^2} p(x,y) \times \sum_{i=1}^{n} \frac{\mathbb{I}\{(X_i, X_{i+1}) \in S^2\}}{p(X_i, X_{i+1})}. \tag{16}$$

Observing that $N_n(S)/n$ converges $\mathbb{P}_\nu$-a.s. to $\lambda(S)\mu(S)$ as $n \to \infty$, an alternative criterion to maximize, independent from the data and asymptotically equivalent to $N_n(S)$, is given by

$$\mathcal{N}_n(S) = n \inf_{(x,y)\in S^2} p(x,y)\ \lambda(S)\mu(S), \tag{17}$$

One gets an empirical counterpart of these quantities by replacing the unknown transition density $p(x,y)$ by an estimate $p_n(x,y)$ in expression (16) or (17), and $\mu(S)$ by the empirical estimator $\widehat{\mu}_n(S) = n^{-1}\sum_{1\leqslant i\leqslant n} \mathbb{I}\{X_i \in S\}$. Actually from a bootstrap viewpoint, the conditional criterion (16) is more pertinent because the rate of convergence of the ARBB distribution is directly related to the effective number of observed regeneration times conditionally to the trajectory. Note furthermore that many nonparametric estimators of the transition density of Harris recurrent chains have been proposed in the literature, among which the standard *Nadaraya-Watson estimator*

$$p_n(x,y) = \frac{\sum_{i=1}^{n} K(h^{-1}(x - X_i))K(h^{-1}(y - X_{i+1}))}{\sum_{i=1}^{n} K(h^{-1}(x - X_i))}, \tag{18}$$

computed from a Parzen-Rosenblatt kernel $K(x)$ and a bandwidth $h > 0$. In the positive recurrent case, their estimation rates have been established under various smoothness assumptions on the density of the joint distribution $\mu(dx)\Pi(x,dy)$ and the one of $\mu(dx)$ (see [2] or [17] and the references therein for instance).

Once $p_n(x,y)$ is computed, calculate its minimum over sets S of the class $\mathcal{S}$ and maximize then the practical empirical criterion over $\mathcal{S}$:

$$S^* = \arg\max_{S\in\mathcal{S}} \widehat{N}_n(S) \tag{19}$$

with

$$\widehat{N}_n(S) = \inf_{(x,y)\in S^2} p_n(x,y) \times \sum_{i=1}^{n} \frac{\mathbb{I}\{(X_i, X_{i+1}) \in S^2\}}{p_n(X_i, X_{i+1})}. \tag{20}$$

On many examples of real valued chains (see section 4 below), it is possible to check at hand that any compact interval $V_{x_0}(\varepsilon) = [x_0 - \varepsilon, x_0 + \varepsilon]$ for a suitably chosen $x_0 \in \mathbb{R}$ and $\varepsilon > 0$ small enough, is small, choosing ϕ as the density $\phi_{V_{x_0}(\varepsilon)}$ of the uniform distribution on $V_{x_0}(\varepsilon)$. For practical purpose, one may perform the optimization over $\varepsilon > 0$, while x_0 is kept fixed (see [7], [8]). But both x_0 and ε may be considered as tuning parameters: searching for (x_0, ε) over a pre-selected grid $\mathcal{G} = \{(x_0(k), \varepsilon(l)),\ 1 \leqslant k \leqslant K,\ 1 \leqslant l \leqslant L\}$ such that $\inf_{(x,y)\in V_{x_0}(\varepsilon)^2} p_n(x, y) > 0$ for any $(x_0, \varepsilon) \in \mathcal{G}$ could lead to the following numerically feasible selection rule. For all $(x_0, \varepsilon) \in \mathcal{G}$, compute the estimated expected number of approximate pseudo-regenerations:

$$\widehat{N}_n(x_0, \varepsilon) = \frac{\delta_n(x_0, \varepsilon)}{2\varepsilon} \sum_{i=1}^{n} \frac{\mathbb{I}\{(X_i, X_{i+1}) \in V_{x_0}(\varepsilon)^2\}}{p_n(X_i, X_{i+1})}, \tag{21}$$

with $\delta_n(x_0, \varepsilon) = 2\varepsilon . \inf_{(x,y)\in V_{x_0}(\varepsilon)^2} p_n(x, y)$. Then, pick $(x_0^*, \varepsilon^*) \in \mathcal{G}$ maximizing $\widehat{N}_n(x_0, \varepsilon)$ over $\mathcal{G}$, corresponding to the set $S^* = [x_0^* - \varepsilon^*,\ x_0^* + \varepsilon^*]$ and the minorization constant $\delta_n^* = \delta_n(x_0^*, \varepsilon^*)$. It remains next to construct the approximate pseudo-blocks using S^*, δ_n^* and p_n as described in § 2.3. We point out that other approaches may be considered for determining practically small sets and establishing accurate minorization conditions, which conditions do not necessarily involve uniform distributions besides. Refer for instance to [40] for Markov diffusion processes.

We end this paragraph by making the following remarks about the practical implementation of the ARBB method. We first emphasize that estimation in specific null recurrent cases (including AR(p) models with unit roots for instance) has been dealt with in [28], which established in particular consistency results for the *Nadaraya-Watson estimator* (18). And it is noteworthy that the procedures described above, the approximate Nummelin construction and the ARBB algorithm, are actually still asymptotically valid in this framework, when applied to adequate functions f. However, in the null recurrent case, the choice of the standardization may be cumbersome. Investigating the asymptotic properties of the ARBB at the first order, which corresponds to choosing $\sigma_n = l_n^{1/2}$ and $\sigma_n^* = l_n^{*1/2}$, may be done using the same approach as in [7] (the study of second order properties is currently in progress). As indicated by the results in [28], accurate estimation of the underlying transition density in the null recurrent case is naturally possible only when a very large data sample is at disposal. In β-null recurrent cases (*i.e.* when the distribution of the return time to the small set has power tail), [16] also established deterministic approximations of l_n (respectively, of l_n^*). To give an insight into the problems encountered in this case, we considered the case of an AR(1) model with a unit root among our simulation studies (see § 4.2): the number l_n of regenerations over a trajectory of length n for the split chain being of order $n^{1/2}$, only large sample sizes n enable us then to get enough (pseudo-) regeneration cycles for

computing significant statistics with our methodology.

Secondly, a natural question arising from the practical considerations discussed above is to determine whether the use of the preliminary estimate p_n and of $\widehat{\mu}_n$ eventually for selecting S and building the pseudo-blocks affect the second order properties of the resulting ARBB distribution. This seems to be a very difficult problem, since by construction the pseudo-regeneration times and the data blocks $\widehat{\mathcal{B}}_j$ they induce, all depend on the whole trajectory now, owing to the transition probability estimation step. A possible construction to avoid this theoretical problem consists in using a *double splitting trick* in a semiparametric sense (see [41]). This amounts first to construct the transition density estimator using the first m_n observations say (with $m_n \to \infty,\ m_n/n \to 0$ as $n \to \infty$), then to drop the next q_n observations (typically $q_n << m_n,\ q_n \to \infty$ as $n \to \infty$) for allowing the split chain to regenerate with overwhelming probability, and finally to build the pseudo-blocks $\widehat{\mathcal{B}}_j$ from the $n - m_n - q_n$ remaining observations. It is easy to understand (but technical to prove) that these blocks are then, asymptotically i.i.d conditionally to the first m_n observations. One may then prove the second order validity of the procedure in both the studentized and unstudentized cases. As shown in [6], this splitting trick entails some loss in the rate of the ARBB distribution, but the latter remains anyway faster than the best rate the MBB may achieve. However, one may argue, as in the semiparametric case, that such a modification of the initial procedure is essentially motivated by our limitations in the analysis of asymptotic properties of the estimators. From our own practical experience, this construction generally deteriorates the finite sample performance of the initial algorithm and estimating $p(x, y)$ from the whole trajectory leads to better numerical results.

4 Simulation studies

We now give two examples, with a view to illustrate the scope of applications of our methodology. The first example presents a regenerative Markov chain described and studied at greater length in [25] (see also [11] and [12]) for modeling storage systems, regenerative chains being widely used in operations research. We point out that our method also applies to the framework of MCMC (Monte-Carlo Markov Chain) in order to control estimates based on regenerative MCMC trajectories (see [32]). In consideration of the recent emphasis on nonlinear models in the time series literature, our second example shows to what extent the ARBB method may apply to a general nonlinear AR model. Further, we point out that the principles exposed in this paper are by no means restricted to the markovian setting, but may apply to any process for which a regenerative extension can be constructed and simulated from the data available (see Chapt. 10 in [44]).

4.1 Example 1 : content-dependent storage systems

We consider a general model for storage, evolving through a sequence of *input times* $(T_n)_{n\in\mathbb{N}}$ (with $T_0 = 0$ by convention), at which the storage system is replenished. Let S_n be the amount of input into the storage system at the n^{th} input time T_n and C_t be the amount of contents of the storage system at time t. When possible, there is withdrawal from the storage system between these input times at the constant rate r and the amount of stored contents that drops in a time period $[T,\, T + \Delta T]$ since the latter input time is equal to $C_T - C_{T+\Delta T} = r\Delta T$, and when the amount of contents reaches zero, it continues to take the value zero until it is replenished at the next input time. If X_n denotes the amount of contents immediately before the input time T_n (*i.e.* $X_n = C_{T_n} - S_n$), we have for all $n \in \mathbb{N}$,

$$X_{n+1} = (X_n + S_n - r\Delta T_{n+1})_+ \,, \tag{22}$$

with $(x)_+ = \sup(x, 0)$, $X_0 = 0$ by convention and $\Delta T_n = T_n - T_{n-1}$ for all $n \geq 1$. Let $K(x, ds)$ be a transition probability on $\mathbb{R}_+$. Assume that, conditionally to $X_1, ..., X_n$, the amounts of input $S_1, ..., S_n$ are independent from each other and independent from the inter-arrival times $\Delta T_1, ..., \Delta T_n$ and that the distribution of S_i is given by $K(X_i, .)$, for $0 \leqslant i \leqslant n$. Under the further assumption that $(\Delta T_n)_{n\geqslant 1}$ is an i.i.d. sequence with common distribution G, independent from $X = (X_n)_{n\in\mathbb{N}}$, the storage process X is a Markov chain with transition probability Π given by $\Pi(X_n,\, \{0\}) = \Gamma(X_n,\, [X_n,\, \infty[)$, $\Pi(X_n,\ ]x,\, \infty[) = \Gamma(X_n,\ ]-\infty,\, X_n - x[)$ for all $x > 0$, where the transition probability Γ is given by the convolution product $\Gamma(x,\,]-\infty,\, y[) = \int_{t=0}^{\infty}\int_{z=0}^{\infty} G(dt)K(x,\, dz)\mathbb{I}\{rt - z < y\}$.

One may check that the chain Π is δ_0-irreducible as soon as $K(x, .)$ has infinite tail for all $x \geqslant 0$. In this case, $\{0\}$ is an accessible atom for X and it can be shown that it is positive recurrent if and only if there exists $b > 0$ and a test function $V : \mathbb{R}_+ \to [0,\, \infty]$ such that $V(0) < \infty$ and for all $x \geqslant 0$:

$$\int \Pi(x, dy)V(y) - V(x) \leqslant -1 + b\mathbb{I}\{x = 0\}. \tag{23}$$

The times at which the storage process X reaches the value 0 are thus regeneration times, and allow to define regeneration blocks dividing the sample path, as shown in Figure 1. Figure 2 below shows a reconstructed RBB data series, generated by a sequential sampling of the regeneration blocks (as described in § 3.1), on which RBB statistics may be based.

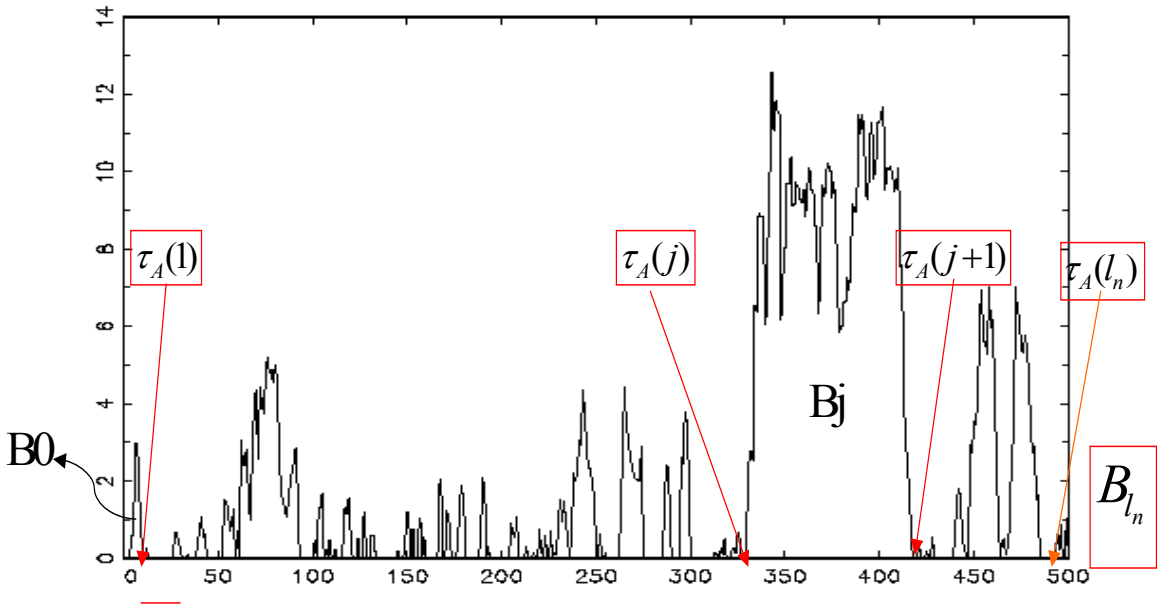

Fig. 1. Dividing the trajectory of the storage process into data blocks corresponding to the regeneration times.

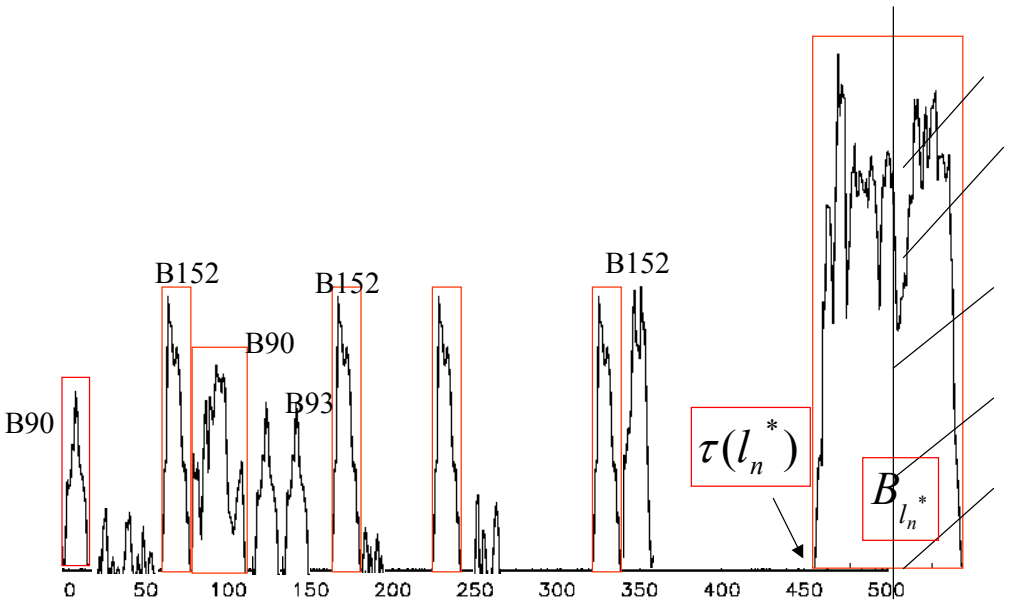

Fig. 2. Reconstruction of a storage process data series using the RBB resampling procedure.

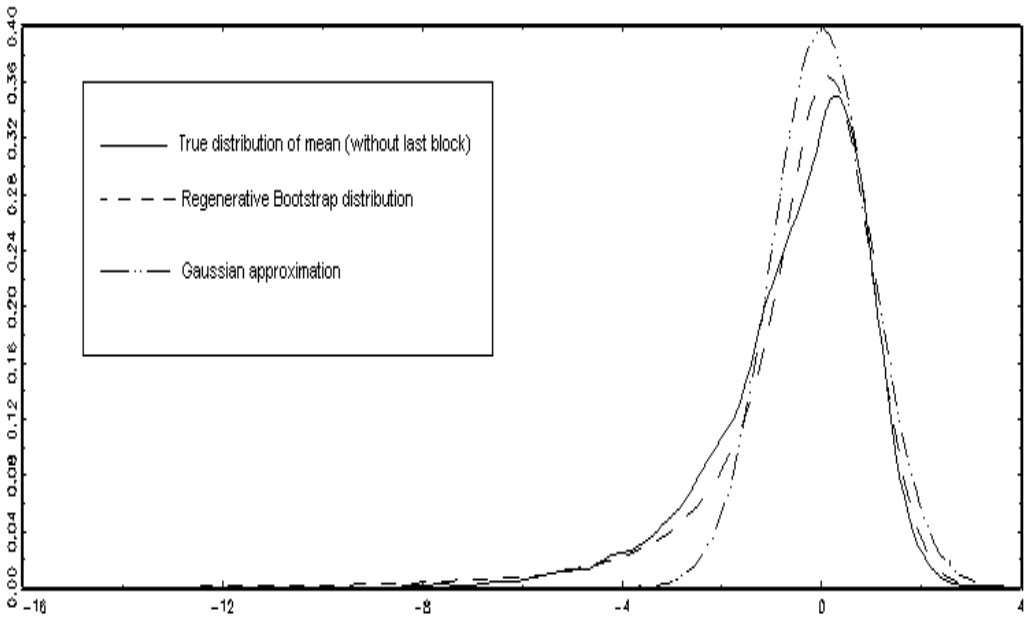

Fig. 3. Distribution estimates.

4.1.1 Simulation results

We simulated two trajectories of respective length $n = 100$ and $n = 200$ drawn from this Markov chain with $r = 1$, $K(x, dy) = Exp_3(dy)$ and $G(dy) = Exp_1(dy)$, denoting by $Exp_\lambda(dy)$ the exponential distribution with mean $1/\lambda > 0$, which is a standard M/M/1 model (see [1] for instance). In Fig. 3 below, a Monte-Carlo estimate of the true distribution of the sample mean standardized by its estimated standard error (as defined in (4)) computed with 10000 simulated trajectories is compared to the RBB distribution (in both cases, Monte-Carlo approximations of RBB estimates are computed from $B = 2000$ repetitions of the RBB procedure) and to the gaussian approximation. Note also that in the ideal case where one *a priori* knows the exact form of the markovian data generating process, one may naturally construct a bootstrap distribution in a parametric fashion by estimating first the parameters of the M/M/1 model, and then simulating bootstrap trajectories based on these estimates. Such an ideal procedure naturally performs very well in practice. In our simulation study, the resulting distribution estimate was actually so close to the true distribution that one could not distinguish one from the other in the plot. Of course, in most applications practitioners have generally no knowledge of the exact form of the underlying Markov model, since this is often one of the major goals of statistical inference.

With the aim of constructing accurate confidence intervals, Table 1 compares

the quantile of order γ of the true distribution, the one of the gaussian approximation (both estimated with 10000 simulated trajectories) and the mean of the quantile of order γ of the RBB distribution over 100 repetitions of the RBB procedure in the tail regions.

The left tail is clearly very well estimated, whereas the right tail gives a better approximation than the asymptotic distribution. The gain in term of coverage accuracy is quite enormous in comparison to the asymptotic distribution. For instance at the level 95%, for $n = 200$, the asymptotic distribution yields a bilateral coverage interval of level 71% only, whereas the RBB distribution yields a level of 92% in our simulation.

n=	100		200		∞	n=	100		200		∞
γ%	TD	RBB	TD	RBB	ASY	γ%	TD	RBB	TD	RBB	ASY
1	-7.733	-7.044	-5.492	-5.588	-2.326	90	1.041	1.032	1.029	1.047	1.282
2	-6.179	-5.734	-4.607	-4.695	-2.054	91	1.078	1.085	1.083	1.095	1.341
3	-5.302	-5.014	-4.170	-4.165	-1.881	92	1.125	1.145	1.122	1.150	1.405
4	-4.816	-4.473	-3.708	-3.757	-1.751	93	1.168	1.207	1.177	1.209	1.476
5	-4.374	-4.134	-3.430	-3.477	-1.645	94	1.220	1.276	1.236	1.277	1.555
6	-4.086	-3.853	-3.153	-3.243	-1.555	95	1.287	1.360	1.299	1.356	1.645
7	-3.795	-3.607	-2.966	-3.045	-1.476	96	1.366	1.453	1.380	1.442	1.751
8	-3.576	-3.374	-2.771	-2.866	-1.405	97	1.433	1.568	1.479	1.549	1.881
9	-3.370	-3.157	-2.606	-2.709	-1.341	98	1.540	1.722	1.646	1.685	2.054
10	-3.184	-2.950	-2.472	-2.560	-1.282	99	1.762	1.970	1.839	1.916	2.326

Table 1 : Comparison of the tails of the true distribution (TD), RBB and gaussian distributions.

4.2 Example 2 : General autoregressive models

Consider now the general heteroscedastic autoregressive model

$$X_{n+1} = m(X_n) + \sigma(X_n)\varepsilon_{n+1},\ n \in \mathbb{N}, \tag{24}$$

where $m : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}_+^*$ are measurable functions, $(\varepsilon_n)_{n\in\mathbb{N}}$ is a i.i.d. sequence of r.v.'s drawn from $g(x)dx$ such that, for all $n \in \mathbb{N}$, ε_{n+1} is

independent from the X_k's, $k \leqslant n$ with $\mathbb{E}[\varepsilon_{n+1}] = 0$ and $var(\varepsilon_{n+1}) = 1$. See [20] for some proposals for bootstrapping such models. The transition density of the chain is given by $p(x, y) = g((y - m(x))/\sigma(x))$, $(x,\ y) \in \mathbb{R}^2$. Assume further that g, m and σ are continuous functions and there exists $x_0 \in \mathbb{R}$ such that $p(x_0, x_0) > 0$. Then, the transition density is uniformly bounded from below over some neighborhood $V_{x_0}(\varepsilon)^2 = [x_0 - \varepsilon, x_0 + \varepsilon]^2$ of $(x_0,\ x_0)$ in $\mathbb{R}^2$: there exists $\delta = \delta(\varepsilon) \in]0,\ 1[$ such that,

$$\inf_{(x,y)\in V_{x_0}^2} p(x, y) \geqslant \delta(2\varepsilon)^{-1}. \tag{25}$$

Any compact interval $V_{x_0}(\varepsilon)$ is thus a small set for the chain X, which satisfies the minorization condition $\mathcal{M}(1, V_{x_0}(\varepsilon), \delta, \mathcal{U}_{V_{x_0}(\varepsilon)})$, where $\mathcal{U}_{V_{x_0}(\varepsilon)}$ denotes the uniform distribution on $V_{x_0}(\varepsilon)$. Hence, in the case when one knows x_0, ε and δ such that (5) holds (this simply amounts to know a uniform lower bound estimate for the probability to return to $V_{x_0}(\varepsilon)$ in one step), one may effectively apply the ARBB methodology to X. In the following, we use the practical criterion $\widehat{N}_n(x_0, \varepsilon)$ with $x_0 = 0$. The choice $x_0 = 0$ is simply motivated by observing that our temporal simulated data fluctuate around 0. Actually, to our own practical experience, optimizing over x_0 does not really improve the performance of the procedure in this case.

In what follows, we shall compare the performance of the ARBB to the one of some reference competitors for bootstrapping time series. In all our simulations the Markov bootstrap (consisting in generating a Markov chains with an estimated transition probability) has performed always worse than all the other methods (due to the difficulty of estimating accurately the transition probability on the whole real line). We do not present the results for this method to alleviate the graphics and tables.

The *sieve bootstrap* is specifically tailored for linear time series (see [13], [14]). The main idea consists in fitting an AR(p) model (eventually with p unknown depending on) first and then applying a residual based resampling method. The fact that it fully exploits the underlying linear structure explains why it performs very well in this framework. When simulating linear time series, we use it as a benchmark for evaluating the pertinence of the ARBB distribution. Recall also that this method requires a preliminary estimation of the order q of the sieve : for this purpose we choose an AIC criterion of the type $AIC(q) = nlog(\widehat{MSE}) + 2q$ in the sequel. In the linear AR(q) model below, this information criterion enables us to pick the right order of the model. And the resulting sieve bootstrap behaves like a parametric bootstrap method in these cases (see [10]), leading to very good numerical results, as soon as the roots of the $AR(q)$ model are far from the unit circle. In contradistinction, we actually experienced problems in our simulations, when dealing with an

AR(1) model with a root close to 1: in such cases, it may happen with high probability that one gets an estimate of the root larger than one, yielding to explosive bootstrap trajectories.

We also compared the ARBB method to the usual MBB. The difficulty for applying the latter method essentially relies in the choice of the block size for estimating the variance and in the choice of the block size for the resampling procedure. As there is actually no reason for these two sizes to be equal, they should be picked separately and the estimator of the variance should be correctly unbiased (see [22]). To our knowledge, the problem of simultaneously calibrating these two quantities has not been treated yet and leads to extremely volatile results. For comparing directly the MBB distribution to the true studentized distribution (11), we have chosen here to standardize all the distributions by the estimator (4), so as to avoid a deteriorating preliminary variance estimation step. The MBB distribution is also correctly centered (at the bootstrap mean). The block size for the MBB is chosen according to the method of [24]. It consists in estimating first the MSE of the MBB distribution corresponding to blocks of size l with a subsampling technique for various size values l and then picking the size corresponding to a minimum MSE estimate. This unfortunately requires to select a subsampling size and a plausible pilot size, which are in their turn also difficult to calibrate (see the discussion in Section 7.3 of [29]): here we have chosen $n^{1/4}$ as pilot size and $b_n = n^{10/21}$ as subsampling size (which is close to $n^{1/2}$ in our simulations and satisfies the conditions needed for the MBB to be asymptotically valid). When standardized this way, the MBB has performed quite well in most simulations, except notably when data exhibit significant nonlinear features and/or nonstationarity. The reason of this misbehavior arises from the fact that, for some drawing of the fixed size blocks, the jumps between the blocks were so important, that the reconstructed series could not be splitted according to our randomized procedure leading to an invalid estimator of the variance. In these case (too few regenerations), we have eliminated the corresponding MBB simulation. Thus the MBB considered here can be considered as a MBB with a Markovian control ensuring that the MBB reconstructed series has some regeneration properties. Such procedure clearly improved the resulting estimated distributions.

4.2.1 Simulation results

Here are empirical evidences for three specific autoregressive models.

The $AR(1)$ model :

$$X_{i+1} = \alpha X_i + \varepsilon_{i+1},\ i \in \mathbf{N}, \tag{26}$$

with i.i.d. $\epsilon_i \sim \mathcal{N}(0,1)$, $\alpha = 0.8$, $X_0 = 0$ and for a trajectory of length $n = 200$.

The $AR(1)$ model with $ARCH(1)$ residuals called *AR-ARCH model*:

$$X_{i+1} = \alpha X_i + (1 + \beta X_i^2)^{1/2} \varepsilon_{i+1,}\, i \in \mathbf{N}, \tag{27}$$

with i.i.d. $\varepsilon_i \sim \mathcal{N}(0,1)$, $\alpha = 0.6$, $\beta = 0.35$, $X_0 = 0$ and for a trajectory of length $n = 200$.

The so called *ExpAR(1) model*

$$X_{i+1} = (\alpha_1 + \alpha_2 e^{-|X_i|^2}) X_{i+1} + \varepsilon_{i+1,}\, i \in \mathbf{N}, \tag{28}$$

with i.i.d. $\varepsilon_i \sim \mathcal{N}(0,1)$, $\alpha_1 = 0.6$, $\alpha_2 = 0.1$, $X_0 = 0$ and for a trajectory of length $n = 200$. Such a chain is recurrent positive under the sole assumption that $|\alpha_1| < 1$, see [43]. This highly nonlinear model behaves like a threshold model: when the chain takes large values, this is almost an AR(1) model with coefficient α_1, whereas for small values, it behaves as an AR(1) model with a larger autoregressive coefficient $\alpha_1 + \alpha_2$.

Here the true distribution of the sample mean is estimated with 10000 simulations. And for a given trajectory, the ARBB distribution is approximated with $B = 1000$ resamplings of the pseudo-blocks. In a previous simulation work, we experienced that the ARBB distribution obtained may strongly fluctuate, depending on the randomization steps (see §2.3). For a given trajectory, this problem may be avoided by repeating the ARBB procedure several times (50 times in our simulations) and averaging the resulting ARBB distribution estimates. According to our experiments, only a small number of repetitions (leading to different ways of dividing the same trajectory) suffices for smoothing the ARBB distribution.

For the ARBB, the sieve and the MBB methods, the whole procedure has been repeated 1000 times. Table 2 below gives the median of the quantiles at several orders γ of the bootstrap distributions over the 1000 replications for each of the three AR models, compared to the true and asymptotic corresponding quantiles.

n=200	AR(1)				AR-ARCH(1)				EXP-AR(1)				
$\gamma\%$	TD	ARBB	Sieve	MBB	TD	ARBB	Sieve	MBB	TD	ARBB	Sieve	MBB	ASY
1	-3.51	-3.61	-3.41	-3.42	-3.03	-3.23	-5.26	-3.16	-4.48	-5.23	-5.59	-9.61	-2.33
2..5	-2.84	-2.78	-2.81	-2.72	-2.41	-2.61	-3.52	-2.52	-3.35	-3.87	-4.76	-6.44	-1.96
5	-2.23	-2.13	-2.11	-2.10	-1.97	-2.14	-2.85	-2.06	-2.58	-2.79	-3.74	-5.00	-1.65
10	-1.62	-1.57	-1.65	-1.55	-1.52	-1.59	-2.25	-1.53	-1.83	-1.98	-2.93	-3.51	-1.28
n=200	AR(1)				AR-ARCH(1)				EXP-AR(1)				
$\gamma\%$	TD	ARBB	Sieve	MBB	TD	ARBB	Sieve	MBB	TD	ARBB	Sieve	MBB	ASY
90	1.62	1.52	1.61	1.61	1.52	1.33	2.26	1.58	1.80	1.89	2.74	2.26	1.28
95	2.21	2.08	2.19	2.14	2.01	1.74	3.04	2.07	2.58	2.68	3.89	3.07	1.65
97.5	2,79	2.71	2.73	2.69	2.37	2.03	3.93	2.44	3.24	3.47	4.79	4.02	1.96
99	3.46	3.73	3.86	3.48	3.11	2.62	5.97	3.22	4.37	5.36	5.92	6.25	2.33

Table 2: Comparison of the tails of the true, ARBB and gaussian distributions for the three models

The small set is selected by maximizing over $\varepsilon > 0$ the empirical criterion $\widehat{N}_n(0,\varepsilon)$ described above. The main steps of the procedure are summarized in the graph panels shown below.

The first figure in Graph panel 1 shows the Nadaraya-Watson (NW) estimator (18), the second one represents $\widehat{N}_n(0,\varepsilon)$ as ε grows (as well as the smoother empirical criterion (17), see the dotted line). It clearly allows to identify an optimal value for the size of the small set. In the case of the AR model for instance, this selection rule leads to pick in mean $\widehat{\varepsilon} = 0.83$ and $\widehat{\delta} = 0.123$. Our empirical criterion tends to overestimate very slightly the size of the "optimal" small set (a phenomenon that we have noticed on several occasions in our simulations). The level sets of the NW estimator, the data points (X_i, X_{i+1}) and the estimated small set are represented in the next graphic. This also shows that the small set chosen may be not that "small" if the transition density is flat around $(x_0, x_0) = (0, 0)$ (in some cases it may be thus preferable to choose $x_0 \neq 0$ so as to be in this situation). In the second line of the panel, the figure on the left hand side represents a sample path of the chain and indicates the pseudo-regenerative blocks obtained by applying the randomization rule with $Ber(1 - \widehat{\delta}(2\varepsilon)^{-1}/\ p_n(X_i, X_{i+1}))$ at times i when $(X_i, X_{i+1}) \in V_0(\varepsilon)^2$. The next figure shows how binded blocks form a typical ARBB trajectory. It is noteworthy that such a trajectory presents less artificial "jumps" than a trajectory reconstructed from a classical MBB procedure: by construction, blocks are

joined end to end at values belonging to the small set. For comparison purpose, the figure on the right hand side displays a typical realization of a MBB trajectory. Finally, on the last line of the panel, the true distribution (green), the ARBB distribution (black), the sieve bootstrap distribution (gray), the MBB distribution (red dotted line) and the asymptotic gaussian distribution (blue dotted line) are compared.

And the last figure shows the QQ-plots $\alpha \in [0,1] \mapsto G_n(H^{-1}(\alpha))$, where H is the true distribution and G_n denotes one of the approximations: this enables us to discriminate between the various approximations in a sharper fashion, especially in the tail regions.

These results clearly indicate that both the sieve and MBB methods perform very well for linear time series. In this case, the ARBB distribution tends to have larger tails. However, when considering nonlinear models, the advantage of the ARBB method over its rivals plainly come into sight: for moderate sample sizes n, the sieve bootstrap tends to choose a too large value $\widehat{q}_n$ for the lag order of the approximate sieve $AR(\widehat{q}_n)$. This problem is less serious for larger sample sizes, as shown in Graph panel 4 (with $n = 500$). In these situations, the MBB may behave very poorly especially when the non-linearity and the non-stationarity is important: we conjecture that it could be possibly improved by investigating further how to tune optimally the block size, especially for standardized distributions.

Pictures in Graph panels 3 and 4 speak volumes: for both nonlinear models, the true distribution is accurately approximated by the ARBB distribution. Note nevertheless the difference in the size of the "optimal small set" and in the number of pseudo-regenerations between these models. We point out that, though remarkable when compared to the gaussian approximation, the gain in accuracy obtained by applying the ARBB methodology to the EXP-AR model is higher than the one obtained for the AR-ARCH type model. As may be confirmed by other simulations, the ARBB method provides less accurate results for a given (moderate) sample size, as one gets closer to a unit root model (*i.e.* as α tends to 1): one may get an insight into this phenomenon by simply noticing that the rate of the number of regenerations (respectively, of the number of visits to the small set) then drastically decreases.

5 Concluding remarks

We finally summarize our empirical findings. We first point out that, in the linear case when roots are much less than 1 in amplitude, the sieve bootstrap clearly surpasses its competitors. But it is noteworthy that both the ARBB and the MBB also provides very good numerical results in this case.

Besides, all these methods seem to break down from a practical viewpoint for an AR(1) model with an autoregressive coefficient α tending to 1 and with a fixed (moderate) sample size: in such a case, too few pseudo-regeneration blocks may be constructed for the ARBB methodology to be practically performant (although it is asymptotically valid). In this respect, the graph of the estimated number of pseudo-regenerations (see Graph panels 1-4) provides a crucial help for diagnosing the success or the failure of the ARBB method. It is also remarkable that the sieve bootstrap can lead to very bad results in this case, due to the fact that the estimated AR model may have a root larger than 1 (generating then explosive sieve bootstrap trajectories). This strongly advocates the use of preliminary tests or constrained estimation procedures (ensuring that the resulting reconstructed series is asymptotically stationary).

And as may be reported from our simulation results, the advantage of the ARBB over the sieve bootstrap, the MBB and the asymptotic distributions, clearly appears when dealing with nonlinear models even if in some case the MBB can still give some good approximation (see the AR-ARCH(1) case, Graph-panel 2). Even if the lag is chosen very large (in mean 85 for the AR-ARCH(1) model and 21 for the EXP-AR model), the linear sieve method is unable to capture the non-linearities and performs very badly for moderate sample sizes. The MBB also performs poorly in some nonlinear setting for moderate sample sizes, whereas the ARBB provides very accurate approximations of the tail distributions in these examples. It should be mentionned that using a moving-block estimator of the variance leads to even worse results. In any case, it is recommended to use all the available methods and to compare the results. The ARBB being much more robust it can be used to check whether the other methods are trustable for the data at hand.

As pointed out by a referee, we restricted our study to the case of Markov models of order 1 (see our framework in section 2). However, by vectorizing, any 1-dimensional Markov model of order p classically boils down to a p dimensional Markov model of order 1. Our theoretical work thus applies in this context. However, statistical problems related to the curse of dimensionality may appear, when considering models of large orders (arising in the choice of the small set or the transition density estimation step), Hence, numerical experiments should be carried out for such models, in order to determine whether theoretical results are still supported by empirical evidence. This is beyond the scope of the present paper, but will certainly be the subject of further investigation. Besides, another challenging line of research could consist in determining whether the ARBB performs well when applied to long memory Markov chains. As a matter of fact, this problem is of different nature. Recall that a Markov model of order 1 may naturally have a long memory: if X is a positive recurrent chain with limiting distribution μ and possesses an atom A, the long memory property for the sequence $\{f(X_n)\}_{n\in\mathbb{N}}$ is equivalent to the condition $\mathbb{E}_A[(\sum_{i\leq\tau_A}\{f(X_i)-\mu(f)\})^2]=\infty$. No theoretical result for the

(A)RBB asymptotic validity in this framework has yet been established, even if one may reasonably expect that it is the case for β-null recurrent chains (which exhibit long range dependence when $\beta < 1$, see [28]).

Akcnowledgements. We thank the referees for their valuable remarks and suggestions.

References

[1] Asmussen, S. (1987). Applied Probabilities and Queues. Wiley.

[2] Athreya, K.B., Atuncar, G.S. (1998). Kernel estimation for real-valued Markov chains. Sankhya, 60, series A, No 1, 1-17.

[3] Athreya, K.B., Fuh, C.D. (1989). Bootstrapping Markov chains: countable case. Tech. Rep. B-89-7, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, ROC.

[4] Bertail, P., Clémençon, S. (2004a). Edgeworth expansions for suitably normalized sample mean statistics of atomic Markov chains. Prob. Th. Rel. Fields, 130, 388-414.

[5] Bertail, P., Clémençon, S. (2004b). Note on the regeneration-based bootstrap for atomic Markov chains. To appear in Test.

[6] Bertail, P. , Clémençon, S. (2004c). Approximate Regenerative Block-Bootstrap for Markov Chains: second-order properties. In Compstat 2004 Proc. Physica Verlag.

[7] Bertail, P. , Clémençon, S. (2006). Regenerative Block Bootstrap for Markov Chains. Bernoulli, 12, No. 4, 689-712.

[8] Bertail, P. , Clémençon, S. (2006). Regeneration-based Statistics for Markov Chains. In Dependence in Probability and Statistics , Eds P. Bertail, P. Doukhan & P. Soulier, 1-53. Springer.

[9] Bölthausen, E. (1980). The Berry-Esseen Theorem for strongly mixing Harris recurrent Markov Chains. Z. Wahr. Verw. Gebiete, 54, 59-73.

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

[11] Brockwell, P.J., Resnick, S.J., Tweedie, R.L. (1982). Storage processes with general release rules and additive inputs. Adv. Appl. Probab., 14, 392-433.

[12] Browne, S., Sigman, K. (1992). Work-modulated queues with applications to storage processes. J. Appl. Probab., 29, 699-712.

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

[14] Bühlmann, P. (2002). Bootstrap for time series. Stat. Sci., 17, 52-72.

[15] Bühlmann, P., Künsch, H. (1999). Block length selection in the bootstrap for time series. Comp. Statist. Data Analysis, 31, 295-310.

[16] Chen, X. (1999). How often does a Harris Recurrent Markov Chain recur? Ann. Probab., 3, 1324-1346.

[17] Clémençon, S. (2000). Adaptive estimation of the transition density of a regular Markov chain. Math. Meth. Stat., 9, No. 4, 323-357.

[18] Datta, S., McCormick W.P. (1993). Regeneration-based bootstrap for Markov chains. Can. J. Statist., 21, No.2, 181-193.

[19] Efron, B. (1979). Bootstrap methods: another look at the jackknife Journal. Ann. Stat. Vol. 7, 1-26

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

[21] Götze, F., Hipp, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. Zeit. Wahrschein. verw. Geb. , 64, 211-239.

[22] Götze, F., Künsch, H.R. (1996). Second order correctness of the blockwise bootstrap for stationary observations. Ann. Statist., 24, 1914-1933.

[23] Hall, P. (1992). The Bootstrap and Edgeworth Expansion. Springer.

[24] Hall, P., Horowitz, J., Jing, B.-Y. (1995), On blocking rules for the bootstrap with dependent data. Biometrika, 82, 561-574.

[25] Harrison, J.M., Resnick, S.J. (1976). The stationary distribution and first exit probabilities of a storage process with general release rule. Math. Oper. Res., 1, 347-358.

[26] Hobert, J.P., Jones, G.L., Presnell, B., Rosenthal, J.S. (2002). On the applicability of regenerative simulation in Markov chain Monte Carlo. Biometrika, 89, 731-743.

[27] Jain, J., Jamison, B. (1967). Contributions to Doeblin's theory of Markov processes. Z. Wahrsch. Verw. Geb., 8, 19-40.

[28] Karlsen, H.A., Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. Ann. Statist., 29 (2), 372-416.

[29] Lahiri, S.N. (2003). Resampling methods for dependent Data. Springer.

[30] Malinovskii, V. K. (1987). Limit theorems for Harris Markov chains I. Theory Prob. Appl., 31, 269-285.

[31] Meyn, S.P., Tweedie, R.L., (1996). Markov chains and stochastic stability. Springer.

[32] Mykland, P., Tierney, L., Yu, B. (1995). Regeneration in Markov chain samplers. J. A.S.A., 90, 233-241.

[33] Nummelin, E. (1978). A splitting technique for Harris recurrent chains. Z. Wahrsch. Verw. Gebiete, 43, 309-318.

[34] Nummelin, E. (1984). General irreducible Markov chains and non negative operators. Cambridge University Press, Cambridge.

[35] Paparoditis, E., Politis, D.N. (2002). The local bootstrap for Markov processes. J. Statist. Plan. Infer., 108, 301-328.

[36] Politis, D.N. (2003). The impact of bootstrap methods on time series analysis. Statistical Science, 18, No. 2, 219-230.

[37] Politis, D.N., White, H. (2004). Automatic Block-Length Selection for the Dependent Bootstrap. Econometric Reviews, Vol. 23, No. 1, 53-70.

[38] Rajarshi, M.B. (1990). Bootstrap in Markov-sequences based on estimates of transition density. Ann. Instit. Statist. Math., 42, No. 2, 253-268.

[39] Revuz, D. (1984). Markov chains. North-Holland, 2nd edition.

[40] Roberts, G.O., Rosenthal, J.S. (1996). Quantitative bounds for convergence rates of continuous time Markov processes. Electr. Journ. Prob., 9, 1-21.

[41] Schick, A. (2001). Sample splitting with Markov Chains. Bernoulli, 7, (1), 33-61.

[42] Smith, W. L. (1955). Regenerative stochastic processes. Proc. Royal Stat. Soc., A, 232, 6-31.

[43] Tjøstheim, D. (1990). Non Linear Time series, Adv. Appl. Prob., 22, 587-611.

[44] Thorisson, H. (2000). Coupling, Stationarity and Regeneration. Springer.

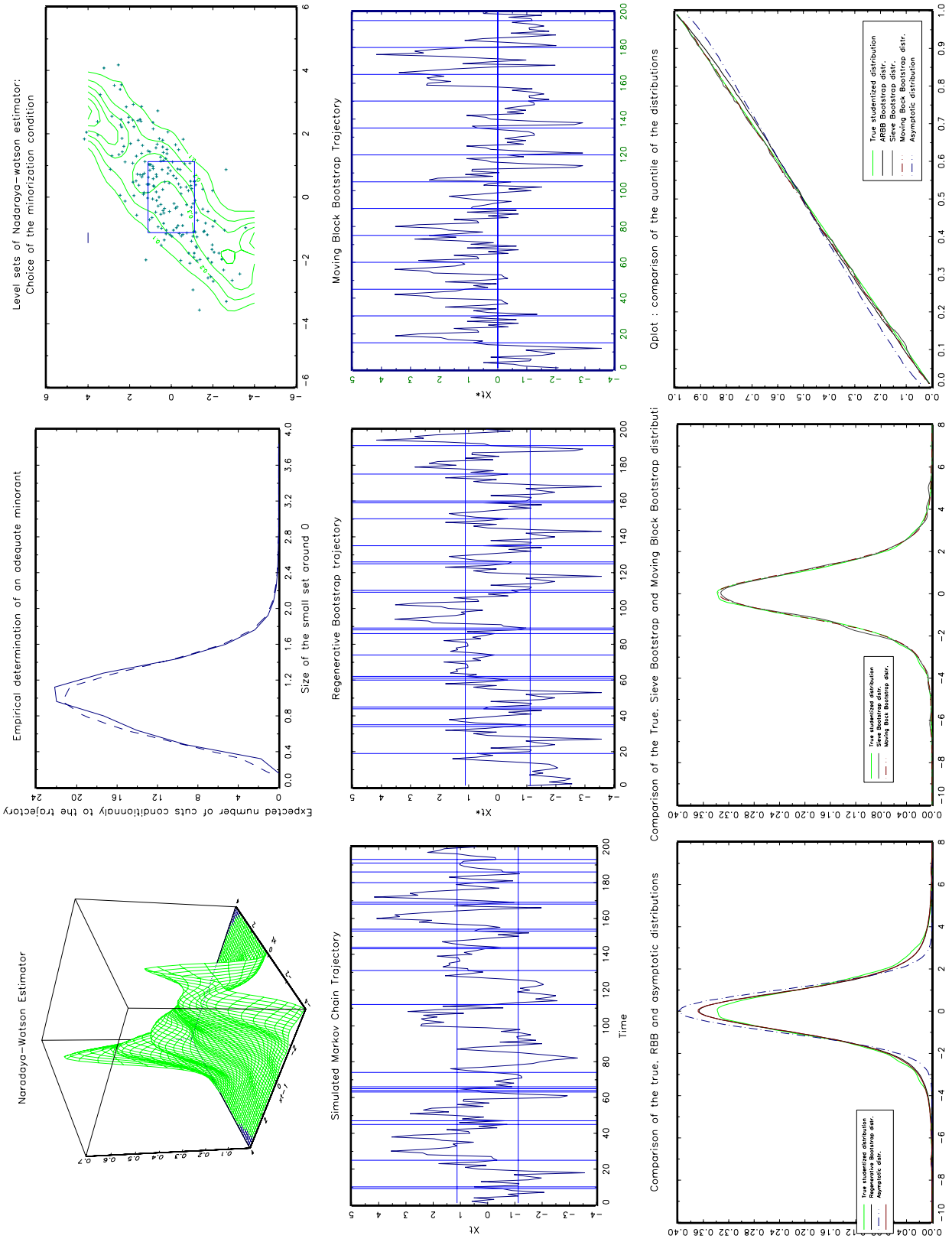

Fig. 4. Graph panel 1: AR(1) model with $\alpha = 0.8,\ n = 200$.

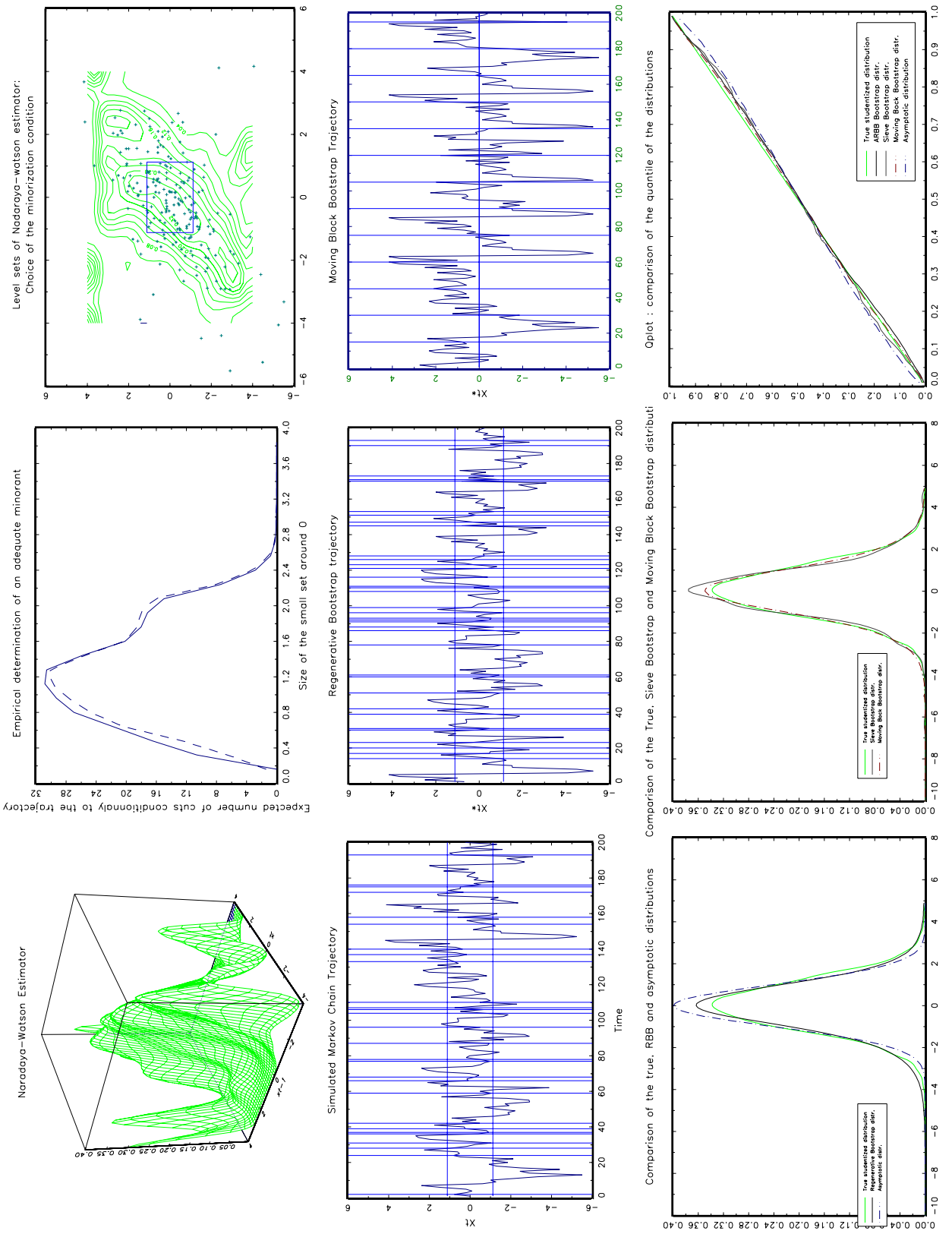

Fig. 5. Graph panel 2: AR(1)-ARCH(1) model with $\alpha = 0.6$ and $\beta = 0.35,\ n = 200$.

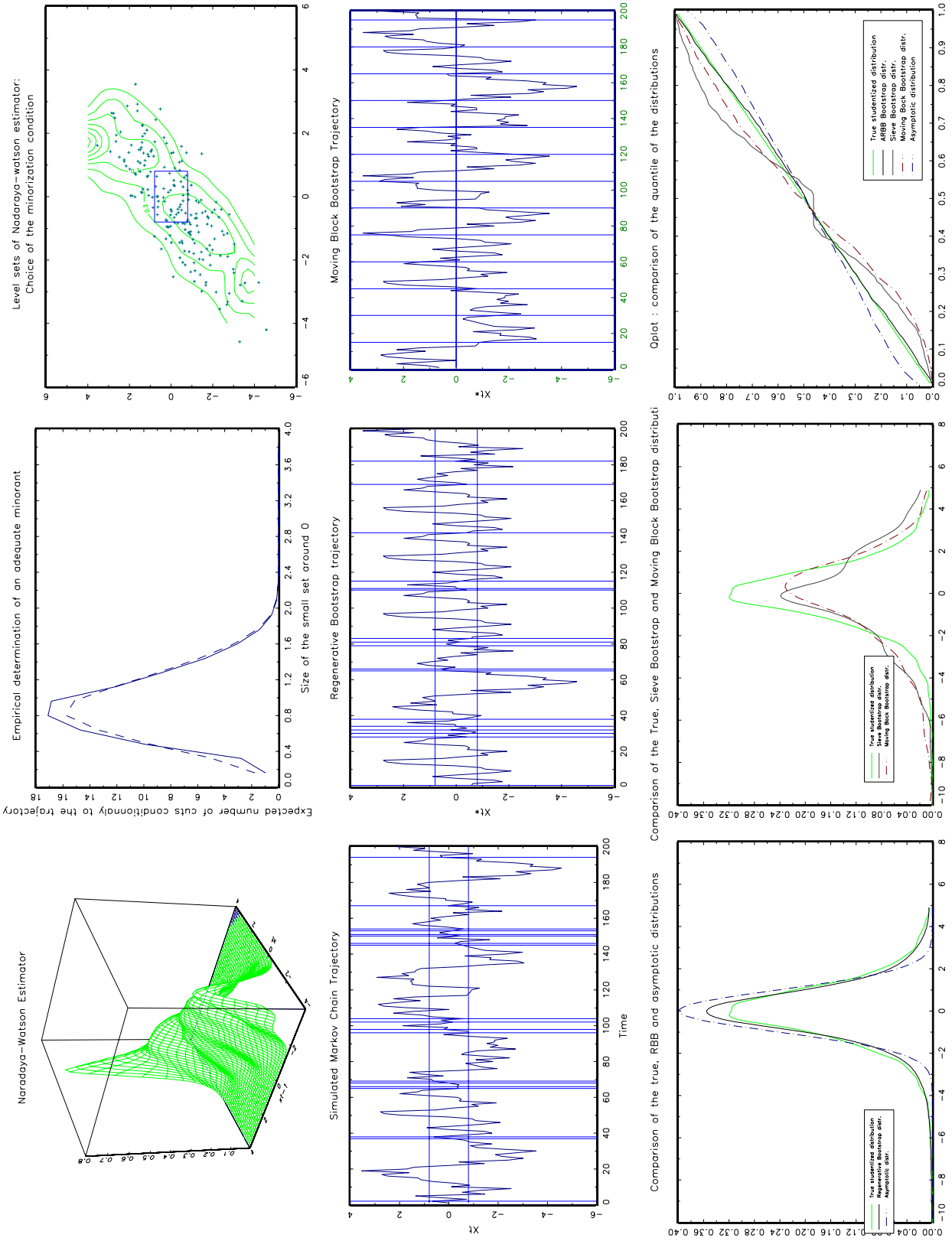

Fig. 6. Graph panel 3: EXP-AR(1) model with $\alpha_1 = 0.8$ and $\alpha_2 = 0.5,\ n = 200$.

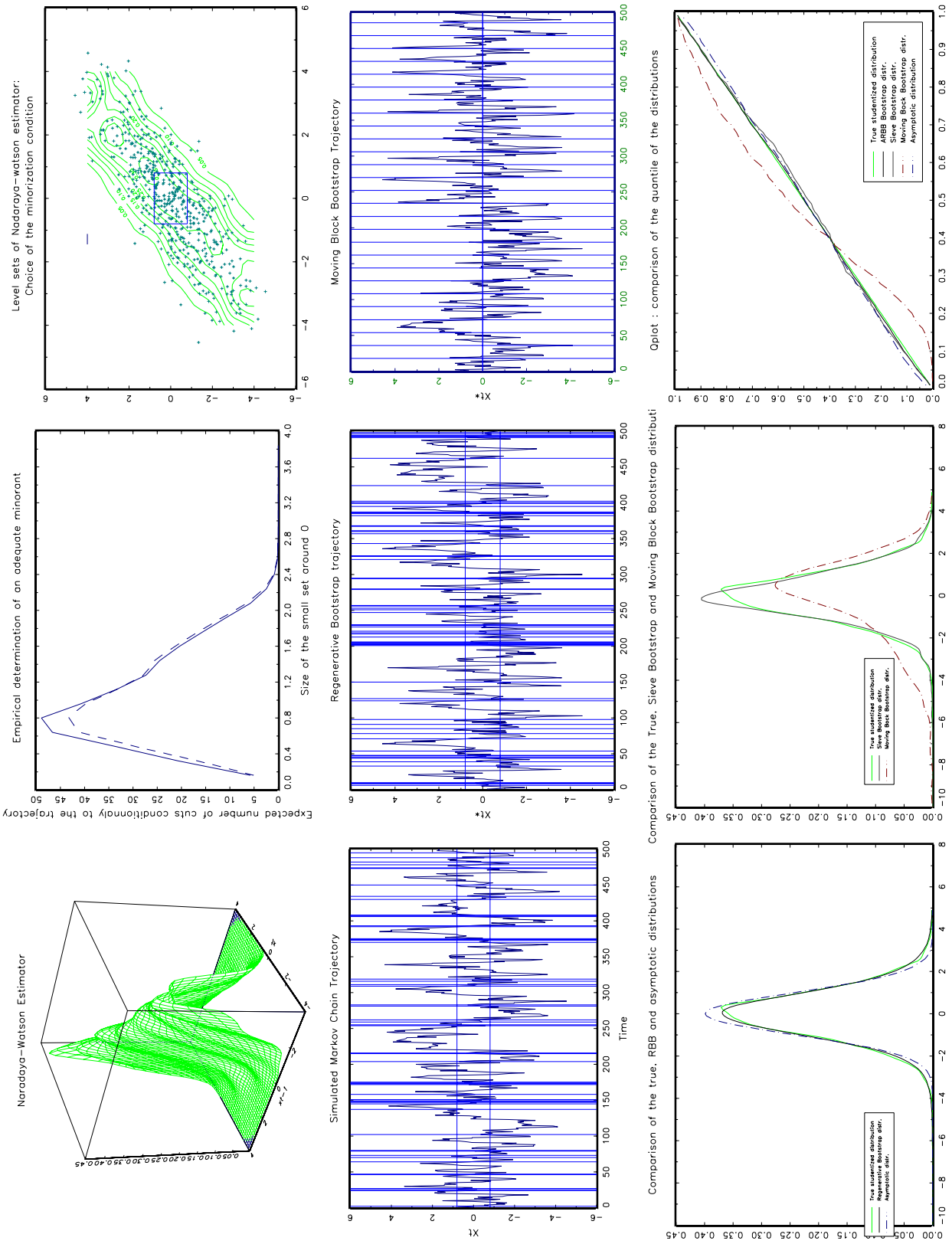

Fig. 7. Graph panel 3: EXP-AR(1) model with $\alpha_1 = 0.8$ and $\alpha_2 = 0.5,\ n = 500$.