



HAL
open science

Regenerative Bootstrap for Markov chains in high dimension

Patrice Bertail, Anna E. Dudek, Karolina Łukasiewicz

► To cite this version:

Patrice Bertail, Anna E. Dudek, Karolina Łukasiewicz. Regenerative Bootstrap for Markov chains in high dimension. 2026. <hal-05556914v2>

HAL Id: hal-05556914

<https://hal.science/hal-05556914v2>

Preprint submitted on 21 May 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Regenerative Bootstrap for Markov chains in high dimension

Patrice Bertail^a, Anna E. Dudek^{b,c}, Karolina Łukasiewicz^{a,1,*}

^a*MODAL'X, Université Paris-Nanterre, Paris, France*

^b*Aix Marseille University, CNRS, I2M, Marseille, France, Marseille, France*

^c*Faculty of Applied Mathematics, AGH University of Krakow, al. Mickiewicza 30, Krakow, 30-059, Poland*

Abstract

Large-dimensional Markov chains appear in many models and many applications. In this paper, we introduce the Multidimensional Approximative Regenerative Block Bootstrap (MARBB), a bootstrap algorithm designed for high-dimensional Markov chains. We focus, in this paper, on a vector autoregressive (VAR(1)) process with a low-rank structure. We first use a reduction algorithm that transforms the original high-dimensional time series into a lower-dimensional Markov chain. Once the chain is reduced, we leverage the regenerative properties of Harris recurrent Markov chains within a general state space, using the Nummelin splitting technique to extend existing results from the one-dimensional settings to the multidimensional case. This approach enables the identification of approximate regeneration times of the lower-dimensional Markov chain, which in turn leads to the splitting of the original high-dimensional Markov chain into approximate regenerative blocks. These blocks are then used to bootstrap relevant statistics based on regenerative blocks. Finally, we give the MARBB consistency results, and we apply our algorithm to simulation data.

Keywords: Multidimensional Markov chains, Approximate Regenerative Block Bootstrap, Dimensionality reduction, Small set, Nummelin splitting technique, VAR process

*Corresponding author

Email addresses: `patrice.bertail@gmail.com` (Patrice Bertail),
`aedudek@agh.edu.pl` (Anna E. Dudek), `karolina.lukasiewicz@agh.edu.pl` (Karolina Łukasiewicz)

1. Introduction and motivations

Markov chains are widely used in stochastic modelling due to their mathematical elegance, computational feasibility, and broad applicability [2, 9, 16]. Markov chains are particularly useful in many practical areas, such as queuing theory, manufacturing processes, and inventory management.

In this paper, we focus on multidimensional Markov chains. The high-dimensional settings are particularly important because they arise naturally in many applications. For example, in the case of studying multiple time series simultaneously, which is crucial in areas such as macroeconomics, electricity load and price forecasting, linguistics (e.g., word2vec models), electroencephalogram (EEG), and social media data.

One goal of this paper is to introduce a tailor-made bootstrap method for multidimensional Markov chains. In the case of dependent data, we can apply one of the well-known block bootstrap methods, for example, the Moving Block Bootstrap (MBB), Nonoverlapping Block Bootstrap (NBB) or Circular Block Bootstrap (CBB) (see [14]). Unfortunately, these methods have some drawbacks: they require the choice of a block length, which can be difficult in practice. An alternative approach is to use bootstrap algorithms specifically designed for Markov chains (see [5, 14]). In the sequel, we will focus on the method proposed in [5], which uses the regenerative properties of Harris recurrent Markov chains. Indeed, for such chains, there always exists an extension of the process with a set of states called an atom such that the process becomes independent of its past after each visit to this set (see [4] and [8]).

The method of [8] was improved by [5], who proposed the Regenerative Block Bootstrap (RBB), a sequential algorithm achieving higher-order accuracy close to the rate $O(n^{-1})$. Moreover, the extension of the RBB method to the general state of space was proposed in [5], and is called the Approximate Regenerative Block Bootstrap (ARBB). A key feature of both RBB and ARBB is the construction of approximate regeneration times, which enables the decomposition of the Markov chain trajectory into blocks that are asymptotically independent. Unlike MBB, NBB, and CBB, the RBB and ARBB methods do not require the choice of block lengths, since the block sizes are determined by the data itself.

The bootstrap method proposed in this paper can be viewed as a multi-dimensional extension of the ARBB. While ARBB was originally developed for one-dimensional Markov chains, its direct application to high-dimensional settings faces significant challenges. To illustrate these difficulties, we briefly recall the idea of the ARBB method. The ARBB algorithm divides the trajectory into blocks using a constructive version of the so-called Nummelin splitting trick (see [18]), based on an estimate of the transition density. These blocks are then used to perform the ARBB by drawing them independently with replacement (see [5]).

However, when Markov chain takes values in high-dimensional space, this constructive splitting trick may become difficult or even infeasible to implement. In particular, the estimator of the transition density may converge very slowly, and identifying an appropriate regeneration set may be challenging (see Subsection 2.3).

The main idea of the proposed MARBB procedure is therefore to combine the ARBB framework with a dimension reduction step. More precisely, we project the original high-dimensional process onto a lower-dimensional space using the Truncated Singular Value Decomposition. This transformation yields a reduced Markov chain for which the Nummelin splitting technique, with estimated parameters, can be applied more efficiently. In this way, we are able to determine approximate regeneration times of the reduced Markov Chains, which appear to be also regeneration times of the original process. It is worth noting that the reduced Markov chain does not need to be one-dimensional. Consequently, the ARBB method must be extended to handle a multidimensional process with a low-rank structure. To the best of our knowledge, this setting has not been previously studied. The paper therefore makes two main contributions. First, we develop a practical method for identifying approximate regeneration times in high-dimensional settings, including the estimation of transition probabilities in reduced dimensions (see Subsection 3.2) and the construction of multidimensional small sets. Second, we establish the asymptotic theory of the resulting procedure: we prove a central limit theorem and its bootstrap counterpart for the sample mean, derive the asymptotic validity of the MARBB, and provide a consistent estimator of the covariance matrix for functionals evaluated along the trajectory of the Markov chain.

In this paper, we consider the Vector AutoRegressive model of order one (VAR (1)), under a low-rank assumption on the coefficient matrix.. However, our ideas may be extended to more general single/multiple-index models (see

Section 5). Recall that VAR(1) processes are widely used for multidimensional time series analysis (see [1, 17]). A natural bootstrap method in this context is the residual bootstrap. However, in high-dimensional VAR processes, the number of parameters to estimate grows significantly, so standard methods, such as mean square estimation, become ineffective. To overcome this issue, regularization methods, such as LASSO, ridge-based VAR estimation, or other shrinkage procedures are often used (see [10]). In this case, the residual bootstrap is no longer consistent due to the dependence between the estimated residuals and the penalized coefficient estimates (see [7]). Our goal is to show that the regenerative bootstrap can still be adapted successfully to high-dimensional settings.

The paper is organized as follows. Section 2 introduces the notations and key definitions. Section 3 presents the main theoretical results and the MARBB algorithm. Section 4 describes the empirical implementation of MARBB method. Section 5 discusses possible extensions of the proposed approach. All proofs and additional materials can be found in the Appendix. A simulation study performed on simulated data is provided in Section 4 in the Supplementary material.

2. Markov chains

In this section, we recall the main definitions and properties of Markov chains used throughout the paper. Then we describe the splitting technique underlying the ARBB algorithm and the regenerative properties of the Markov chains.

2.1. Regenerative Markov chains

We begin with several basic definitions and properties regarding Markov chains. For more details, we refer to [2, 9, 16].

Consider a countably generated state space (E, \mathcal{E}) . By capital letters A, B, C, \dots , we denote elements of the σ -algebra \mathcal{E} , and we call them sets or sets of states. The states in E are denoted by small letters x, y, z, \dots . Each state is represented as a vector $x = (x_1, \dots, x_d)^\top$, where $d \in \mathbb{N}$ is fixed and $(\cdot)^\top$ denotes the transposition. We write $\mathbb{I}\{\cdot\}$ for an indicator function.

Let $\mathbf{X} = (X_1, X_2, \dots, X_n, \dots) = (X_t)_{t \in \mathbb{N}}$, with $X_t = (X_{1,t}, \dots, X_{d,t})^\top$, be a homogeneous d -dimensional, positive recurrent Markov chain on (E, \mathcal{E}) with initial distribution ν , transition probability Π and stationary distribution μ . For every $t \in \mathbb{N}$ and every measurable set $B \in \mathcal{E}$, the Markov property

states that

$$P_\nu(X_{t+1} \in B | X_0, \dots, X_t) = \Pi(X_t, B) \quad \text{almost surely.} \quad (1)$$

We write \mathbb{P}_ν to denote the probability measure under which the initial state X_0 is drawn from the distribution ν , while \mathbb{P}_x represents the probability measure where $X_0 = x \in E$ with probability one. The corresponding expectations are denoted by $\mathbf{E}_\nu(\cdot)$ and $\mathbf{E}_x(\cdot)$, respectively.

Additional definitions used later in the paper (such as irreducibility, aperiodicity, atoms, and Harris recurrence) are collected in the Supplementary Material; see Definition 1.1.

Below, based on [5], we recall some standard results for atomic Markov chains to explain the relation between the regenerative structure of the process and the way of creating the blocks used to obtain bootstrap samples.

Let $\mathbf{X}_n = (X_1, \dots, X_n)$ be a realisation of an aperiodic, atomic, and Harris recurrent d -dimensional Markov chain. Such a chain possesses a regenerative structure, that is, there exist regeneration times at which the process forgets about the past and "renews" (see [16]). This regenerative structure plays a crucial role in bootstrap procedures for dependent data. Indeed, regeneration times allow the trajectory of the Markov chain to be decomposed into blocks that are i.i.d. In the atomic case, these regeneration times correspond to the instants when the Markov chain hits the atom A . Let us define the sequence of regeneration times as $(\tau_A(j))_{j \geq 1}$. Let $\tau_A = \tau_A(1) = \inf \{t \geq 1 : X_t \in A\}$ be the first time when the chain hits the regeneration set A and $\tau_A(j) = \inf \{t > \tau_A(j-1), X_t \in A\}$, for $j \geq 2$ are the consecutive visiting times to the regeneration set A . From the strong Markov property (1) and the definition of the atom (see Definition 1.1), given any initial law ν , the sample paths \mathbf{X}_n can be divided into i.i.d. blocks corresponding to the consecutive visits of the chain to the atom A . The segments of data, also called cycles, are given by

$$\mathbf{B}_j = (X_{1+\tau_A(j)}, \dots, X_{\tau_A(j+1)}), \text{ for } j \geq 1$$

and take values in the torus $\cup_{k=1}^\infty E^k$.

Let $l_n = \sum_{i=1}^n \mathbb{I}\{X_i \in A\}$ be the total number of consecutive visits to the regeneration set A up to time n . We make the convention that $\mathbf{B}_{l_n}^{(n)} = \emptyset$, whenever $\tau_A(l_n) = n$. Furthermore, we denote the length of j -th regeneration block by $l(\mathbf{B}_j) = \tau_A(j+1) - \tau_A(j)$, $j \geq 1$, and $\mathbf{E}_A(\tau_A)$ be the mean inter-renewal time.

2.2. Splitting technique for general Harris recurrent Markov chains

When the chain is not atomic, it is always possible for Harris recurrent chains, including nonstationary ones, to extend the original time series to obtain an atomic Markov chain and to approximate regenerative blocks via the notion of a small set (see [18]). The idea is known as the *splitting technique*. It extends the probabilistic structure of the chain by creating an artificial regeneration set in the general Harris recurrent case. Below, we recall the crucial definition of the small set.

Definition 2.1 (Small set). Consider an aperiodic, ψ -irreducible Markov chain on a general state space (E, \mathcal{E}) with transition probability Π . A set $S \in \mathcal{E}$ is said to be *small*, if there exist $m \in \mathbb{N}$, $\delta > 0$ and a probability measure Φ supported by S such that, for all $x \in S$, $B \in \mathcal{E}$,

$$\Pi^m(x, B) \geq \delta\Phi(B), \quad (2)$$

where Π^m is the m -th iterate of Π .

When (2) holds, we say that the chain satisfies the *minorization condition* $\mathcal{M}(m, S, \delta, \Phi)$. Thus, if the Markov chain starts in a small set S , then it will return in m -steps with nonzero probability. It is worth mentioning that such a construction of a small set is always possible because an accessible small set always exists for irreducible Markov chains (for details, see Proposition 2.6 in [2]).

Remark 2.1. Throughout the remainder of the paper, we assume that the minorization condition (2) holds with $m = 1$. In this case, the regeneration blocks constructed via the splitting technique are i.i.d., which simplifies both the bootstrap procedure and the theoretical analysis. If the condition (2) is satisfied for some $m > 1$, then the resulting blocks are 1-dependent. In that case, two approaches may be considered. The first is to use only odd or only even blocks, which restores independence (see [15]). The second is to work with the vectorized process $\{W_t, t \in \mathbb{N}\}$, where $W_t = (X_{mt}, \dots, X_{(t+1)m-1})^\top$, for which a one-step minorization condition is satisfied (see [16]). However, this latter approach increases the dimension of the process, which is exactly the difficulty we aim to avoid in high-dimensional settings.

Assumption 1. *Minorization condition $\mathcal{M}(1, S, \delta, \Phi)$ holds for some set S such that $\phi(S) > 0$.*

The aim of the splitting technique is to enlarge the state space so that the extended chain becomes atomic (for more information, see Section 2 of the Supplementary Material). To simplify the implementation of the splitting technique, we make the following

Assumption 2. *Assume that the initial distribution ν , the transition probability Π and measure Φ are absolutely continuous with respect to a σ -finite measure λ , with $\nu(dy) = f(y)\lambda(dy)$, $\Pi(x, dy) = p(x, y)\lambda(dy)$ for each $x \in E$ and $\Phi(dy) = \phi(y)dy$.*

Under Assumptions 1 and 2, the minorization condition (2) can be rewritten equivalently as

$$p(x, y) \geq \delta\phi(y) \quad \lambda(dy) - a.s. \quad (3)$$

Remark 2.2. Assumption 2 is imposed primarily for notational convenience. It allows the minorization condition to be expressed in terms of densities, which leads to a more transparent formulation of the splitting method. The original Nummelin splitting technique (see [18]) can also be stated directly in terms of probability measures, without requiring the existence of densities. However, such a formulation would result in substantially more involved notation.

Then the Nummelin splitting trick may be implemented as follows (see [5]). Let $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ be a binary random sequence such that the distribution of Y_n conditionally to $\mathbf{X}_{n+1} = (X_1, \dots, X_{n+1})$ is a tensor product of Bernoulli distribution given, for $\bar{\theta}_n = (\theta_1, \dots, \theta_n) \in \{0, 1\}^n$, by

$$P_\nu(\mathbf{Y}_n = \bar{\theta}_n | \mathbf{X}_{n+1} = \mathbf{x}_{n+1}) = \prod_{i=1}^n P_\nu(Y_i = \theta_i | X_i = x_i, X_{i+1} = x_{i+1}),$$

where $\mathbf{x}_{n+1} = (x_1, \dots, x_{n+1}) \in \mathbb{R}^{d \times (n+1)}$ is a vector of states corresponding to $\mathbf{X}_{n+1} = (X_1, \dots, X_{n+1})$. More precisely, for $i \in 1, \dots, n$ we have

$$\begin{aligned} P_\nu(Y_i = 1 | X_i = x_i, X_{i+1} = x_{i+1}) &= \delta, \quad \text{if } x_i \notin S, \\ P_\nu(Y_i = 1 | X_i = x_i, X_{i+1} = x_{i+1}) &= \frac{\delta\phi(x_{i+1})}{p(x_i, x_{i+1})}, \quad \text{if } x_i \in S. \end{aligned}$$

We denote this conditional distribution by $\mathbf{L}^{(n)}(p, S, \delta, \phi, \mathbf{x}_{n+1})$.

2.3. Curse of dimensionality

In this paragraph, we describe the main problems arising in high-dimensional settings and explain why generalising the ARBB algorithm from the 1-dimensional case to the multidimensional one is non-trivial.

In a countable state space, we are looking for a d -dimensional atom of the form $A = (a_1, \dots, a_d)^\top$. When d is large, the probability of returning exactly to the same state decreases significantly, making regenerative blocks larger and limiting the opportunities to split the process.

In a general state space, constructing the split chain may be practically impossible. This is because we need to construct the split chain, with distribution $\mathbf{L}^{(n)}(p, S, \delta, \phi, \mathbf{x}_{n+1})$. In practice, the transition probability is unknown and must be estimated. For this purpose, one can use a multivariate kernel density estimator. However, the convergence rate of this estimator in a Besov space with regularity s , under some additional regularity conditions, is typically of the order $\left(\frac{\log n}{n}\right)^{\frac{s}{(2s+d)}}$ (see [19]). This makes the algorithm quite inefficient when the transition has a low regularity and/or the dimension d is large. Moreover, choosing the small set in this case can be very difficult.

A solution to these challenges is to transform the process from a high-dimensional space to a lower-dimensional setting.

3. Main results and algorithms

In this section, we describe the main results of this paper, that is, the dimensionality reduction procedure for the VAR(1) process (see the Supplementary material for more information), and introduce the algorithm to bootstrap a multidimensional Markov chain in a general state space. Additionally, we show how to adapt the Nummelin splitting technique to high-dimensional settings. Finally, we derive consistency results for our method.

3.1. Reduced Markov Chain

We consider a d -dimensional VAR(1) process of the form

$$X_t = \mathbf{F}X_{t-1} + \epsilon_t, \quad t \in \mathbb{Z},$$

where \mathbf{F} is a $(d \times d)$ coefficient matrix and $(\epsilon_t)_{t \in \mathbb{Z}}$ is an i.i.d. sequence of error vectors with mean zero and covariance matrix $\mathbf{\Omega}$. It is well known that the VAR(1) process is second-order stationary and causal when all eigenvalues of \mathbf{F} have modulus strictly smaller than one. In what follows, we assume that the VAR(1) process satisfies the following condition.

Assumption 3. *The VAR(1) process $X_t = \mathbf{F}X_{t-1} + \epsilon_t$, satisfy the following conditions:*

1. *it is a second-order stationary and causal process. Thus, it is a positive recurrent Markov chain,*
2. *the residuals of the model ϵ_t^\top 's, for $t = 1, \dots, n$ are i.i.d., and have a non-degenerate distribution with an absolutely continuous part (thus, \mathbf{X} is a ψ -irreducible Markov chain),*
3. *the covariance matrix $\mathbf{\Omega}$ of the innovation term ϵ_t is invertible,*
4. *the matrix \mathbf{F} is a low-rank matrix with known rank k .*

Point 2 excludes purely discrete innovation distributions, which may cause difficulties in verifying irreducibility and mixing properties (see [3]). Moreover, the assumption in point 4, that the reduced dimension k is known is mainly adopted for technical convenience. In practical applications, the effective rank is typically unknown and must be estimated from the observed data. For the class of VAR(1) processes considered in this paper, a data-driven estimation procedure for k is described in Section 3.2 of the Supplementary Material.

The idea of dimensionality reduction is based on the Truncated Singular Value Decomposition (TSVD) (see [11], page 535), which we adapt to our setting. Let $\mathbf{F} = \alpha\beta^\top$, where $\alpha = \mathbf{U}\mathbf{D}_k^{1/2}$ and $\beta = \mathbf{V}\mathbf{D}_k^{1/2}$ are matrices of the size $(d \times k)$ and \mathbf{D}_k is a diagonal matrix $(k \times k)$ of singular values arranged in decreasing order. Here, \mathbf{U} and \mathbf{V} are $(d \times k)$ matrices that contain the left and right singular vectors, respectively, associated with $\mathbf{D}_k^{1/2}$. We define the reduced process as $\tilde{\mathbf{X}}_n = (\tilde{X}_1, \dots, \tilde{X}_n) = \beta^\top \mathbf{X}_n$. Note that $\tilde{\mathbf{X}}_n$ is by construction a k -dimensional Markov chain since we have $\tilde{X}_t = \beta^\top \alpha \tilde{X}_{t-1} + \beta^\top \epsilon_t$.

In practice, the matrix \mathbf{F} is unknown, but can be estimated by $\hat{\mathbf{F}}_n$ using the Ordinary Least Squares (OLS) method. Thus, we use the TSVD to decompose $\hat{\mathbf{F}}_n = \hat{\mathbf{U}}\hat{\mathbf{D}}_k\hat{\mathbf{V}}^\top$. Taking $\hat{\alpha}_n = \hat{\mathbf{U}}\hat{\mathbf{D}}_k^{1/2}$ and $\hat{\beta}_n = \hat{\mathbf{V}}\hat{\mathbf{D}}_k^{1/2}$ we have $\hat{\mathbf{F}}_n = \hat{\alpha}_n (\hat{\beta}_n)'$.

Using the estimated $(\hat{\beta}_n)'$ we construct an estimated k -dimensional process $\check{\mathbf{X}}_n = (\check{X}_1, \dots, \check{X}_n) = (\hat{\beta}_n)' \mathbf{X}_n$. Since we assumed that k is much smaller than d , the process $\check{\mathbf{X}}$ can be seen as an "estimated" reduced Markov chain.

3.2. Estimating β and reducing dimension

Due to the identifiability problem of β , we assume that for $j = 1, \dots, k$ we have standardized the estimator such that

$$\left[\left(\hat{\beta}_n \right)' \right]_{1,1} > 0 \quad \text{and} \quad \left\| \left[\left(\hat{\beta}_n \right)' \right]_j \right\| = \sqrt{\sum_{i=1}^d \left| \left[\left(\hat{\beta}_n \right)' \right]_{i,j} \right|^2} \leq 1, \quad (4)$$

where $\left[\hat{\beta}_n^\top \right]_{i,j}$ is the element (i, j) of $\left(\hat{\beta}_n \right)'$ and $\left[\left(\hat{\beta}_n \right)' \right]_j$ is a j -th column of $\left(\hat{\beta}_n \right)'$.

The following Theorem 3.1 gives a rate of convergence of $\left(\hat{\beta}_n \right)' \mathbf{X}_n$ to $\beta^\top \mathbf{X}_n$ for the Frobenius norm given by $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m a_{i,j}^2}$, for the general $(n \times m)$ -matrix \mathbf{A} , with running elements $a_{i,j}$.

Theorem 3.1. *Let \mathbf{X}_n be a realisation of the VAR(1) process satisfying Assumption 3 and having finite fourth moments, that is, $\mathbf{E} |\varepsilon_{it} \varepsilon_{jt} \varepsilon_{lt} \varepsilon_{mt}| < \infty$, for all i, j, l, m . In addition, assume that $\mathbf{E} \|\mathbf{X}_n\|_F^2 < \infty$, and the estimator $\left(\hat{\beta}_n \right)'$ is obtained by the OLS method under condition (4). Additionally, let \mathbf{F} has singular values $\sigma_1 \geq \dots \geq \sigma_{\min\{k,n\}}$. Assume that for some fixed $m, l \in \{1, 2, \dots, k\}$ with $m \leq l$, we have $\min(\sigma_{m-1}^2 - \sigma_m^2, \sigma_l^2 - \sigma_{l+1}^2) > 0$, where $\sigma_0^2 := \infty$ and $\sigma_{\text{rank}(\mathbf{F})+1}^2 := -\infty$. Then we have*

$$\mathbf{E} \left\| \beta^\top \mathbf{X}_n - \left(\hat{\beta}_n \right)' \mathbf{X}_n \right\|_F = O_P(n^{-1/2}).$$

It should be mentioned that the condition $\min(\sigma_{m-1}^2 - \sigma_m^2, \sigma_l^2 - \sigma_{l+1}^2) > 0$ in Theorem 3.1 means that there are at least two gaps between the eigenvalues. Thus, at least three eigenvalues have to be different.

It is worth noting that the proof of Theorem 3.1, which can be found in Appendix, is non-trivial due to the challenge of matrix rotation under perturbation when several eigenvalues are equal. Specifically, the matrix $\hat{\mathbf{F}}_n$ is a perturbed version of \mathbf{F} , meaning that small changes in the estimated matrix can affect both its singular values and singular vectors. Especially, the singular vectors of matrix $\hat{\mathbf{F}}_n$ may not have the same direction as singular vectors of matrix \mathbf{F} . This phenomenon is known as a *rotation* problem

(for more details, see [20]). To address this issue, we use the Davis-Kahan theorem, which was originally formulated to quantify the rotation of eigenvectors under perturbations of Hermitian matrices and extend it to estimate the rotation of singular vectors in the context of singular value decomposition ([21], page 7, Theorem 4).

The next theorem ensures that the Nummelin splitting technique can be applied to the reduced Markov chain $\tilde{\mathbf{X}}_n$.

Theorem 3.2. *Under Assumptions 1, 2, and the assumptions of Theorem 3.1, the reduced k -dimensional reduced Markov chain $\tilde{\mathbf{X}}_n$ derived from \mathbf{X}_n is ψ -irreducible and positive recurrent.*

By Theorem 3.2 and since every ψ -irreducible Markov chain possesses a small set (see [16], Theorem 5.2.2, page 105), the k -dimensional Markov chain $\tilde{\mathbf{X}}_n$ satisfies a minorization condition of the form $\mathcal{M}(1, \tilde{S}, \tilde{\delta}, \tilde{\phi})$, with some parameters $\tilde{S}, \tilde{\delta}$ and $\tilde{\phi}$.

Hence, after reducing the dimension from d to $k \ll d$, we overcome the large dimensionality problem, described in Subsection 2.3. Therefore, we can now apply the splitting technique (see Section 2.2) to split the Markov chain $\tilde{\mathbf{X}}_n$. We construct the split chain $\tilde{\mathbf{X}}^{\mathcal{M}(n)} = \left(\begin{pmatrix} \tilde{X}_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} \tilde{X}_2 \\ Y_2 \end{pmatrix}, \dots, \begin{pmatrix} \tilde{X}_n \\ Y_n \end{pmatrix} \right)$, which has a distribution $\mathbf{L}^{(n)}(p^\beta, \tilde{S}, \tilde{\delta}, \tilde{\phi}, \tilde{\mathbf{x}}_{n+1})$, with an atom $A_{\mathbf{M}}$ and the corresponding theoretical regeneration times $\tau_{A_{\mathbf{M}}}$.

Note that the reduced k -dimensional Markov chain $\tilde{\mathbf{X}}_n$ depends on the unknown parameter β . Thus, in practice, we consider the reduced k -dimensional chain $\check{\mathbf{X}}_n$, constructed with the estimator $(\hat{\beta}_n)'$. Additionally, we consider the transition probability kernel $P^\beta(x, y) = P(\beta^\top X_{t+1} \leq y | \beta^\top X_t = x)$ for $(x, y) \in (\mathbb{R}^k)^2$. We estimate the density $p^\beta(x, y)$ by Nadaraya-Watson estimator $p_n^\beta(x, y)$. For a fixed β , it is given by

$$p_n^\beta(x, y) = \frac{f_n^\beta(x, y)}{f_n^\beta(x)} \text{ where,} \tag{5}$$

$$f_n^\beta(x, y) = \frac{1}{nh_n^{2k}} \sum_{i=1}^{n-1} \mathcal{K}\left(\frac{\beta^\top X_{i+1} - y}{h_n}\right) \mathcal{K}\left(\frac{\beta^\top X_i - x}{h_n}\right),$$

$$f_n^\beta(x) = \frac{1}{nh_n^k} \sum_{i=1}^n \mathcal{K}\left(\frac{\beta^\top X_i - x}{h_n}\right).$$

Here, $\mathcal{K}(\cdot)$ is a multiplicative Gaussian kernel and $h_n \sim n^{-1/(4+k)}$ is an optimal bandwidth (see [12], page 102). Since the estimator of the transition kernel depends on the parameter β , which is in practice unknown, we also consider the estimator of the transition probability based on $(\hat{\beta}_n)' \mathbf{X}_n$, of the form

$$p_n^{\hat{\beta}}(x, y) = \frac{f_n^{\hat{\beta}}(x, y)}{f_n^{\hat{\beta}}(x)}, \quad (6)$$

where (6) is defined analogously to the estimator (5) with the difference that we replace the unknown parameter β by its estimator $(\hat{\beta}_n)'$.

The following lemma establishes the convergence rate of the estimator (6) relative to the quantity (5).

Lemma 3.3. *Under the assumptions of Theorem 3.2 the following mean square error (MSE), measured by the L^∞ loss over the small set $\tilde{S} \times \tilde{S}$, satisfies*

$$\mathbf{E} \sup_{x, y \in \tilde{S}} \left| p_n^{\hat{\beta}}(x, y) - p_n^\beta(x, y) \right| = O_P(n^{-1/2}). \quad (7)$$

This lemma ensures that we can apply the Nummelin splitting trick with the estimated transition. For this, we generate an "approximated" random binary vector $\hat{\mathbf{Y}}_n = (\hat{Y}_1, \dots, \hat{Y}_n)$ from the distribution $\mathbf{L}^{(n)}(p_n^{\hat{\beta}}, \tilde{S}, \tilde{\delta}, \tilde{\phi}, \check{\mathbf{x}}_{n+1})$.

In the following subsection, we present the MARBB for high-dimensional Markov chain with fixed parameters $(\tilde{\delta}, \tilde{S}, \tilde{\gamma})$.

3.3. MARBB algorithm

Algorithm 3.1 (Multidimensional Approximate Regenerative Block Bootstrap).

1. Pick a small set \tilde{S} satisfying condition $\mathcal{M}(1, \tilde{S}, \tilde{\delta}, \tilde{\phi})$, typically a cuboid in \mathbb{R}^k . In the special case where $\tilde{\phi}$ is the uniform distribution, an optimal data-driven choice of \tilde{S} is given later in Algorithm 4.1.

2. Based on the projected values $\check{\mathbf{X}}_n = \left(\hat{\beta}_n\right)' \mathbf{X}_n$, construct an estimator $p_n^{\hat{\beta}}(x, y)$ according to the formula (6), satisfying the following conditions (which, in the case of the uniform measure, essentially mean that the small set should be chosen in a region where there is some data):

$$p_n^{\hat{\beta}}(x, y) \geq \tilde{\delta}\tilde{\phi}(y), \text{ for } \lambda(dy)\text{-almost everywhere,} \quad (8)$$

$$p_n^{\hat{\beta}}(X_i, X_{i+1}) > 0, \text{ for } i \in \{1, \dots, n-1\}. \quad (9)$$

3. From $\mathbf{L}^{(n)} \left(p_n^{\hat{\beta}}, \tilde{S}, \tilde{\delta}, \tilde{\phi}, \check{\mathbf{x}}_{n+1}\right)$ generate a binary vector $\hat{Y}_1, \dots, \hat{Y}_n$ conditionally to the trajectory, that is,

$$\begin{aligned} P_\nu \left(\hat{Y}_i = 1 | X_i = x_i, X_{i+1} = x_{i+1} \right) &= \tilde{\delta}, \text{ if } \check{x}_i = \left(\hat{\beta}_n\right)' x_i \notin \tilde{S} \\ P_\nu \left(\hat{Y}_i = 1 | X_i = x_i, X_{i+1} = x_{i+1} \right) &= \frac{\tilde{\delta}\tilde{\phi}(X_{i+1})}{p_n^{\hat{\beta}}(X_i, X_{i+1})}, \text{ otherwise.} \end{aligned} \quad (10)$$

4. Compute $\hat{l}_n = \sum_{i=1}^n \mathbb{I}\{\check{X}_i \in \tilde{S}, \hat{Y}_i = 1\}$, the number of visits of the "estimated" reduced Markov chain to the small set $A_{\mathbf{M}} = \tilde{S} \times \{1\}$ up to the time n . Then divide the original trajectory $\mathbf{X}_n = (X_1, \dots, X_n)$ into $\hat{l}_n + 1$ approximate regeneration blocks, which are constructed every time the approximate split chain $(\check{\mathbf{X}}_n, \hat{\mathbf{Y}}_n)$ visits $\tilde{S} \times \{1\}$, that is,

$$\begin{aligned} \hat{\mathbf{B}}_0 &= \left(X_1, \dots, X_{\hat{\tau}_{A_{\mathbf{M}}}(1)} \right), \hat{\mathbf{B}}_1 = \left(X_{\hat{\tau}_{A_{\mathbf{M}}}(1)+1}, \dots, X_{\hat{\tau}_{A_{\mathbf{M}}}(2)} \right), \dots, \\ \hat{\mathbf{B}}_{\hat{l}_n-1} &= \left(X_{\hat{\tau}_{A_{\mathbf{M}}}(\hat{l}_n-1)+1}, \dots, X_{\hat{\tau}_{A_{\mathbf{M}}}(\hat{l}_n)} \right), \hat{\mathbf{B}}_{\hat{l}_n}^{(n)} = \left(X_{\hat{\tau}_{A_{\mathbf{M}}}(\hat{l}_n)+1}, \dots, X_n \right). \end{aligned}$$

where $\hat{\tau}_{A_{\mathbf{M}}}(1) = \inf\{t \geq 1, \check{X}_t \in \tilde{S}, \hat{Y}_t = 1\}$ and, for $j \geq 1$, $\hat{\tau}_{A_{\mathbf{M}}}(j+1) = \inf\{n > \hat{\tau}_{A_{\mathbf{M}}}(j), \check{X}_t \in \tilde{S}, \hat{Y}_t = 1\}$.

5. Drop the non-regenerative blocks i.e., the first block $\hat{\mathbf{B}}_0$, and also the last block $\hat{\mathbf{B}}_{\hat{l}_n}^{(n)}$ if $\hat{\tau}_{A_{\mathbf{M}}}(\hat{l}_n) < n$. Sample randomly with replacement independent blocks $\mathbf{B}_1^*, \dots, \mathbf{B}_k^*$ from the empirical distribution $\hat{\mathbf{L}}_n = (\hat{l}_n - 1)^{-1} \sum_{j=1}^{\hat{l}_n-1} \delta_{\hat{\mathbf{B}}_j}$, until $l^*(k) = \sum_{j=1}^k l(\mathbf{B}_j^*) \geq n$. Define $l_n^* = \inf\{k \geq 1, l^*(k) \geq n\}$.
6. Construct the bootstrap pseudo-trajectory $\mathbf{X}_n^* = \left(\mathbf{B}_1^*, \dots, \mathbf{B}_{l_n^*-1}^*\right)$ by concatenating the resampled blocks $\mathbf{B}_1^*, \dots, \mathbf{B}_{l_n^*}^*$.

Remark 3.1. We aim to clarify why the approximate regeneration times of the split chain $(\check{\mathbf{X}}_n, \hat{\mathbf{Y}}_n)$ also determine the approximate regeneration times of the original process \mathbf{X} . Observe that $X_t = \alpha\beta^\top X_{t-1} + \epsilon_t = g(\tilde{X}_{t-1}) + \epsilon_t$, for $t = 2, \dots, n$, where $g : \mathbb{R}^k \rightarrow \mathbb{R}^d$ is a linear map. Hence, the dynamics of \mathbf{X} are driven by the reduced process $\tilde{\mathbf{X}}$. Therefore, if the trajectory $\tilde{\mathbf{X}}_n$ is divided into i.i.d. blocks, then the resulting blocks of \mathbf{X} are also i.i.d. Thus, the proposed procedure constructs (asymptotically) i.i.d. blocks for the original high-dimensional process through regeneration of its lower-dimensional representation.

In the next subsection, we prove that the approximated regeneration times obtained by this approximated Nummelin splitting technique converge to the true regeneration times. This result is fundamental for the asymptotic validity of the proposed bootstrap procedure.

3.4. Coupling of estimated Markov chain $\tilde{\mathbf{X}}_n$ and the true one \mathbf{X}_n

In this section, we demonstrate that the probability of an incorrect split using $(\check{\mathbf{X}}_n, \hat{\mathbf{Y}}_n)$ is small and bounded by the rate of convergence of the estimator $p_n^\beta(x, y)$.

Theorem 3.4. *Assume that:*

1. *the assumptions of Theorem 3.2 hold;*
2. *the small set \tilde{S} satisfies the minorization condition $\mathcal{M}(1, \tilde{S}, \tilde{\delta}, \tilde{\phi})$ for some $\tilde{\delta}$ and measure $\tilde{\phi}$;*
3. *the transition density is estimated by $p_n^\beta(\cdot, \cdot)$ and satisfies (8), (9). Moreover, its uniform MSE measured by the L^∞ loss over the small set $\tilde{S} \times \tilde{S}$, has a rate $\kappa_n \rightarrow 0$ as $n \rightarrow \infty$, that is,*

$$\mathbf{E}_\nu \left(\sup_{(x, y) \in \tilde{S}^2} |p_n^\beta(x, y) - p^\beta(x, y)|^2 \right) = O(\kappa_n).$$

Then, the probability that we cut the trajectory at the wrong places can be bounded as follows

$$P_\nu \left(\hat{\mathbf{Y}}_n \neq \mathbf{Y}_n \mid \mathbf{X}_{n+1} = \mathbf{x}_{n+1} \right) \leq \left(\tilde{\delta} \inf_{x \in \tilde{S}} \tilde{\phi}(x) \right)^{-1} \left((\kappa_n)^{1/2} + n^{-1/2} \right).$$

Theorem 3.4 shows that the probability of splitting the "estimated" reduced Markov chain at incorrect time points is of order $O_p\left(n^{-1/2} + \kappa_n^{1/2}\right)$. This rate depends on the uniform rate of convergence of $p_n^\beta(x, y)$ to the true transition density $p^\beta(x, y)$ in the reduced space, as well as on the parameters of the minorization condition. Note that κ_n is the convergence rate of the transition density for the reduced Markov chain. Hence, it is substantially faster than the corresponding convergence rate for the original d -dimensional Markov chain.

3.5. Asymptotic properties of the MARBB

Once we split the realisation \mathbf{X}_n into regeneration blocks using Algorithm 3.1, the asymptotic properties of the MARBB are analogous to the asymptotic results of the ARBB (see [5], Section 3.5).

3.5.1. Estimators of the mean and variance

In this subsection, we introduce estimators for the mean and variance, together with their counterparts obtained using the MARBB method.

Let $f : \mathbb{R}^d \rightarrow \mathbb{R}^q$ be a μ -integrable function. We are interested in estimating

$$\mu(f) = \mathbf{E}_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}})^{-1} \mathbf{E} \sum_{i=1}^{\tau_{A_{\mathcal{M}}}} f(X_i)$$

and the covariance matrix of size $(q \times q)$

$$\Sigma_f^2 = \mathbf{E}_{A_{\mathcal{M}}}(\tau_{A_{\mathcal{M}}})^{-1} \mathbf{E}_{A_{\mathcal{M}}} \left[\left(\sum_{i=1}^{\tau_{A_{\mathcal{M}}}} (f(X_i) - \mu(f)) \right) \left(\sum_{i=1}^{\tau_{A_{\mathcal{M}}}} (f(X_i) - \mu(f)) \right)^\top \right],$$

where $\tau_{A_{\mathcal{M}}} = \inf\{t \geq 1, (X_t, Y_t) \in S \times \{1\}\}$ and $\mathbf{E}_{A_{\mathcal{M}}}(\cdot)$ denotes expectation conditional on $(X_0, Y_0) \in A_{\mathcal{M}} = S \times \{1\}$. Since the true regeneration times are unknown, we consider the estimators of asymptotic mean and asymptotic variance based on the *pseudo-regeneration times* (i.e., times i when $(\check{X}_i, \hat{Y}_i) \in \tilde{S} \times \{1\}$), eliminating the first and last approximate blocks. This leads to the

estimators

$$\widehat{\mu}_n(f) = \widehat{n}_{A_M}^{-1} \sum_{j=1}^{\widehat{l}_n-1} f(\widehat{\mathbf{B}}_j), \quad (11)$$

$$\widehat{\Sigma}_n^2(f) = \widehat{n}_{A_M}^{-1} \sum_{j=1}^{\widehat{l}_n-1} \left(f(\widehat{\mathbf{B}}_j) - \widehat{\mu}_n(f)l(\widehat{\mathbf{B}}_j) \right) \left(f(\widehat{\mathbf{B}}_j) - \widehat{\mu}_n(f)l(\widehat{\mathbf{B}}_j) \right)^\top \quad (12)$$

of $\mu(f)$ and Σ_f^2 , respectively, where $\widehat{n}_{A_M} = \widehat{\tau}_{A_M}(\widehat{l}_n) - \widehat{\tau}_{A_M}(1) = \sum_{j=1}^{\widehat{l}_n-1} l(\widehat{\mathbf{B}}_j)$ and $f(\widehat{\mathbf{B}}_j) = \sum_{i=1+\widehat{\tau}_A(j)}^{\widehat{\tau}_A(j+1)} f(X_i)$. In the case where there are no blocks or only one block (i.e., $\widehat{l}_n \leq 1$), we assume that both $\widehat{\mu}_n(f)$ and \widehat{n}_{A_M} are equal to 0.

We do not use the observations from the first block to construct these estimators because the initial block may depend on the initial distribution, which can introduce significant bias into the computed statistics. Similarly, the last block, $\widehat{\mathbf{B}}_n^{(n)}$, may not end within a small set, which could also create bias due to the discontinuity in the constructed pseudo-trajectory. Nevertheless, the effect of including the first and last block on the statistics is asymptotically negligible (for further details, see [5]).

3.5.2. Asymptotic results for mean and variance estimator

We now state the asymptotic properties of the estimators introduced in Subsection 3.5.1. To proceed, we first impose the following assumptions:

$\mathcal{H}_1(f, \rho, \nu)$: The small set \tilde{S} satisfies

$$\sup_{\beta^\top x \in \tilde{S}} \mathbf{E}_x \left(\sum_{i=1, \dots, \tau_{\tilde{S}}} \|f(X_i)\| \right)^\rho < \infty \text{ and } \mathbf{E}_\nu \left(\sum_{i=1}^{\tau_{\tilde{S}}} \|f(X_i)\| \right)^\rho < \infty.$$

$\mathcal{H}_2(\rho, \nu)$: The set \tilde{S} is such that $\sup_{\beta^\top x \in \tilde{S}} \mathbf{E}_x \left(\tau_{\tilde{S}}^\rho \right) < \infty$ and $\mathbf{E}_\nu \left(\tau_{\tilde{S}}^\rho \right) < \infty$. The conditions $\mathcal{H}_1(f, \rho, \nu)$ and $\mathcal{H}_2(\rho, \nu)$ are independent of the specific choice of the small set \tilde{S} . If these assumptions hold for some accessible small set \tilde{S} , then they also hold for all accessible small sets of the chain. Furthermore, when $\mathcal{H}_1(\mathbf{k}, \nu)$ is satisfied, then $\mathcal{H}_1(\mathbf{f}, \mathbf{k}, \nu)$ is automatically fulfilled for any measurable bounded function f (see [5]).

Below, we state the asymptotic properties of the estimators (11) and (12).

Theorem 3.5. *Let the assumptions of Theorem 3.4 hold. Moreover, assume that $\mathcal{H}_1(f, \rho, \nu)$, $\mathcal{H}_2(\rho, \nu)$ are satisfied for some $\rho \geq 4$. Then, we have*

$$\widehat{\Sigma}_n^2(f) \xrightarrow{P_\nu} \Sigma_f^2 \quad \text{and} \quad \widehat{n}_{\text{AM}}^{1/2} (\widehat{\mu}_n(f) - \mu(f)) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

as $n \rightarrow \infty$. If in addition Σ_f^2 is invertible, then

$$\widehat{n}_{\text{AM}} (\widehat{\mu}_n(f) - \mu(f))^\top \widehat{\Sigma}_n^{-1} (\widehat{\mu}_n(f) - \mu(f)) \xrightarrow{d} \chi^2(q),$$

as $n \rightarrow \infty$.

3.6. Validity of MARBB

In this subsection, we define the bootstrap versions of the mean and the variance introduced in Subsection 3.5.1, computed from reconstructed pseudo-trajectories \mathbf{X}_n^* (see Algorithm 3.1). Recall that $(\mathbf{B}_1^*, \dots, \mathbf{B}_{l_n-1}^*)$ is a pseudo-trajectory obtained by resampling the approximate regenerative blocks $\widehat{\mathbf{B}}_1, \dots, \widehat{\mathbf{B}}_{l_n-1}$. Let $n_{\text{AM}}^* = \sum_{j=1}^{l_n-1} l(\mathbf{B}_j^*)$ be a length of \mathbf{X}_n^* . Then

$$\begin{aligned} \mu_n^*(f) &= n_{\text{AM}}^{*-1} \sum_{j=1}^{l_n^*-1} f(\mathbf{B}_j^*), \\ \Sigma_n^{*2}(f) &= n_{\text{AM}}^{*-1} \sum_{j=1}^{l_n^*-1} (f(\mathbf{B}_j^*) - \mu_n^*(f)l(\mathbf{B}_j^*)) (f(\mathbf{B}_j^*) - \mu_n^*(f)l(\mathbf{B}_j^*))^\top \end{aligned} \quad (13)$$

are the bootstrap counterparts of the mean and variance estimators, respectively. Define now respectively the distributions of the standardised and studentized sample means

$$\begin{aligned} H_\nu^U(x) &= P_\nu \left(\widehat{n}_{\text{AM}}^{1/2} (\widehat{\mu}_n(f) - \mu(f)) \leq x \right), \\ H_\nu^S(x) &= P_\nu \left(\widehat{n}_{\text{AM}}^{1/2} \widehat{\Sigma}_n(f)^{-1} (\widehat{\mu}_n(f) - \mu(f)) \leq x \right). \end{aligned}$$

Using the bootstrap estimators, we derive the corresponding MARBB distributions:

$$\begin{aligned} H_{\text{MARBB}}^U(x) &= P^* \left(n_{\text{AM}}^{*1/2} (\mu_n^*(f) - \widehat{\mu}_n(f)) \leq x \mid \mathbf{X}_{n+1} \right), \\ H_{\text{MARBB}}^S(x) &= P^* \left(n_{\text{AM}}^{*1/2} \Sigma_n^*(f)^{-1} (\mu_n^*(f) - \widehat{\mu}_n(f)) \leq x \mid \mathbf{X}_{n+1} \right). \end{aligned}$$

The following Theorem establishes the asymptotic validity of the MARBB method, extending the results of ARBB to high dimensions.

Theorem 3.6. *Under the assumptions of Theorem 3.5, we have the following convergence in P_ν -probability*

$$\begin{aligned}\Delta_n^U &= \sup_{x \in \mathbb{R}} |H_{MARBB}^U(x) - H_\nu^U(x)| \rightarrow 0, \text{ as } n \rightarrow \infty, \\ \Delta_n^S &= \sup_{x \in \mathbb{R}} |H_{MARBB}^S(x) - H_\nu^S(x)| \rightarrow 0, \text{ as } n \rightarrow \infty.\end{aligned}$$

Once we have shown that the approximate regenerative times of the splitting chain converge to the true regenerative times (see Theorem 3.4), Theorem 3.6 follows as a natural generalization of the one-dimensional result (see Theorem 3.3 in [5]) to the multidimensional setting.

It is worth mentioning that the regenerative blocks obtained from the MARBB algorithm may also be used to derive the distribution of general statistics, not only the sample mean, which we considered here.

4. Empirical algorithm for choosing a small set

Since the performance of the MARBB algorithm strongly depends on the choice of a small set for the reduced chain $\tilde{\mathbf{X}}$, in this subsection, we propose a practical data-driven method to select these parameters. A standard choice of the small set is the hyper-rectangle $V_{x_0}(\epsilon) = \prod_{j=1}^k [x_0^{(j)} - \epsilon^{(j)}, x_0^{(j)} + \epsilon^{(j)}]$ for a well-chosen $x = (x_0^{(1)}, \dots, x_0^{(k)})$ and $\epsilon = (\epsilon^{(1)}, \dots, \epsilon^{(k)})$ sufficiently small. As a minorization measure, we may choose the uniform distribution on $V_{x_0}(\epsilon)$ denoted by $\mathcal{U}_{V_{x_0}(\epsilon)}$. We make the following assumption to ensure that the process visits the neighbourhood of x_0 many times.

Assumption 4. *Assume that there exists $x_0 \in \mathbb{R}^k$, such that $p^\beta(x_0, x_0) > 0$, and that $p^\beta(x, y)$ is continuous at (x_0, x_0) .*

This assumption is automatically satisfied when the residuals of the VAR(1) process have a continuous density. The following lemma ensures that the reduced process satisfies a minorization condition.

Lemma 4.1. *Let $\tilde{\mathbf{X}}_n$ be the reduced VAR(1) process satisfying the assumptions of Theorem 3.1, and assume that $p_n^\beta(x, y)$ satisfies Assumption 4. Then there exists some $\tilde{\delta} \in (0, 1)$ such that, for $(x, y) \in V_{x_0}^2$*

$$p_n^\beta(x, y) \geq \tilde{\delta} \left(\prod_{i=1}^k 2\epsilon^{(i)} \right)^{-1}. \quad (14)$$

The main challenge is to choose x_0 and ϵ so as to obtain a large number of i.i.d blocks. Let us note that if ϵ is too small, then the small set is narrow and only few visits by the chain occur. On the other hand, increasing the size of $V_{x_0}(\epsilon)$ lowers the uniform lower bound on the transition density (14), resulting in a decreasing number of pseudo-regeneration times. A strategy, following [5], is to choose the parameters x_0 , ϵ , and $\delta(\epsilon)$ which maximise the expected number of pseudo-regenerative blocks conditionally to the given trajectory of the process, that is,

$$N_n(\epsilon) = \tilde{\delta} \left(\prod_{i=1}^k 2\epsilon^{(i)} \right)^{-1} \sum_{i=1}^n \frac{\mathbb{I} \left\{ \left(\tilde{X}_i, \tilde{X}_{i+1} \right) \in V_{x_0}(\epsilon)^2 \right\}}{p^\beta(X_i, X_{i+1})}.$$

Since $p^\beta(X_i, X_{i+1})$ and $\tilde{\delta}$ are unknown, we replace them by $p_n^{\hat{\beta}}(x, y)$ and $\tilde{\delta}(\epsilon) = \inf_{(x, y) \in V_{x_0}(\epsilon)^2} p_n^{\hat{\beta}}(x, y) \prod_{i=1}^k 2\epsilon^{(i)}$ in the preceding expression. The following algorithm summarizes the empirical selection of the small set.

Algorithm 4.1 (MARBB algorithm with empirical choice of the small set).

1. Choose the mean or the median (of each row) for x_0 .
2. Construct a grid of the sizes of the hyper-rectangle, used for searching the optimal values of $\epsilon = (\epsilon^{(1)}, \dots, \epsilon^{(k)})^\top$. Let $G_\epsilon^{(i)} = \{\epsilon^{(i)}(j) : 0 \leq \epsilon^{(i)}(j) \leq J^i, j = 1, \dots, L_\epsilon\}$ for $i = 1, \dots, k$, where L_ϵ is the discretisation size and J^i are some bounds depending on the range of component i . Then consider $\mathbf{G}_\epsilon = \{(\epsilon^{(1)}, \dots, \epsilon^{(k)}), \epsilon^{(i)} \in G_\epsilon^{(i)}, \text{ for } i = 1, \dots, k\}$ which can be represented as a matrix of size $(L_\epsilon \times k)$.
3. Compute the estimator $p_n^{\hat{\beta}}(x, y)$ over the grid (see (6)) and estimate the minorization parameters $\tilde{\delta}$ using Lemma (4.1). Compute the estimated expectation number of the pseudo-regenerative blocks.
4. Select x_0^* and ϵ^* which maximise the expected number of the pseudo-regenerative blocks and keep the corresponding minorization constant $\delta_n^* = \delta_n(x_0^*, \epsilon^*)$.

5. Future directions

This section discusses the possible extension of our method to semiparametric settings. Specifically, we would like to generalize our methodology to

nonlinear models, with a focus on multiple-index models of the form:

$$X_t = g(\beta^\top X_{t-1}) + \varepsilon_t, \quad (15)$$

where $t \in \mathbb{N}$, X_t is a d -dimensional vector, β is a $(d \times k)$ matrix, $g : \mathbb{R}^k \rightarrow \mathbb{R}^d$, $k \ll d$, and ε is a d -dimensional noise vector with distribution $N_d(0, \mathbf{\Omega})$.

For this model, our objective remains the same, that is, to estimate the unknown function g and the matrix β and, then to use β to achieve dimensionality reduction and obtain a k -dimensional Markov chain, $\tilde{\mathbf{X}} = \beta^\top X_t$ satisfying

$$\beta^\top X_t = \beta^\top g(\beta^\top X_{t-1}) + \beta^\top \varepsilon_t.$$

A particular case of (15) is the single-index model, which has received considerable attention in the literature (see for example, [13]). The single-index model is based on β , which is a vector of length d , and $g : \mathbb{R} \rightarrow \mathbb{R}^d$. In this setting, the vector β enables the construction of a 1-dimensional Markov chain.

In an ongoing work, we investigate a slight modification of the algorithm considered here and prove the validity of the regenerative block bootstrap for both single-index and multiple-index models. The underlying ideas and the structure of the algorithm remain the same but the proofs are different.

Appendix A. Proofs of the main results

Proof of Theorem 3.1. Throughout the proof, let $\|\mathbf{F}\|_{op} = \sigma_{\max}(\mathbf{F})$, where σ denotes the largest singular value of \mathbf{F} , and $\|\cdot\|_{op}$ is the operator norm. Using the property of the Frobenius norm, we bound the MSE

$$\mathbf{E} \left\| \beta^\top \mathbf{X}_n - (\hat{\beta}_n)' \mathbf{X}_n \right\|_F \leq \mathbf{E} \left[\left\| \beta^\top - (\hat{\beta}_n)' \right\|_F \|\mathbf{X}_n\|_F \right].$$

By the Cauchy-Schwarz inequality, we obtain that

$$\mathbf{E} \left[\left\| \beta^\top - (\hat{\beta}_n)' \right\|_F \|\mathbf{X}_n\|_F \right]^2 \leq \mathbf{E} \left\| \beta^\top - (\hat{\beta}_n)' \right\|_F^2 \mathbf{E} \|\mathbf{X}_n\|_F^2.$$

Under the assumptions that $\mathbf{E} \|\mathbf{X}_n\|_F^2 < \infty$, it is enough to show that $\mathbf{E} \left\| \beta^\top - (\hat{\beta}_n)' \right\|_F^2 < \infty$. To achieve this, let us recall that $(\hat{\beta}_n)' = \hat{\mathbf{V}} \hat{\mathbf{D}}_k^{1/2}$,

$\hat{\alpha}_n = \hat{\mathbf{U}}\hat{\mathbf{D}}_k^{1/2}$ and $\alpha = \mathbf{U}\mathbf{D}^{1/2}$, $\beta^\top = \mathbf{V}\mathbf{D}_k^{1/2}$ (see Subsection 3.1). By the Singular Value Decomposition, there exist orthogonal matrices $\hat{\mathbf{O}}_1, \hat{\mathbf{O}}_2 \in \mathbb{R}^{d \times d}$ (see [21], page 10), such that $\hat{\mathbf{O}}_1^\top \hat{\mathbf{V}}^\top \mathbf{V} \hat{\mathbf{O}}_2 = \text{diag}(\cos \theta_1, \dots, \cos \theta_d)$, where $\theta_1, \dots, \theta_d$ are the principal angles between the column space of \mathbf{V} and $\hat{\mathbf{V}}$. Let $\hat{\mathbf{O}} = \hat{\mathbf{O}}_1 \hat{\mathbf{O}}_2^\top$ and from the property of orthogonality, we can write

$$\left(\hat{\beta}_n\right)' = \hat{\mathbf{V}}\hat{\mathbf{O}}_1^\top \hat{\mathbf{O}}_1 \hat{\mathbf{D}}_k^{1/2} = \hat{\mathbf{V}}_1 \hat{\mathbf{D}}_1 \text{ and } \beta^\top = \mathbf{V}\hat{\mathbf{O}}_2^\top \hat{\mathbf{O}}_2 \mathbf{D}_k^{1/2} = \hat{\mathbf{V}}_2 \hat{\mathbf{D}}_2,$$

where $\hat{\mathbf{V}}_1 = \hat{\mathbf{V}}\hat{\mathbf{O}}_1^\top$, $\hat{\mathbf{D}}_1 = \hat{\mathbf{O}}_1 \hat{\mathbf{D}}_k^{1/2}$, $\hat{\mathbf{V}}_2 = \mathbf{V}\hat{\mathbf{O}}_2^\top$ and $\hat{\mathbf{D}}_2 = \hat{\mathbf{O}}_2 \mathbf{D}_k^{1/2}$. Then, we have that

$$\begin{aligned} \left\| \beta^\top - \left(\hat{\beta}_n\right)' \right\|_F &= \left\| \hat{\mathbf{V}}_1 \hat{\mathbf{D}}_1 - \hat{\mathbf{V}}_2 \hat{\mathbf{D}}_2 \right\|_F = \left\| \hat{\mathbf{V}}_1 \hat{\mathbf{O}} \hat{\mathbf{O}}^\top \hat{\mathbf{D}}_1 - \hat{\mathbf{V}}_2 \hat{\mathbf{D}}_2 \right\|_F \\ &\leq \left\| \left(\hat{\mathbf{V}}_1 \hat{\mathbf{O}} - \hat{\mathbf{V}}_2 \right) \hat{\mathbf{O}}^\top \hat{\mathbf{D}}_1 + \hat{\mathbf{V}}_2 \hat{\mathbf{O}}^\top \hat{\mathbf{D}}_1 - \hat{\mathbf{V}}_2 \hat{\mathbf{D}}_2 \right\|_F \\ &\leq \left\| \hat{\mathbf{V}}_1 \hat{\mathbf{O}} - \hat{\mathbf{V}}_2 \right\|_F \left\| \hat{\mathbf{O}}^\top \right\|_F \left\| \hat{\mathbf{D}}_1 \right\|_F + \left\| \hat{\mathbf{V}}_2 \right\|_F \left\| \hat{\mathbf{O}}^\top \hat{\mathbf{D}}_1 - \hat{\mathbf{D}}_2 \right\|_F. \end{aligned} \quad (\text{A.1})$$

Let us note, that $\left\| \hat{\mathbf{O}}^\top \right\|_F = 1$ and $\left\| \hat{\mathbf{V}}_2 \right\|_F \leq 1$. For the first term of (A.1) we use Theorem 4 from [21] and the assumption on singular values assumed in Theorem 3.1. This gives us the result for $c := l - m + 1$, where $m, l \in \{1, 2, \dots, k\}$ are fixed such that $m \leq l$ and $\min(\sigma_{m+1}^2 - \sigma_m^2, \sigma_l^2 - \sigma_{l+1}^2) > 0$ that

$$\left\| \hat{\mathbf{V}}\hat{\mathbf{O}} - \hat{\mathbf{V}}_2 \right\|_F \quad (\text{A.2})$$

$$\leq \frac{2^{3/2} \left(2\sigma_1 + \left\| \mathbf{F} - \hat{\mathbf{F}}_n \right\|_{op} \right) \min \left(d^{1/2} \left\| \mathbf{F} - \hat{\mathbf{F}}_n \right\|_{op}, \left\| \mathbf{F} - \hat{\mathbf{F}}_n \right\|_F \right)}{\min(\sigma_{m-1}^2 - \sigma_m^2, \sigma_l^2 - \sigma_{l+2}^2)}. \quad (\text{A.3})$$

From the CLT for the $\hat{\mathbf{F}}_n$ (see, Proposition 2, point 4), we have that $\left\| \mathbf{F} - \hat{\mathbf{F}}_n \right\|_{op} = O_P(n^{-1/2})$ and $\left\| \mathbf{F} - \hat{\mathbf{F}}_n \right\|_F = O_P(n^{-1/2})$. Under the assumptions on singular values given in Theorem 3.1 there exists some positive constant C_4 , such that $\min(\sigma_{m+1}^2 - \sigma_m^2, \sigma_l^2 - \sigma_{l+1}^2) = C_4 > 0$. Thus,

$$\left\| \hat{\mathbf{V}}\hat{\mathbf{O}} - \hat{\mathbf{V}}_2 \right\|_F = \frac{C_1(C_2 + O_P(n^{-1/2}))C_3 O_P(n^{-1/2})}{C_4} = O_P(n^{-1/2}),$$

where $C_i, i = 1, \dots, 4$ are some positive constants. Furthermore, using the fact that the Frobenius norm of any orthogonal matrix is equal to 1 and using the definition of the Frobenius norm, the TSVD decomposition of the matrix $\hat{\mathbf{F}}_n = \hat{\mathbf{U}}\hat{\mathbf{D}}_k\hat{\mathbf{V}}^\top$ satisfies $\|\hat{\mathbf{F}}_n\|_F^2 = \|\hat{\mathbf{D}}_k\|_F^2$. Thus, the second term of (A.1) is equal to

$$\begin{aligned} \|\hat{\mathbf{D}}_1\|_F &= \|\hat{\mathbf{O}}_1\|_F \|\hat{\mathbf{D}}_k^{1/2}\|_F = \|\hat{\mathbf{F}}_n\|_F \\ &= \|\hat{\mathbf{F}}_n - \mathbf{F} + \mathbf{F}\|_F \leq \|\hat{\mathbf{F}}_n - \mathbf{F}\|_F + \|\mathbf{F}\|_F \leq O_P(n^{-1/2}) + C_5, \end{aligned}$$

where C_5 is some positive constant. Since the process \mathbf{X}_n is second-order stationary and the matrix \mathbf{F} has all singular values smaller than 1, then $\|\mathbf{F}\|_F \leq C_5 < \infty$. Finally, the last term of (A.1) is equal to

$$\begin{aligned} \|\hat{\mathbf{O}}^\top \hat{\mathbf{D}}_1 - \hat{\mathbf{D}}_2\|_F &= \|\hat{\mathbf{O}}_2 \hat{\mathbf{O}}_1^\top \hat{\mathbf{O}}_1 \hat{\mathbf{D}}_k^{1/2} - \hat{\mathbf{O}}_2 \hat{\mathbf{D}}_k\|_F \\ &= \|\hat{\mathbf{O}}_2\|_F \|\hat{\mathbf{D}}_k^{1/2} - \hat{\mathbf{D}}_k\|_F = \|\hat{\mathbf{F}}_n - \mathbf{F}\|_F = O_P(n^{-1/2}). \end{aligned}$$

Therefore, we obtain $\mathbf{E} \left\| \beta^\top - \left(\hat{\beta}_n \right)' \right\|_F = O_P(n^{-1/2})$. \square

Proof of Theorem 3.2. We have to prove that the k -dimensional Markov chain $\tilde{\mathbf{X}} = \beta^\top X_t = \beta^\top \alpha \beta^\top X_{t-1} + \beta^\top \epsilon_t$ is ψ -irreducible and positive recurrent. Notice that $\tilde{\mathbf{X}}$ is now a VAR(1) model of the form $\tilde{\mathbf{X}} = \beta^\top \alpha \tilde{\mathbf{X}}_{t-1} + \beta^\top \epsilon_t$. To prove irreducibility and positive recurrence, we essentially have to prove that the residuals here are non-degenerate and that the eigenvalues of $\beta^\top \alpha$ are smaller than 1.

Firstly, recall that under Assumption 3 the ϵ_t has a continuous distribution f_ϵ , satisfying $f_\epsilon(y) > 0$, where $y \in \mathbb{R}^d$. Thus, the $\beta^\top \epsilon_t$ also has a non-degenerate distribution with a continuous part, else it would mean that the components are linked, and hence the covariance matrix is not invertible.

Secondly, we demonstrate that if the characteristic polynomial associated to the matrix $\mathbf{F} = \alpha \beta^\top$ has roots outside the unit circle, this is equivalent to the fact that the characteristic polynomial of the matrix $\mathbf{G} = \beta^\top \alpha$ also has its roots outside the unit circle. Let us note that if the polynomial $\det(\mathbf{I}_d - \mathbf{F}\lambda)$ has roots outside the unit circle, then equivalently $\frac{1}{\lambda} \in \sigma(\mathbf{F})$, where $\sigma(\cdot)$ is a spectrum of the matrix.

Therefore, we need to show that $\frac{1}{\lambda} \in \sigma(\mathbf{F}) \Leftrightarrow \frac{1}{\lambda} \in \sigma(\mathbf{G})$.

(\Rightarrow) There exists vector $v \neq 0$, such that $\alpha\beta^\top v = \frac{1}{\lambda}v$. Consequently, we have $\beta^\top \alpha\beta^\top v = \frac{1}{\lambda}\beta^\top v$. Note that $\tilde{v} = \beta^\top v$ is a nonzero vector. If $\beta^\top v = 0$, then the previous equation would become $0 = \frac{1}{\lambda}\beta^\top v$, which implies $v = 0$ or $\beta^\top = 0$. Both cases lead to a contradiction. Then, we have $\mathbf{G}\tilde{v} = \frac{1}{\lambda}\tilde{v}$. Thus, $\frac{1}{\lambda} \in \sigma(\mathbf{G})$. Now let's prove the reverse implication. There exists vector $w \neq 0$, such that $\beta^\top \alpha w = \frac{1}{\lambda}w$. Consequently, $\alpha\beta^\top \alpha w = \frac{1}{\lambda}\alpha w$. Let $\tilde{w} = \alpha w$ be a nonzero vector. Then, we get $\mathbf{G}\tilde{w} = \frac{1}{\lambda}\tilde{w}$. Finally, we obtain that $\det(\mathbf{I}_d - \lambda\mathbf{F}) = 0 \Leftrightarrow \det(\mathbf{I}_d - \lambda\mathbf{G}) = 0$. Therefore, the process $\tilde{\mathbf{X}}_t$ is irreducible, positive recurrent (and thus causal) and second-order stationary. \square

Proof of lemma 3.3. We now derive L_2 uniform bound estimation for $p_n^{\hat{\beta}}(x, y) - p_n^\beta(x, y)$ over the small set $\tilde{S} \times \tilde{S}$ using the mean value theorem. We have by the Cauchy-Schwarz inequality

$$\begin{aligned} \mathbf{E} \sup_{x, y \in \tilde{S}} \left| p_n^{\hat{\beta}}(x, y) - p_n^\beta(x, y) \right| &\leq \mathbf{E} \sup_{x, y \in \tilde{S}} \left\| \frac{\partial p_n^\beta}{\partial \beta} \Big|_{\beta=\tilde{\beta}} \right\|_F \|\hat{\beta} - \beta\|_F \\ &\leq \left(\mathbf{E} \left(\sup_{x, y \in \tilde{S}} \left\| \frac{\partial p_n^\beta}{\partial \beta} \Big|_{\beta=\tilde{\beta}} \right\|_F \right)^2 \right)^{1/2} \left(\mathbf{E} \left(\|\hat{\beta} - \beta\|_F \right)^2 \right)^{1/2}, \end{aligned}$$

where $\tilde{\beta} \in (\beta, \hat{\beta}_n)$, what can be understood as $\tilde{\beta}_{i,j} \in (\beta_{i,j}, \hat{\beta}_{i,j})$ for $i = 1, \dots, d$, $j = 1, \dots, k$ and where $(\cdot)_{i,j}$ denotes the element from the i -th row and j -th column. Since we assumed in (4) that $\left\| \left[\hat{\beta}_n^\top \right]_j \right\| \leq 1$, for $j = 1, \dots, k$, then

$$\|\tilde{\beta}\|_F^2 \leq \sum_{j=1}^k \left\| \left[\tilde{\beta} \right]_j \right\|_2^2 \leq \sum_{j=1}^k \left\| \left[\hat{\beta} \right]_j \right\|_2^2 \leq k,$$

where k is a fixed number. Moreover, since $(x, y) \in \tilde{S}^2$, which is a compact set and the kernel function is a Gaussian kernel, which is an absolutely continuous function, then the derivative $\frac{\partial p_n}{\partial \beta}$ is bounded by some constant.

From Theorem 3.1, the second term $\mathbf{E} \left\| \hat{\beta} - \beta \right\|_F^2$ is $O_P(n^{-1})$. Finally, we have $\mathbf{E} \left(\sup_{x, y \in \tilde{S}} \left| p_n^{\hat{\beta}}(x, y) - p_n^\beta(x, y) \right| \right) = O_P(n^{-1/2})$. \square

Proof. Proof of Theorem 3.4

The proof is similar to the proof of Theorem 3.1 in [6], although we use

$p_n^{\hat{\beta}}(x, y)$ instead of $p_n^{\beta}(x, y)$. By $\tau_S = \tau_S(1) = \inf\{t \geq 1, X_t \in S\}$ and $\tau_S(j) = \{t > \tau_S(j-1), X_t \in S\}$ we denote the first and subsequent visiting times to the small set S . Let us take $\mathbf{X}_n = (X_1, \dots, X_{\tau_S}, \dots, X_{\tau_S(L_n)}, \dots, X_n)$, where $L_n = \sum_{i=1}^n \mathbb{I}\{X_i \in S\}$ is the number of visits to the small set by the chain. Let us write the distribution of (\mathbf{Y}_i) when the Markov chain is in small set as follows

$$\begin{aligned} Y_{\tau_S(k)} &\sim \text{Ber}(\delta\phi(X_{\tau_S(k)+1}) / p^{\beta}(X_{\tau_S(k)}, X_{\tau_S(k)+1})), \\ \widehat{Y}_{\tau_S(k)} &\sim \text{Ber}(\delta\phi(X_{\tau_S(k)+1}) / p_n^{\hat{\beta}}(X_{\tau_S(k)}, X_{\tau_S(k)+1})), \text{ for } i \in [1, n] \end{aligned}$$

Let us consider two cases. Firstly, if $p^{\beta}(X_{\tau_S(k)}, X_{\tau_S(k)+1}) \leq p_n^{\hat{\beta}}(X_{\tau_S(k)}, X_{\tau_S(k)+1})$, then we have

$$\begin{aligned} P(\widehat{Y}_{\tau_S(k)} = 0, Y_{\tau_S(k)} = 1 \mid \mathbf{X}_{n+1}) &= \frac{\delta\phi(X_{\tau_S(k)+1})}{p^{\beta}(X_{\tau_S(k)}, X_{\tau_S(k)+1})} - \frac{\delta\phi(X_{\tau_S(k)+1})}{p_n^{\hat{\beta}}(X_{\tau_S(k)}, X_{\tau_S(k)+1})}, \\ P(\widehat{Y}_{\tau_S(k)} = 1, Y_{\tau_S(k)} = 0 \mid \mathbf{X}_{n+1}) &= 0, \text{ for } k \in \{1, \dots, L_n\}. \end{aligned}$$

In the second case, if $p^{\beta}(X_{\tau_S(k)}, X_{\tau_S(k)+1}) \geq p_n^{\hat{\beta}}(X_{\tau_S(k)}, X_{\tau_S(k)+1})$, we have

$$\begin{aligned} P(\widehat{Y}_{\tau_S(k)} = 1, Y_{\tau_S(k)} = 0 \mid \mathbf{X}_{n+1}) &= \frac{\delta\phi(X_{\tau_S(k)+1})}{p_n^{\hat{\beta}}(X_{\tau_S(k)}, X_{\tau_S(k)+1})} - \frac{\delta\phi(X_{\tau_S(k)+1})}{p^{\beta}(X_{\tau_S(k)}, X_{\tau_S(k)+1})}, \\ P(\widehat{Y}_{\tau_S(k)} = 0, Y_{\tau_S(k)} = 1 \mid \mathbf{X}_{n+1}) &= 0, \text{ for } k \in \{1, \dots, L_n\}. \end{aligned}$$

Following Nummelin's construction, if the chain is not in the small set, then take $i \sim \text{Ber}(\delta)$. Therefore, by combining the two cases considered above, we can compute the probability of missing the true generation times as follows

$$\begin{aligned} &P(\widehat{Y}_{\tau_S(k)} \neq Y_{\tau_S(k)} \mid \mathbf{X}_{n+1}) \\ &= \left| \frac{\delta\phi(X_{\tau_S(k)+1})}{p^{\beta}(X_{\tau_S(k)}, X_{\tau_S(k)+1})} - \frac{\delta\phi(X_{\tau_S(k)+1})}{p_n^{\hat{\beta}}(X_{\tau_S(k)}, X_{\tau_S(k)+1})} \right| \text{ a.s., for } 1 \leq k \leq L_n. \end{aligned}$$

Let us note that if we are in the next step $k+1$ in the small set, then we can obtain the following approximation

$$\begin{aligned} &\left| \frac{\delta\phi(X_{k+1})}{p^{\beta}(X_k, X_{k+1})} - \frac{\delta\phi(X_{k+1})}{p_n^{\hat{\beta}}(X_k, X_{k+1})} \right| \\ &\leq \mathbb{I}\{X_{k+1} \in S\} \left| \frac{p_n^{\hat{\beta}}(X_k, X_{k+1}) - p^{\beta}(X_k, X_{k+1})}{p^{\beta}(X_k, X_{k+1})} \right|, \end{aligned}$$

since $\frac{\delta\phi(X_{k+1})}{p_n^{\hat{\beta}}(X_k, X_{k+1})} > 1$. Next, from Lemma 3.3 and the condition 3 from the Assumption 3 of Theorem 3.4 we obtain

$$\begin{aligned} & \sup_{(X_k, X_{k+1}) \in \tilde{S}^2} \left| p_n^{\hat{\beta}}(X_k, X_{k+1}) - p^\beta(X_k, X_{k+1}) \right| \\ & \leq \sup_{(X_k, X_{k+1}) \in \tilde{S}^2} \left| p_n^{\hat{\beta}}(X_k, X_{k+1}) - p_n^\beta(X_k, X_{k+1}) \right| \\ & \quad + \sup_{(X_k, X_{k+1}) \in \tilde{S}^2} \left| p_n^\beta(X_k, X_{k+1}) - p^\beta(X_k, X_{k+1}) \right| \leq (\kappa_n)^{1/2} + n^{-1/2}. \end{aligned}$$

Moreover, under the condition (8), we obtain the desired bound, which ends the proof. \square

Proof. Proof of Lemma 4.1

Since $p^\beta(x_0, x_0) > 0$ for some $x_0 \in \mathbb{R}^k$ (Assumption 4), the transition density function is positive in a neighbourhood $V_{x_0}(\epsilon) = [x_0 - \epsilon, x_0 + \epsilon]^k$, where $\epsilon > 0$ is small. Thus, there exists some $\delta \in (0, 1)$, such that $\inf_{(x,y) \in V_{x_0}(\epsilon)^2} p^\beta(x, y) \geq \delta$. Consequently, for fixed small set $V_{x_0}(\epsilon)$ we get $p^\beta(x, y) \geq \delta$. Therefore, $p^\beta(x, y) \geq \tilde{\delta} \left(\prod_{i=1}^k 2\epsilon^{(i)} \right)^{-1} \mathbb{I}\{(x, y) \in V_{x_0}(\epsilon)^2\}$, where, for a fixed ϵ and x_0 , we put $\tilde{\delta} = \delta \prod_{i=1}^k 2\epsilon^{(i)}$ and choose the minorization condition $\mathcal{M} \left(1, V_{x_0}(\epsilon), \tilde{\delta}, \mathcal{U}_{V_{x_0}(\epsilon)} \right)$. \square

Acknowledgments

The authors are grateful to an anonymous referee and the associate editor for their insightful comments, which greatly improved the quality of the paper.

Anna E. Dudek acknowledges the financial support provided by the French government under the France 2030 investment plan, through the Initiative d'Excellence d'Aix-Marseille Université – AMIDEX (grant no. AMX-22-CEX-057).

This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01). P. Bertail and K. Łukasiewicz (Marek) would like to thank the Labex for their financial support.

Research of Anna.E.Dudek and K. Łukasiewicz (Marek) was partially supported by the AGH University of Krakow under grant no. 16.16.420.054, funded by the Polish Ministry of Science and Higher Education.

References

- [1] Alquier, P., Marie, N., 2019. Matrix Factorization for Multivariate Time Series Analysis. *Electronic Journal of Statistics* 13, 4346–4366.
- [2] Andrews, D.W.K., 1984a. Non-strong mixing autoregressive processes. volume 21.
- [3] Andrews, D.W.K., 1984b. Non-strong mixing autoregressive processes. *Journal of Applied Probability* 21, 930–934.
- [4] Athreya, K., Fuh, C., 1992. Bootstrapping markov chains: countable case. *Journal of Statistical Planning and Inference* 33, 311–331.
- [5] Bertail, P., Cléménçon, S., 2006. Regeneration-based statistics for harris recurrent markov chains, in: *Dependence in Probability and Statistics*. Springer, New York, NY. volume 187 of *Lecture Notes in Statistics*, pp. 3–54.
- [6] Bertail, P., Portier, F., 2019. Rademacher complexity for Markov chains: Applications to kernel smoothing and Metropolis–Hastings. *Bernoulli* 25, 3912–3938.
- [7] Chatterjee, A., Lahiri, S.N., 2011. Bootstrapping lasso estimators. *Journal of the American Statistical Association* 106, 608–625.
- [8] Datta, S., McCormick, W.P., 1993. Regeneration-based bootstrap for Markov chains. *Can. J. Stat.* 21, 181–193.
- [9] Douc, R., Moulines, E., Priouret, P., Soulier, P., 2018. *Markov Chains: Basic Definitions*. Springer International Publishing, Cham.
- [10] Han, F., Lu, H., Liu, H., 2015. A direct estimation of high dimensional stationary vector autoregressions. *Journal of Machine Learning Research* 16, 3115–3150.
- [11] Hansen, P.C., 1987. The truncated svd as a method for regularization. *BIT Numerical Mathematics* 27, 534–553.
- [12] Härdle, W.K., Müller, M., Sperlich, S., Werwatz, A., 2006. *Nonparametric and Semiparametric Models*.

- [13] Ichimura, H., 1993. Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of Econometrics* 58, 71–120.
- [14] Lahiri, S.N., 2003. *Resampling Methods for Dependent Data*. Springer, New York, NY, USA.
- [15] Levental, S., 1988. Uniform limit theorems for Harris recurrent Markov chains. *Probab. Theory Related Fields* 80, 101–118.
- [16] Meyn, S.P., Tweedie, R.L., 2012. *Markov chains and stochastic stability*. Springer Science & Business Media.
- [17] Negahban, S., Wainwright, M.J., 2011. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics* 39, 1069–1097.
- [18] Nummelin, E., 1978. A splitting technique for harris recurrent markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 43, 309–318.
- [19] Sart, M., 2014. Estimation of the transition density of a markov chain, *Gauthier-Villars*. pp. 1028–1068.
- [20] Stewart, G., 2001. *Stochastic perturbation theory*. *SIAM Review* 32.
- [21] Yu, Y., Wang, T., Samworth, R.J., 2014. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika* 102, 315–323.