Stochastic Modeling and Applications
Vol.21, No. 1(June, 2017), 31-42

MUK PUBLICATIONS
Open Access Publisher

FULL INTERACTION PARTITION ESTIMATION IN STOCHASTIC PROCESSES

M. FERNÁNDEZ, JESÚS E. GARCÍA, V.A. GONZÁLEZ-LÓPEZ, AND M.L. LANFREDI VIOLA

Abstract. Consider X_t as being a multivariate Markov process on a finite alphabet A. The marginal processes of X_t interact depending on the past states of X_t. We introduce in this paper a consistent strategy to find the groups of independent marginal processes, conditioned to parts of the state space, in which the strings in the same part, of the state space, share the same transition probability to a next symbol on the alphabet A. The groups of conditionally independent marginal processes will be the *structure of interaction of* X_t. The theoretical results introduced in this paper ensure, through the Bayesian Information Criterion, that for a sample size large enough the estimation strategy allows to recover the true conditional structure of interaction of X_t. Moreover, by construction, the strategy is also capable to catch mutual independence between the marginal processes of X_t.

1. Introduction

The study of dependence between sources is a topic investigated by several authors in a broad range of contexts, for example see [15] and [21]. This knowledge is useful to guide decisions and define steps that allow to create corrective settings. It is necessary to highlight some aspects, the complexity of the models to address such situations and the variety of types of dependence, to represent the relationship between sources. Just as it is complex, the representation of a source (or a process), see for instance the large range of models in this line [17], [20], [14], [1], [12] and [9]. In this paper we present a strategy for identifying groups of coordinates from a multivariate process, which are independent, when the joint process is conditioned to certain strings of observations, also, the space of strings that allow this property can be identified. Moreover, this strategy can be used before being defined a model for the process, facilitating the implementation of dependence structures as set forth in [15] and [21]. In [7] this problem was studied and for an appropriated contextualization was chosen a low cost family of Markov models, as a natural environment. This is the family of partition Markov models, see [5]. As showed in [7], partition Markov models are convenient to develop an estimation strategy of the interaction structure. The strategy is based on the Bayesian Information Criterion (see [18]) which allows a consistent estimation of the interaction structure and also a consistent estimation of partition Markov

2010 *Mathematics Subject Classification.* 62M05; 62Hxx.
Key words and phrases. Multivariate Markov chains; Independence; Partition Markov models; Bayesian Information Criterion.

models. However [7] is only concerned with pairwise independence and cannot detect situations of mutually independence, already registered on the literature. In this paper we introduce a new estimation strategy which provides a full solution of the problem to find independent groups of coordinates, being preferable than the strategy proposed in [7].

The partition Markov models are being investigated for different purposes, see for example [4]. For instance, the strategy used in [8] is to combine through a copula (see [16] and [13]) the optimal partitions of the marginal state spaces, coming from the marginal processes and the optimal partition of the joint state space coming from the multivariate process. The goal in [8] is to define a natural correction for the estimator of the transition probabilities of a multivariate Markov process, it is helpful when the sample size is not large enough to ensure a good quality of the estimator derived only using the Bayesian Information Criterion in the multivariate process. This procedure shows desirable theoretical properties that will be essential to increase the predictive ability of the estimation. In [8] the procedure was applied to multivariate Brazilian financial data in order to show how this new estimator allows to work with a longer past (order of the process), than the one allowed by the usual procedure of estimation, when the data size is not large enough to ensure reliable results.

Here, we describe the topics covered in this paper. In section 2 we introduce the concept of Markov chain with partition $\mathcal{L}$, which is a partition of the state space defined through a stochastic equivalence between strings of the state space. Also we introduce the concept *structure of interaction*, defined over the previous partition of the state space. In section 3 we describe the consistent model selection procedure for choosing the structure of interaction, that is based on the Bayesian Information criterion. In section 4 we obtain groups of series which are conditionally independent. In the case, we use 4 series which are representative in the Brazilian market: VALE (VALE5), ITAUUNIBANCO (ITUB4), AMBEV S/A (ABEV3) and BRADESCO (BBDC4), period: January 02, 2006-July 31, 2015. We conclude this paper with a discussion in section 5. The auxiliary result's proof is in the Appendix 7.

2. Preliminaries

Let $X_t = (X(1)_t, ..., X(k)_t)$ be the state of the set of k processes at time t, where $X(i)_t \in B$ and it is the state of the i-process at time t for $i = 1, ..., k$; $X_t \in A = B^k$, B is a finite and enumerable set of elements. We will assume that X_t is an order M, Markov chain, with $M < \infty$. The concatenation of elements from A, $a_m a_{m+1} \dots a_n$ where $a_i \in A$, $m \leq i \leq n$, is denoted by a_m^n. Given the state space of strings of size M which is $\mathcal{S} = A^M$, for each $s \in \mathcal{S}$, $a \in A$, we denote the conditional probability of the k-variate process by $P(a|s) = \text{Prob}(X_t = a|X_{t-M}^{t-1} = s)$. The next definition introduces a notion of equivalence $\sim$ between strings from the state space $\mathcal{S}$, induced by the joint probability of the process. Also, we introduce the notion of partition $\mathcal{L}$ corresponding to $\sim$.

Definition 2.1. Let X_t be an order M, k-variate Markov chain, with alphabet $A = B^k$ and state space $\mathcal{S} = A^M, M < \infty$, (i) for each $s, r \in S$, $s \sim r$ if

$P(a|s) = P(a|r)\ \forall a \in A$; (ii) X_t has partition $\mathcal{L}$ if this partition is the one defined by the equivalence relationship $\sim$ introduced by i.

Then, the set of parameters for a Markov chain over the alphabet A with partition $\mathcal{L}$ is given by the set of conditional probabilities $\{P(a|L) : a \in A, L \in \mathcal{L}\}$, where $P(a|L) = P(a|s)$, for any $s \in L$. Thus, the total number of parameters will be $|\mathcal{L}|(|A| - 1)$. The model integrated by such parameters is called Partition Markov Model (PMM). Now we show, an introduction of the interaction structure under investigation. This structure was explored in [7] under the perspective of other strategy of estimation considering the PMM models and in [19] considering Variable Length Markov Chain models. The target is to obtain for each part of the partition of the state space, a partition of the set of coordinates of the multivariate process. The partition of coordinates will split the set of coordinates in independent subsets, called in this paper as *Structure of Interaction.* For a collection of coordinates $u = \{u_1, ...u_l\} \subseteq \{1, 2, ..., k\}$ and $a = (a_1, ..., a_k) \in A$, define, $a^u = (a_{u_1}, ..., a_{u_l})$ that is a vector composed only by the u coordinates of a. For each part $L \in \mathcal{L}$ define the transition probability from L to a vector a^u, of u coordinates, $P(a^u|L) = \text{Prob}\Big((X(u_1)_t, ..., X(u_l)_t) = a^u|X_{t-M}^{t-1} = s\Big),\ \forall s \in L$. The previous equality holds because L is a part of the partition $\mathcal{L}$ following Definition 2.1-i., i.e. all the strings in L share the same transition probability. Given $L \in \mathcal{L}$, consider $\mathcal{I}_L$ a partition of $\{1, \ldots, k\}$ such that

$$\forall a \in A,\ P(a|L) = \prod_{C \in \mathcal{I}_L} P(a^C|L). \tag{2.1}$$

We define the optimal partition which satisfies the previous relation.

Definition 2.2. Let X_t be an order M and k-variate Markov chain, with alphabet $A = B^k$, state space $\mathcal{S} = A^M$, $M < \infty$ and partition of the state space $\mathcal{L}$,

i. for each $L \in \mathcal{L}$ define $\mathcal{D}_L$ as the largest partition of $\{1, 2, ..., k\}$, such that
$$P(a|L) = \prod_{C \in \mathcal{D}_L} P(a^C|L),\ \ \forall a \in A.$$
ii. $\mathcal{D}_{\mathcal{L}} = \{\mathcal{D}_L\}_{L \in \mathcal{L}}$ is the *structure of interaction* of the process X_t.

iii. $\mathcal{M} = \{M_P\}_{P \in \mathcal{D}_{\mathcal{L}}}$ with $M_P = \cup_{\{L \in \mathcal{L}: \mathcal{D}_L = P\}} L$ and $\mathcal{D}_L$ given by i., is the *partition of interactions of* $\mathcal{S}$.

According to Definition 2.2-i., we note that the number of parameters needed to specify the transition probability from L to a is $\sum_{C \in \mathcal{D}_L}(|B|^{|C|} - 1), \forall a \in A$.

Example 2.3. Suppose the partition of $\mathcal{S}$ (Definition 2.1)-ii is given by 4 parts, say $\mathcal{L} = \{L_1, L_2, L_3, L_4\}$, in a process with $k = 5$ coordinates. Suppose also the structure of interaction is given by Table 1. The elements (list of parts of $\mathcal{L}$) of the

Table 1. Parts of the partition related to $\sim$ and its structure of interaction.

Part	L_1	L_2	L_3	L_4
$\mathcal{D}_L$	$\{1, 2\}, \{3, 4, 5\}$	$\{1, 2, 3\}, \{4, 5\}$	$\{1, 2\}, \{3, 4, 5\}$	$\{1, 2\}, \{3\}, \{4, 5\}$

TABLE 2. Composition of $\mathcal{M}$, the partition of interactions.

Part of $\mathcal{M}$	$M_{\mathcal{D}_{L_1}}$	$M_{\mathcal{D}_{L_2}}$	$M_{\mathcal{D}_{L_4}}$
Parts of $\mathcal{L}$	$L_1 \cup L_3$	L_2	L_4

partition of interactions $\mathcal{M}$ (Definition 2.2-iii) are exposed in Table 2. This means that all the strings in the state space $\mathcal{S}$ included in the parts L_1 and L_3 share the same interaction between the coordinates, and split the 5 coordinates in two independent sets $\{1,2\}$ and $\{3,4,5\}$. If we consider the part L_i, for $i = 1,3$, we have $P(a|L_i) = P(a^{1,2}|L_i)P(a^{3,4,5}|L_i)$, with $a^{1,2} = (a_1, a_2), a^{3,4,5} = (a_3, a_4, a_5)$, $\forall a = (a_1, a_2, a_3, a_4, a_5)$.

3. Estimation

In this section we will introduce the methodology of estimation, which consists of the strategic use of the Bayesian Information Criterion to produce consistent estimates of the structure of interaction. We give an overview of the process of estimating the partition $\mathcal{L}$, and in the sequence, the article is devoted to the estimation of the structure of interaction.

3.1. Estimation of $\mathcal{L}$, a Review. Consider a sample x_1^n of the process X_t, $a \in A$ and $s \in \mathcal{S}$. We will denote by $N_n(s)$ the number of occurrences of s in the sample and by $N_n(s,a)$ the number of occurrences of s followed by a in the sample, $N_n(s) = \sum_{m=M+1}^{n+1} 1_{\{x_{m-M}^{m-1}=s\}}$, $N_n(s,a) = \sum_{m=M+1}^{n} 1_{\{x_{m-M}^{m-1}=s, x_m=a\}}$.

Definition 3.1. Let x_1^n be a sample of X_t, for any $s, r \in \mathcal{S}$,

$$\begin{aligned} d_n(s,r) \quad = \quad & \frac{2}{(|A|-1)\ln(n)} \sum_{a \in A} \left\{ N_n(s,a) \ln\left(\frac{N_n(s,a)}{N_n(s)}\right) \right. \\ & + N_n(r,a) \ln\left(\frac{N_n(r,a)}{N_n(r)}\right) \\ & \left. - (N_n(\{s,r\},a) \ln\left(\frac{N_n(\{s,r\},a)}{N_n(s)+N_n(r)}\right) \right\}, \end{aligned}$$

with $N_n(\{s,r\},a) = N_n(s,a) + N_n(r,a)$.

d_n can be generalized to subsets of $\mathcal{S}$ and it has the property of being equivalent to the Bayesian Information Criterion to decide if $s \sim r$ for any $s, r \in \mathcal{S}$. See [5] for a more complete explanation.

Remark 3.2. (i) As a consequence of Theorem 2.1 proved in [5], if X_t is a discrete time, order M, Markov chain on a finite alphabet A and x_1^n is a sample of the process, then for n large enough, for each $s, r \in \mathcal{S}$, $d_n(r,s) < 1$ if and only if s and r belong to the same class.

(ii) The algorithm introduced in [6] (using d_n) returns the true partition for the source, this means that under the assumptions of Theorem 2.1 [5], $\hat{\mathcal{L}}_n$ given by the algorithm converges almost surely eventually to $\mathcal{L}$, where $\mathcal{L}$ is the partition of $\mathcal{S}$ given by Definition 2.1-ii.

(iii) For each $a \in A$ and $s \in \mathcal{S}$, the estimator of $P(a|s)$ or $P(a|\hat{L})$ is defined by $\sum_{r\in\hat{L}} N_n(r,a)/\sum_{r\in\hat{L}} N_n(r)$, such that $s \in \hat{L}$ and $\hat{L}$ is a part of $\hat{\mathcal{L}}_n$.

We complete this brief exposition of the estimation process, used in this paper, with the following references regarding the estimation of variable length Markov chains and Markov chains both included in the PMM class, see [2], [3] and [10].

3.2. Consistent Estimation of the Structure of Interaction. In this section we present the maximum likelihood expression that allows the estimation of the underlying structure $D_{\mathcal{L}}$, introduced by Definition 2.2. Based on an estimate of the partition of the state space $\mathcal{S}$, the Bayesian Information Criterion enables to obtain an estimated structure that converges eventually almost surely to the true structure of interaction. To estimate the probabilities we introduce some complementary notation, for $s \in \mathcal{S}$ the number of occurrences in x_1^n of the string s followed by a vector that has the coordinates listed by u equal to a^u, $N_n(s,a^u) = \big|\{t : M < t \leq n, x_{t-M}^{t-1} = s, (x(u_1)_t, ..., x(u_l)_t) = a^u\}\big|$, where each element x_t is such that $x_t = (x(1)_t, x(2)_t, \ldots, x(k)_t)$. In addition for each part $L \in \mathcal{L}$, $N_n^{\mathcal{L}}(L,a^u) = \sum_{s\in L} N_n(s,a^u)$ and $N_n^{\mathcal{L}}(L) = \sum_{s\in L} N_n(s)$. Given a sample of the process x_1^n, if we write $P(x_1^n) = \text{Prob}(X_1^n = x_1^n)$, we obtain under the assumption of a hypothetical partition $\mathcal{L}$ of $\mathcal{S}$, Definition 2.1-ii, and the structure of interaction $\mathcal{D}_{\mathcal{L}}$, given by Definition 2.2-ii., $P(x_1^n) = P(x_1^M)\prod_{L\in\mathcal{L},a\in A}\prod_{C\in\mathcal{D}_L} P(a^C|L)^{N_n^{\mathcal{L}}(L,a)}$. The maxima for $\prod_{L\in\mathcal{L},a\in A}\prod_{C\in\mathcal{D}_L} P(a^C|L)^{N_n^{\mathcal{L}}(L,a)}$ is

$$\text{ML}(\mathcal{L},\mathcal{D}_{\mathcal{L}},x_1^n) = \prod_{L\in\mathcal{L},a\in A}\prod_{C\in\mathcal{D}_L}\left(\frac{N_n^{\mathcal{L}}(L,a^C)}{N_n^{\mathcal{L}}(L)}\right)^{N_n^{\mathcal{L}}(L,a)},$$

and the Bayesian Information Criterion expression under this formulation will be given by the next definition.

Definition 3.3. Let x_1^n a sample from the k-variate Markov chain X_t, of order M, $M < \infty$ with alphabet $A = B^k$ and state space $\mathcal{S} = A^M$. Suppose $\mathcal{L}$ is the partition introduced by Definition 2.1-ii. and $\mathcal{D}_{\mathcal{L}}$ is the structure of interaction given by Definition 2.2-ii., then, the corresponding BIC is $\text{BIC}(\mathcal{L},\mathcal{D}_{\mathcal{L}},x_1^n) = \ln(\text{ML}(\mathcal{L},\mathcal{D}_{\mathcal{L}},x_1^n)) - \sum_{L\in\mathcal{L}}\sum_{C\in\mathcal{D}_L}(|B|^{|C|}-1)\frac{\ln(n)}{2}$.

The model selection methodology is consistent as we show in the next result. A preliminary version of this result is displayed in [7] (without proof). Here we present all the results with proofs. We will use to prove those results the divergence measure between two probability distributions, $D(P||Q) = \sum_{b\in C} P(b)\ln\{\frac{P(b)}{Q(b)}\}$ for P and Q probability distributions on C.

Theorem 3.4. *Let X_t be a k-variate Markov chain, of order $M, M < \infty$, with alphabet $A = B^k$ and state space $\mathcal{S} = A^M$. With partition of the state space $\mathcal{L}$ (Definition 2.1-ii.) and structure of interaction $\mathcal{D}_{\mathcal{L}}$ (Definition 2.2-ii.). Define, $\hat{\mathcal{D}} = \arg\max_{\mathcal{D}\in\mathbf{D}}\{BIC(\mathcal{L},\mathcal{D},x_1^n)\}$, where D is the set of all possible structures under the condition (2.1) and $BIC(\mathcal{L},\mathcal{D},x_1^n)$ is given by Definition 3.3 in $\mathcal{D} \in D$. Then, $\hat{\mathcal{D}} = \mathcal{D}_{\mathcal{L}}$ eventually almost surely as $n \to \infty$.*

Proof. Suppose $\hat{\mathcal{D}} = \{\hat{D}_{L_1}, \ldots, \hat{D}_{L_{|\mathcal{L}|}}\} \neq \mathcal{D}_{\mathcal{L}}\ \forall n$. Then, from Definition 2.2 $\exists\ \tilde{\mathcal{D}} = \{\tilde{D}_{L_1}, \ldots, \tilde{D}_{L_{|\mathcal{L}|}}\}$ such that at least one part (say L_1) is such that $|\tilde{D}_{L_1}| \geq |\hat{D}_{L_1}|$. Note that the indexing of $\hat{\mathcal{D}}$ and $\tilde{\mathcal{D}}$ is the same, and equal to $|\mathcal{L}|$, because $\mathcal{L} = \{L_1, \ldots, L_{|\mathcal{L}|}\}$ is the true partition. By simplicity we suppose $\tilde{D}_{L_i} = \hat{D}_{L_i}, \forall i \neq 1$. We have $\hat{D}_{L_1} = \{G_1, G_2, \ldots, G_{J_{L_1}}\}$ where G_i is a collection of coordinates. Also we assume $\tilde{D}_{L_1} = \{\tilde{G}_{11}, \tilde{G}_{12}, G_2, \ldots, G_{J_{L_1}}\}$. It means that G_1 according to $\hat{D}_{L_1}$ represents a block of coordinates, which is split into two sets $G_1 = \tilde{G}_{11} \cup \tilde{G}_{12}$ according to $\tilde{D}_{L_1}$. $P(a^{G_1}|L_1) = P(a^{\tilde{G}_{11}}|L_1)P(a^{\tilde{G}_{12}}|L_1), \forall a \in A$, then by Proposition 7.1, there exists n_0 such that $BIC(\mathcal{L}, \tilde{\mathcal{D}}, x_1^n) > BIC(\mathcal{L}, \hat{\mathcal{D}}, x_1^n)$, when $n > n_0$, which is a contradiction. On the other hand if there is a value $a \in A$ such that $P(a^{G_1}|L_1) \neq P(a^{\tilde{G}_{11}}|L_1)P(a^{\tilde{G}_{12}}|L_1)$, define $\delta := D\Big(P(\cdot^{G_1}|L_1)||P(\cdot^{\tilde{G}_{11}}|L_1)P(\cdot^{\tilde{G}_{12}}|L_1)\Big)$ and $\delta_n := D\Big(\hat{P}(\cdot^{G_1}|L_1)||\hat{P}(\cdot^{\tilde{G}_{11}}|L_1)\hat{P}(\cdot^{\tilde{G}_{12}}|L_1)\Big)$, observe that $\delta > 0$ and $\hat{P}$ is constructed following remark 3.2-iii. Following the same ideas from the proof of Proposition 7.1 we can write $\frac{BIC(\mathcal{L},\tilde{\mathcal{D}},x_1^n) - BIC(\mathcal{L},\hat{\mathcal{D}},x_1^n)}{n} = -\delta_n + c_0\frac{\ln(n)}{n}$, for some positive constant value c_0. (i) Given a constant value $K > 1$ and $\epsilon = \frac{\delta}{K}$ there is a value $n_1 > 1$ such that $\forall n > n_1 : c_0\frac{\ln(n)}{n} < \frac{\delta}{K}$, (ii) because $\delta_n \to \delta$, when $n \to \infty$, then given $\epsilon = (1 - \frac{2}{K})\delta$ there is a value $n_2 >$ such that $\forall n > n_2 : \delta_n > \frac{2}{K}\delta$. From (i) and (ii), if $n > \max(n_1, n_2)$, $\frac{BIC(\mathcal{L},\tilde{\mathcal{D}},x_1^n) - BIC(\mathcal{L},\hat{\mathcal{D}},x_1^n)}{n} = -\delta_n + c_0\frac{\ln(n)}{n} < -\frac{\delta}{K}$. As a consequence $BIC(\mathcal{L}, \tilde{\mathcal{D}}, x_1^n) - BIC(\mathcal{L}, \hat{\mathcal{D}}, x_1^n) < -\frac{\delta}{K}n < 0$, when $n > \max(n_1, n_2)$. So, it is not possible for a partition $\tilde{\mathcal{D}}$ under this condition, to be the arg max of the BIC. □

3.3. Full Estimation of the Structure of Interaction. In this section, we will estimate the structure of interaction using the Bayesian Information Criterion. The strategy proposed in this paper, to estimate the structure of interaction, exceeds the detection capability of the proposal in [7]. In [7] it is showed a consistent algorithm to detect pairwise independence between the coordinates of a multivariate process. Nevertheless that pairwise consistent algorithm is not appropriate to situations as the exposed by the next example. In contrast, the strategy introduced by this paper is specially designed to detect situations such as the introduced in the example 3.5, since this strategy is jointly consistent.

Example 3.5. (see [11]) Let (X_1, X_2, X_3) be a vector with joint probability mass $p(x_1, x_2, x_3) = \frac{1}{4}$ if $(x_1, x_2, x_3) \in \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\}$ and $p(x_1, x_2, x_3) = 0$ otherwise. As a consequence, the pairwise probability mass function is $p_{ij}(x_i, x_j) = \frac{1}{4}$ if $(x_i, x_j) \in \{(0, 0), (1, 0), (0, 1), (1, 1)\}$ and $p_{ij}(x_i, x_j) = 0$ otherwise, if $i \neq j$, $i, j \in \{1, 2, 3\}$. Even more, $p_i(x_i) = \frac{1}{2}$ if $x_i = 0, 1$ and $p_i(x_i) = 0$ otherwise. Then, the variables X_i and X_j are pairwise independent, because $p_{ij}(x_i, x_j) = p_i(x_i)p_j(x_j)$, when $i \neq j$ with $i, j \in \{1, 2, 3\}$. But $p(x_1, x_2, x_3) \neq p_1(x_1)p_2(x_2)p_3(x_3)$ and this means that X_1, X_2 and X_3 are not mutually independent.

Now, we show a result which support the strategy of estimation, based on Theorem 3.4.

Theorem 3.6. *Let X_t be under the assumptions of Theorem 3.4 and consider x_1^n a sample of the k-variate process X_t. Define $\hat{\mathcal{D}}_n = \arg\max_{\mathcal{D}\in\mathbf{D}}\{BIC(\hat{\mathcal{L}}_n, \mathcal{D}, x_1^n)\}$. where $\hat{\mathcal{L}}_n$ is the estimator of $\mathcal{L}$ - see Remark 3.2-ii - and $\mathbf{D}$ is the set of all possible structures under the condition (2.1), $BIC(\hat{\mathcal{L}}_n, \mathcal{D}, x_1^n)$ is given by Definition 3.3 and evaluated in $\hat{\mathcal{L}}_n$ and $\mathcal{D} \in D$. Then, eventually almost surely as $n \to \infty$, $\hat{\mathcal{D}}_n = \mathcal{D}_{\mathcal{L}}$.*

Proof. According to [6]: given a sample x_1^n of X_t, $\exists\ n_1$ such that $\hat{\mathcal{L}}_n = \mathcal{L},\ \forall n > n_1$ with probability 1. Therefore, $\hat{\mathcal{D}}_n = \arg\max_{\mathcal{D}\in\mathbf{D}}\{BIC(\mathcal{L}, \mathcal{D}, x_1^n)\}, \forall n > n_1$. From Theorem 3.4, $\hat{\mathcal{D}}_n \to \mathcal{D}_{\mathcal{L}}$ eventually almost surely when $n \to \infty$. This means $\exists\ n_2$ such that $\hat{\mathcal{D}}_n = \mathcal{D}_{\mathcal{L}}, \forall n > n_2$, with probability 1. □

The following relation supports the estimation strategy, derived from Theorem 3.6.

3.3.1. *Strategy.* From Definition 3.3 for a fixed partition $\mathcal{L}$, $BIC(\mathcal{L}, \mathcal{D}_{\mathcal{L}}, x_1^n)$ is

$$\sum_{L\in\mathcal{L}}\Big(\sum_{C\in\mathcal{D}_L}\Big[\sum_{a\in A} N_n^{\mathcal{L}}(L,a)\ln\left(\frac{N_n^{\mathcal{L}}(L,a^C)}{N_n^{\mathcal{L}}(L)}\right) - (|B|^{|C|}-1)\frac{\ln(n)}{2}\Big]\Big), \tag{3.1}$$

the contribution of each part L in $\mathcal{L}$ will be denoted by $BIC_L(\mathcal{D}_L, x_1^n)$ where $\mathcal{D}_L$ is an element in the structure of interaction $\mathcal{D}_{\mathcal{L}}$,

$$BIC_L(\mathcal{D}_L, x_1^n) = \sum_{C\in\mathcal{D}_L}\Big[\sum_{a\in A} N_n^{\mathcal{L}}(L,a)\ln\left(\frac{N_n^{\mathcal{L}}(L,a^C)}{N_n^{\mathcal{L}}(L)}\right) - (|B|^{|C|}-1)\frac{\ln(n)}{2}\Big]. \tag{3.2}$$

From equations (3.1)-(3.2),

$$BIC(\mathcal{L}, \mathcal{D}_{\mathcal{L}}, x_1^n) = \sum_{L\in\mathcal{L}} BIC_L(\mathcal{D}_L, x_1^n), \tag{3.3}$$

then, for each $L \in \mathcal{L}$,

$$\arg\max_{\mathcal{D}\in\mathbf{D}}\{BIC(\mathcal{L}, \mathcal{D}_{\mathcal{L}}, x_1^n)\}|_L = \arg\max_{\mathcal{D}_L\in\mathbf{D}^L}\{BIC_L(\mathcal{D}_L, x_1^n)\}. \tag{3.4}$$

Where $\mathbf{D}^L$ is the set of all possible structures following the condition (2.1) over $\{1, ..., k\}$ of L. And $\arg\max_{\mathcal{D}\in\mathbf{D}}\{BIC(\cdot)\}|_L$ denotes the arg max related to the L part - left side of equation (3.4). The strategy is expressed by the equation (3.4), that is, using the term on the right of this equation, we can find by the Bayesian Information Criterion, the optimal structure of interaction for each estimated part. See the procedure: (1) obtain $\hat{\mathcal{L}}_n$; (2) for each $\hat{L} \in \hat{\mathcal{L}}_n$ obtain $\hat{\mathcal{D}}_{\hat{L}}$; (3) define $\hat{\mathcal{D}} = \{\hat{\mathcal{D}}_{\hat{L}} : \hat{L} \in \hat{\mathcal{L}}_n\}$. This procedure being performed for each element $\hat{L}$, is equivalent to performing the maximization for all elements of $\hat{\mathcal{L}}_n$ simultaneously - equation (3.3). Furthermore, the convergence of this strategy is assured by the Theorem 3.6. The strategy represented by the equations (3.3)-(3.4) is specially designed to detect situations such as example 3.5, since this strategy is jointly consistent.

4. Application

In the application we study four series of the Brazilian financial market: VALE (VALE5), ITAUUNIBANCO (ITUB4), AMBEV S/A (ABEV3) and BRADESCO (BBDC4). For each marginal process we study the returns of the closing prices. We define $X_t(i) = 1$ if $P_{t+1}(i) \geq P_t(i)$ and $X_t(i) = 0$ otherwise ($B = \{0, 1\}$), where $P_t(i)$ is the closing price, at time t of the financial series i, with $i = 1, 2, 3, 4$. Table 3 identifies each series. The four series considered here compound the BOVESPA

Table 3. Four of the main BOVESPA index's stocks. Period: January, 02-2006–July, 31-2015, sample size=2368.

Coordinate (i)	Company	Code	Sector
1	VALE	VALE5	Basic Materials/Mining
2	ITAUUNIBANCO	ITUB4	Financial
3	AMBEV S/A	ABEV3	Non-Cyclical Consumer Goods
4	BRADESCO	BBDC4	Financial

index (IBOVESPA). IBOVESPA is the main indicator of the Brazilian stock market's performance. In Table 4 we present the results obtained by applying the Bayesian Information Criterion. The memory of the joint process is $\lfloor \log_{|A|}(2367) \rfloor$ =2, with $|A|$=16. In order to illustrate the estimation strategy introduced here, in Table 4 we selected the top 3 Bayesian Information Criterion values for each of the estimated parts, obtained from Definition 3.1, and on the left side of each value of the Bayesian Information Criterion, we show the structure of interaction for each case. The composition of each part $\hat{L}_i$ member of $\hat{\mathcal{L}}_n$ can be obtained in `www.ime.unicamp.br/~jg/sm/` and Table 5 shows two of them. It is concluded from the Table 4 that the partition of the state space $\mathcal{M}$ is composed by 2 elements, the estimates are given by the junction of parts 1 and 2 in one case, $\hat{M}_1 = \hat{L}_1 \cup \hat{L}_2$ and parts 3 and 4 on the other, $\hat{M}_2 = \hat{L}_3 \cup \hat{L}_4$, meaning that for all strings in $\hat{M}_1$ the 4 series are dependent and for all strings in $\hat{M}_2$ the series 1 and 2 will behave independently of series 3 and 4. These results are summarized in Table 6. For instance if we focus on the state $x_{t-1} = (1, 1, 1, 0)$, we have that for an arbitrary state $x_t = (x_t(1), x_t(2), x_t(3), x_t(4))$, $P(x_t|x_{t-2}^{t-1}) = P(x_t(1), x_t(2)|x_{t-2}^{t-1})P(x_t(3), x_t(4)|x_{t-2}^{t-1})$, for $x_{t-2} = (0, 1, 1, 1), (1, 0, 0, 1), (0, 1, 0, 1)$ (cases listed in Table 5). Since the strings $(0, 1, 1, 1)(1, 1, 1, 0)$ and $(1, 0, 0, 1)(1, 1, 1, 0) \in \hat{L}_3$ and $(0, 1, 0, 1)(1, 1, 1, 0) \in \hat{L}_4$, we have

$$\begin{aligned} P(x_t|(0,1,1,1)(1,1,1,0)) &= P(x_t|(1,0,0,1)(1,1,1,0)) \\ &\neq P(x_t|(0,1,0,1)(1,1,1,0)). \end{aligned}$$

For all the strings in $\hat{M}_2$, the two series from the financial sector are separated each other. Price increases for the financial series ITUB4 are dependent on the values of VALE5, from the sector *Basic Materials/Mining*. Similarly, increases in the price of the financial series BBDC4 are dependent of ABEV3, coming from the sector *Non-Cyclical Consumer Goods*.

TABLE 4. $\hat{\mathcal{D}}_{\hat{L}_i}$ associated with each estimated part $\hat{L}_i, i = 1, 2, 3, 4$. In boldface letter the best result of the Bayesian Information Criterion and at its left the winning partition of coordinates. $\hat{L}_i$ estimated by the distance given by Definition 3.1.

Part of $\hat{\mathcal{L}}_n$	$\hat{\mathcal{D}}_{\hat{L}}$	BIC value
$\hat{L}_1$	$\{\mathbf{1,2,3,4}\}$	**-5152.6078**
	$\{1,3,4\}\{2\}$	-5315.7120
	$\{1,2,3\}\{4\}$	-5341.8196
$\hat{L}_2$	$\{\mathbf{1,2,3,4}\}$	**-189.9074**
	$\{1,2,4\}\{3\}$	-235.9355
	$\{1,2,3\}\{4\}$	-242.3120
$\hat{L}_3$	$\{\mathbf{1,2}\}\{\mathbf{3,4}\}$	**-103.6107**
	$\{1,3\}\{2,4\}$	-103.6399
	$\{1,4\}\{2,3\}$	-105.6649
$\hat{L}_4$	$\{\mathbf{1,2}\}\{\mathbf{3,4}\}$	**-111.9021**
	$\{1,4\}\{2,3\}$	-115.8587
	$\{1,3\}\{2,4\}$	-116.3862

TABLE 5. Composition of two parts.

Part	Strings
$\hat{L}_3$	(0,0,0,0)(0,1,1,0); (0,0,0,1)(0,0,0,0); (0,1,1,1)(1,1,1,0)
	(1,0,0,1)(1,1,1,0); (1,0,1,1)(0,0,1,1); (1,1,0,1)(1,0,1,1)
$\hat{L}_4$	(0,0,1,1)(1,0,1,1); (0,1,0,1)(1,0,0,0); (0,1,0,1)(1,1,1,0)
	(0,1,1,1)(1,0,1,0); (1,0,0,0)(1,0,0,0); (1,0,0,0)(1,0,1,0)

TABLE 6. Structure of Interaction by parts which share the same type of interaction between the series.

$\hat{M}_i$	Interaction in $\hat{M}_i$
$\hat{M}_1 = \hat{L}_1 \cup \hat{L}_2$	{VALE5, ITUB4, ABEV3, BBDC4}
$\hat{M}_2 = \hat{L}_3 \cup \hat{L}_4$	{VALE5, ITUB4}, {ABEV3, BBDC4}

5. Conclusion

In this article we review the concept of interaction, introduced in [7]. We introduce a strategy to find groups of coordinates, conditionally independent. We show in Theorem 3.6 (see also Theorem 3.4) that is possible to consistently estimate the structure of interaction. Also, based on the construction of the overall Bayesian Information Criterion is possible to quantify the contribution of each part (members of $\mathcal{L}$) to the overall Bayesian Information Criterion. From that, we propose a strategy, based on Theorem 3.6 to produce an efficient estimate which detects the mutual independence. In this way we ended the problem introduced in [7] in the context of partition Markov models and by [19] in the context of variable length Markov models (particular cases of partition Markov models). Finally, in the application to real data we analysed four main financial series of the Brazilian market, we observe that there is a portion of the state space in which can be identified two independent groups of series, those pairs of series are: {VALE (VALE5), ITAUUNIBANCO (ITUB4)} and {AMBEV S/A (ABEV3), BRADESCO (BBDC4)}, respectively.

6. Acknowledgments

The authors wish to express their gratitude to two referees and the editor for their helpful comments on an earlier draft of this paper.

References

[1] Buhlmann, P. and Wyner, A. : Variable length Markov chains, *The Annals of Statistics* **27** (1999), 480-513.

[2] Csiszár, I. and Shields, P. C.: The consistency of the BIC Markov order estimator, *The Annals of Statistics* **28** (2000), (6) 1601-1619.

[3] Csiszár, I. and Talata, Z.: Context tree estimation for not necessarily finite memory processes, via BIC and MDL, *IEEE Trans. Inform. Theory* **52** (2006), 1007-1016.

[4] Farcomeni. A.: Hidden Markov partition models, *Statistics & Probability Letters* **81** (2011), (12) 1766-1770.

[5] García, Jesús E. and González-López, V.A.: Minimal Markov Models, arXiv preprint *arXiv:1002.0729*, 2010.

[6] García, Jesús E. and González-López, V.A.: Minimal Markov Models, in: *Proceedings of the Fourth Workshop on Information Theoretic Methods in Science and Engineering,* **1** (2011) 25-28, Helsinki.

[7] García, Jesús E. and González-López, V.A.: Interaction partition. Structure of interacting coordinates for a multivariate stochastic process, in: *SMTDA2014 Conference Proceedings,* G-L (2014) 1.

[8] García, Jesús E., González-López, V.A. and Hirsh, I. D.: Copula-based prediction of economic movements, in: *International Conference of Numerical Analysis and Applied Mathematics 2015: ICNAAM 2015.* Vol. 1738, No. 1, AIP Publishing (2016) pp 140005.

[9] García, Jesús E., González-López, V.A. and Viola, M.L.L.: Robust Model Selection for Stochastic Processes, *Communications in Statistics-Theory and Methods* **43** (2014) (10-12), 2516-2526.

[10] Gusfield, D.: *Algorithms on strings, trees and sequences: computer science and computational biology,* Cambridge university press, 1997.

[11] Hogg, R. V., McKean, J. W. and Craig, A. T.: *Introduction to Mathematical Statistics,* Upper Saddle River, N.J: Pearson Education, 2005.

[12] Jääskinen, V., Xiong, J., Corander, J. and Koski, T.: Sparse markov chains for sequence data, *Scandinavian Journal of Statistics* **41** (2014) (3) 639-655.

[13] Joe, H.: *Multivariate models and multivariate dependence concepts,* CRC Press, 1997.

[14] Manning, C. D. and Schütze, H.: *Foundations of statistical natural language processing,* Vol. 999. Cambridge: MIT press, 1999.

[15] Messaoud, S.B. and Aloui, C.: Measuring Interdependence and Contagion: A Copula Approach, *International Review of Business Research Papers* **11** (2015) 1.

[16] Nelsen, R. B.: *An Introduction to Copulas,* Springer: Berlin/Heidelberg, Germany, 2006.

[17] Rissanen, J.: A universal data compression system, *IEEE Trans. Inform. Theory* **29** (1983) (5) 656-664.

[18] Schwarz, G.: Estimating the dimension of a model, *The Annals of Statistics* **6** (1978) (2) 461-464.

[19] Viola, M.L: Tópicos em seleção de modelos markovianos, PhD Tesis, 2011. Accessed on 5 September 2016. http://www.bibliotecadigital.unicamp.br/document/?code=000844289.

[20] Weinberger, M. J., Rissanen, J. J. and Feder, M.: A universal finite memory source, *IEEE Trans. Inform. Theory* **41** (1995) (3) 643-652.

[21] Yang, L., Cai, X. J., Li, M. and Hamori, S.: Modeling dependence structures among international stock markets: Evidence from hierarchical Archimedean copulas, *Economic Modelling* **51** (2015) 308-314.

7. Appendix

Proposition 7.1. *Given $(X,Y) \sim P$, $\{(x_i,y_i)\}_{i=1}^n$ i.i.d. sample of (X,Y) with values in $A=\{0,1\}^2$. Then, if X and Y are independent, $\exists n_0 > 1$:*

$$Dif(\{(x_i,y_i)\}_{i=1}^n) = BIC(\{(x_i,y_i)\}_{i=1}^n, I) - BIC(\{(x_i,y_i)\}_{i=1}^n) > 0, \forall n > n_0.$$

Where $BIC(\{(x_i,y_i)\}_{i=1}^n, I)$ denotes the BIC under the assumption of independence between X and Y.

Proof. If $r_n(x,y) = \frac{\sum_{i=1}^n 1_{\{(x_i,y_i)=(x,y)\}}}{n}$, $r_{n,1}(x) = \frac{\sum_{i=1}^n 1_{\{x_i=x\}}}{n}$ and $r_{n,2}(y) = \frac{\sum_{i=1}^n 1_{\{y_i=y\}}}{n}$. In general,

$$BIC(\{(x_i,y_i)\}_{i=1}^n) = \sum_{(x,y)\in A} nr_n(x,y)\ln(r_n(x,y)) - \frac{3\ln(n)}{2},$$

because $A=\{0,1\}^2$.
On the other hand, if X and Y are independent, $BIC(\{(x_i,y_i)\}_{i=1}^n, I)$ is

$$\sum_{x\in\{0,1\}} nr_{n,1}(x)\ln(r_{n,1}(x)) + \sum_{y\in\{0,1\}} nr_{n,2}(y)\ln(r_{n,2}(y)) - \frac{2\ln(n)}{2}$$

because in this case we have only 2 parameters to specify which are the marginal distributions $P_X(0)=\sum_y P(0,y)$ and $P_Y(0)=\sum_x P(x,0)$. Then,

$$\begin{aligned} Dif(\{(x_i,y_i)\}_{i=1}^n) &= \sum_{x\in\{0,1\}} nr_{n,1}(x)\ln(r_{n,1}(x)) + \sum_{y\in\{0,1\}} nr_{n,2}(y)\ln(r_{n,2}(y)) \\ &\quad - \sum_{(x,y)\in A} nr_n(x,y)\ln(r_n(x,y)) + \frac{\ln(n)}{2} \\ &= \sum_{(x,y)\in A} nr_n(x,y)\ln\Big(\frac{r_{n,1}(x)r_{n,2}(y)}{r_n(x,y)}\Big) + \frac{\ln(n)}{2}. \end{aligned}$$

Then, $\frac{BIC(\{(x_i,y_i)\}_{i=1}^n, I) - BIC(\{(x_i,y_i)\}_{i=1}^n)}{n} > 0$ if and only if,

$$\frac{\ln(n)}{2n} > D\Big(r_n(\cdot,\cdot)||r_{n,1}(\cdot)r_{n,2}(\cdot)\Big).$$

From [3] we have $D\Big(r_n(\cdot,\cdot)||r_{n,1}(\cdot)r_{n,2}(\cdot)\Big) \le \frac{\sum_{(x,y)\in A}\Big(r_{n,1}(x)r_{n,2}(y)-r_n(x,y)\Big)^2}{r_{n,1}(x)r_{n,2}(y)}$.
Because X and Y are independent, the numerator $r_{n,1}(x)r_{n,2}(y) - r_n(x,y)$ goes to zero when $n\to\infty$, and the left side will be arbitrary small when $n\to\infty$. □

M. Fernández: BM&FBovespa, São Paulo, SP, CEP 01010-901, Brazil
E-mail address: mariela.fer@gmail.com

Jesús E. García: Department of Statistics, University of Campinas, Campinas, SP, CEP: 13083-859, Brazil
E-mail address: jg@ime.unicamp.br

V.A. González-López: Department of Statistics, University of Campinas, Campinas, SP, CEP: 13083-859, Brazil
E-mail address: veronica@ime.unicamp.br

M.L. Lanfredi Viola: Federal University of São Carlos, São Carlos, SP, CEP: 13565-905, Brazil
E-mail address: lanfredi@ufscar.br