UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

GENERALIZED MULTIVARIATE MARKOV CHAINS: ESTIMATION, INFERENCE AND IMPLEMENTATION IN **R**

MASTER

APPLIED ECONOMETRICS AND FORECASTING

MASTER'S FINAL WORK

DISSERTATION

CAROLINA MICAEL DE ABREU E VASCONCELOS

SUPERVISION:

PROFESSOR BRUNO DAMÁSIO

NOVEMBER - 2021

Document exclusively elaborated for obtaining a master's degree

UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

GENERALIZED MULTIVARIATE MARKOV CHAINS: ESTIMATION, INFERENCE AND IMPLEMENTATION IN **R**

MASTER

APPLIED ECONOMETRICS AND FORECASTING

MASTER'S FINAL WORK

DISSERTATION

CAROLINA MICAEL DE ABREU E VASCONCELOS

SUPERVISION:

PROFESSOR BRUNO DAMÁSIO

NOVEMBER - 2021

Document exclusively elaborated for obtaining a master's degree

Acronyms

CRAN Comprehensive R Archive Network. i, 16

DGP Data Generating Process. i, 17, 20

DJIA Dow Jones Industrial Average. i, 23, 25, 27

FRED Federal Reserve Economic Data. i, 24

GLIM Generalized Linear Interactive Modeling. i, 11

GMMC Generalized Multivariate Markov Chain. i, 12, 16

HOMC High-Order Markov chain. i, 4, 8, 10, 11

MLE Maximum Likelihood Estimator. i, 12, 13, 28

MMC Multivariate Markov Chains. i, 1, 3–7, 11, 12, 14, 28

MTD Mixture Transition Distribution. i, 3, 4, 6–8, 10, 11, 13, 15, 28

TPM Transition Probability Matrix. i, 9

WoS Web of Science. i, 1

RESUMO

Este trabalho propõe uma nova generalização do modelo de Cadeias de Markov Multivariadas. Tipicamente, uma cadeia de Markov é descrita pelos valores passados do processo, a generalização proposta neste trabalho permiritá também considerar variáveis exógenas. Especificamente, iremos incorporar os efeitos dos valores passados do processo e os efeitos de variáveis pré-determinadas ou exógenas no modelo. Deste modo, será considerada uma cadeia de Markov não-homogénea em vez de uma cadeia de Markov homogénea. Os resultados da simulação de Monte Carlo mostraram que o modelo proposto detectou uma cadeia de Markov não-homogénea e detectou valores específicos dos parâmetros. Porém, quando esses valores eram baixos em magnitude, os resultados da simulação mostraram que o modelo tinha baixo poder de teste. Portanto, para estimativas de baixa magnitude, dever-se-á considerar um nível de significância mais alto ao testar a significância individual dos parâmetros. Adicionalmente, uma ilustração empírica demonstrou a relevância deste novo modelo, ao estimar a matriz de transição de probabilidade, para diferentes valores de uma variável exógena. Uma contribuição adicional e prática deste trabalho é o desenvolvimento de uma *package* **R** com esta generalização.

KEYWORDS: Cadeias de Markov Multivariadas; Mixture transition distribution Model; Cadeias de Markov de Elevada Ordem; Cadeias de Markov Multivaridas com variáveis exógenas; R package

ABSTRACT

This essay proposes a new generalization of Multivariate Markov Chains (MMC) model. Typically, a Markov chain is described by the process' past values, the generalization proposed in this work will also consider exogenous variables. Specifically, we will incorporate the effects of the process' past values and the effects of pre-determined or exogenous covariates in the model. This is achieved by considering a non-homogeneous Markov chain instead of an homogeneous Markov chain. The findings from the Monte Carlo simulation showed that the model proposed detected a non-homogeneous Markov chain and it detected specific values of the parameters. However, when these values were small in magnitude, the results from the simulation showed that the model had low power of test. Hence, for estimates with small magnitude, one should use a higher significance level when testing for individual significance of the parameters. Moreover, an empirical illustration demonstrated the relevance of this new model, by estimating the probability transition matrix, for different values of the exogenous variable. An additional and practical contribution of this work is the development of a novel **R** package with this generalization.

KEYWORDS: Multivariate Markov chains; Mixture transition distribution Model; High order Markov chains; Multivariate Markov chains with exogenous variables; R package

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

I would like to express my very great appreciation to my supervisor, Professor Bruno Damásio, for his guidance, enthusiastic encouragement, and helpful critiques and suggestions.

I would also like to thank my mother for the support given through my academic years and for always saying the right words.

Finally, I would like to thank my boyfriend, for listening and for the happy moments to rest my mind outside of this work.

1 INTRODUCTION

This essay proposes a new model for Multivariate Markov Chains (MMC) that incorporates the effect of exogenous variables. Instead of only considering the process' past values to explain the present, this model will also consider pre-determined or exogenous covariates.

Multivariate Markov chains have a wide range of applications, in various fields. A bibliometric analysis of all the relevant publications regarding multivariate Markov chains, collected from Web of Science (WoS), showed that, between 1977 and 2020, 1,708 articles mentioned "Multivariate Markov Chains" in the title, abstract, or keywords. These records were authored by 3,441 individuals and were published in 540 journals (unique ISSNs). Figure 1 displays the time evolution of the MMC-related articles. There is a first publication in the year 1977, followed by a gap in the sample until 1984. There is a take-off in 1997, from which the levels of published articles start to increase. In 2020, MMC-related articles reached the maximum number of published articles since 1977.

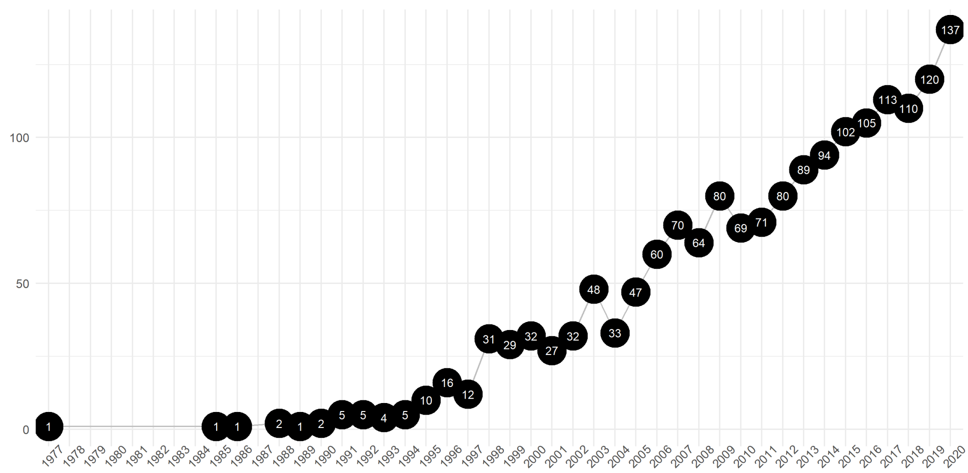


FIGURE 1: Number of articles per year

Moreover, the number of countries participating in the research output follows the same pattern as the number of articles published. In the 90's we see a take-off in the number of countries participating in the production and, we reach a global maximum in 2020, where 45 different countries contributed to the production of the articles of MMC. Therefore, throughout time, there has been increased production and international diffusion of Multivariate Markov chains.

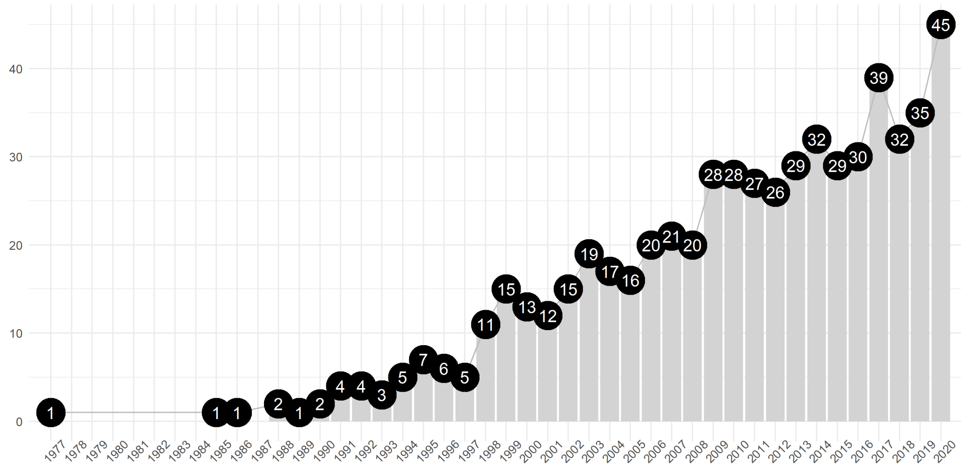


FIGURE 2: Number of countries participating per year

Regarding the number of articles per country, US is in the lead with 566 published articles, followed by China with 150 and UK with 139. In Figure 3, we have the top 25 most productive countries that comprise all the continents. However, European countries are more common.

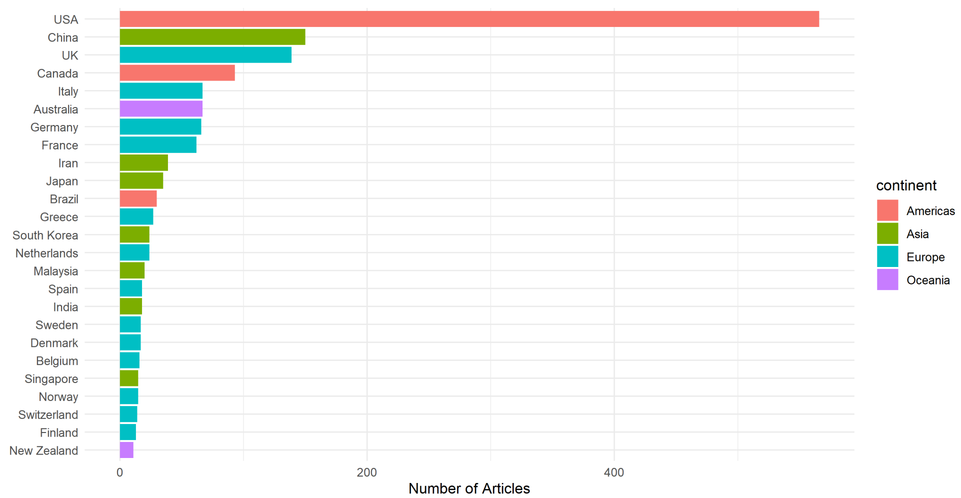


FIGURE 3: Number of articles by country

Finally, regarding the most predominant categories, Mathematics takes the lead in the number of articles per year. In the second half of the '90s, we see an increased interest in other fields, such as Mathematical & Computational Biology, Engineering, and Computer Science. Although these categories show an increased number of articles throughout the years, Mathematics is still in the lead with a substantial difference. In 2020, the most common categories were Mathematics, Computer Science, and Engineering.

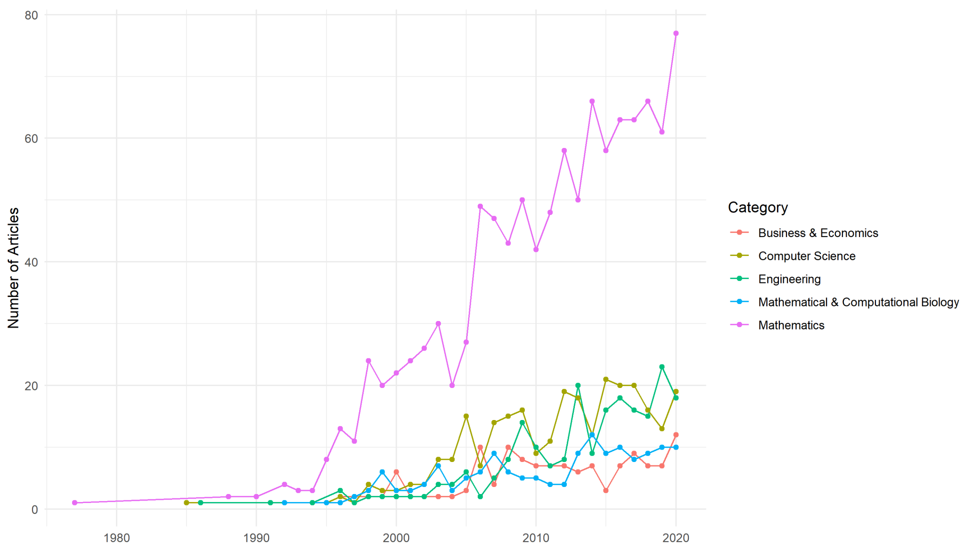


FIGURE 4: Time evolution of most relevant categories

Since the first record in 1977, there has been clear rise and interest in published MMC-related articles, with applications extended, not only to mathematics, but also to biology, engineering, and computer science. Thus, a new approach to the usual MMC models, that allows incorporating the effect of continuous covariates and, consequently, provide more precise estimates, would be a relevant contribution.

A well known model for MMC is the Mixture Transition Distribution (MTD) model, proposed by Raftery (1985). Ching, Fung, and Ng (2002) improved the MTD model, by proposing an alternative with fewer parameters. This was relevant since, for a high number of data sequences, order or states, it was unfeasible to model MMC through the MTD model. This work proposes a generalization of this alternative model proposed by Ching, Fung, and Ng (2002), that considers the effect of covariates. The generalization is achieved by considering non-homogeneous Markov chains, instead of homogeneous Markov chains. Hence, the research question is: Is it feasible to model Ching, Fung, and Ng (2002) MMC model with non-homogeneous Markov chains?. This essay intends to contribute to the field of MMC models, by answering the previous research question, through a Monte Carlo simulation that demonstrates the properties of the generalization proposed and an illustration for stock market returns, highlighting the main advantages of this approach. Finally, we implement the methods proposed in an **R** package.

The remain of this essay is organized as follows: Section 2 contains the literature review of this essay, followed by the theoretical framework, with some basic concepts for the fundamentals of the MMC model (Section 3). Section 4 presents the generalization proposed, followed by a description of its implementation in R, a Monte Carlo simulation study (Section 5), and an illustration of the methods proposed (Section 6). Lastly, Section 7 contains some concluding remarks regarding the developed work.

2 LITERATURE REVIEW

2.1 Multivariate Markov Chains

Markov chains can be appropriate for representing dependencies between successive observations of a random variable. However, when the order of the chain or the number of possible values increases, Markov chains have lack parsimony. In this context, several models for High-Order Markov chain (HOMC) were proposed, such as Jacobs and Lewis (1978), Pegram (1980), and Logan (1981). Notwithstanding these developments, the Mixture Transition Distribution model (Raftery (1985)) was proved to be more suitable to model HOMC, and it overshadowed the previously proposed models.

Several relevant extensions of the MTD model emerged: the Multimatrix MTD (Berchtold (1995), Berchtold (1996)), which allowed modeling the MTD by using a different $m \times m$ transition matrix for each lag, the Infinite-Lag MTD model that assumes an infinite lag order ($l = \infty$), which was first considered by Mehran (1989) and later developed by Le, Martin, and Raftery (1996) in a more general context. Finally, the MTD with General State Spaces allowed modeling more general processes with an arbitrary space state (Martin and Raftery (1987), Adke and Deshmukh (1988) and Wong and Li (2001)).

Although the MTD model presents a more parsimonious approach to model Markov chains with order higher than one, it has weaknesses. Namely, when considering more than one data sequence, one represents the MMC as a HOMC, by expanding the state-space. This approach could result in a more complex probability transition matrix. Consequently, this can make the estimation unfeasible as the order, states, and the number of data sequences increase. Additionally, the model assumes the same transition matrix for each lag.

In this setting, Ching, Fung, and Ng (2002) determined an alternative to handle the unfeasibility of the conventional multivariate Markov chain (MMC) by proposing a model with fewer parameters. The model developed is essentially the same as the MTD, however, it considers a different $m \times m$ transition matrix for each lag and considers more than one data sequence.

In the proposed multivariate Markov chain model, Ching, Fung, and Ng (2002) assume the following relationship:

Let $x_t^{(j)}$ be the state vector of the jth sequence at time t. If the jth sequence is in state

j at time t then

$$x_t^{(j)} = (0, \ldots, 0, \underbrace{1}_{\text{jth entry}}, 0, \ldots, 0)^t$$

The model is given by

$$x_{t+1}^{(j)} = \sum_{k=1}^{s} \lambda_{jk} P^{(jk)} x_t^{(k)}, \text{for } j = 1, 2, \ldots, s \tag{1}$$

where $\lambda_{jk} \geq 0$ for $1 \leq j, k \leq s$ and $\sum_{k=1}^{s} \lambda_{jk} = 1$ for $j = 1, 2, \ldots, s$.

The state probability distribution of the kth sequence at time $(t+1)$ depends on the weighted average of $P^{(jk)} x_t^{(k)}$. Here $P^{(jk)}$ is a transition probability matrix from the states in the kth sequence to the states in the jth sequence and $x_t^{(k)}$ is the state probability distribution of the kth sequences at time t. In matrix form:

$$\underline{x}_{t+1}^{(j)} \equiv \begin{bmatrix} x_{t+1}^{(1)} \\ \vdots \\ x_{t+1}^{(s)} \end{bmatrix} = \begin{bmatrix} \lambda_{11} P^{(11)} & \ldots & \lambda_{1s} P^{(1s)} \\ \vdots & \ddots & \vdots \\ \lambda_{s1} P^{(s1)} & \ldots & \lambda_{ss} P^{(ss)} \end{bmatrix} \begin{bmatrix} x_t^{(1)} \\ \vdots \\ x_t^{(s)} \end{bmatrix} \equiv Q \underline{x}_t \tag{2}$$

where Q is an $ms \times ms$ block matrix ($s \times s$ blocks of $m \times m$ matrices) and x_t is a stacked ms column vector (s vectors, each one with m rows).

For each data sequence, the matrices $P^{(jk)}$ can be estimated by counting the transition frequency from the states in the kth sequence to the states in the jth sequence, obtaining the transition frequency matrix for the data sequence. After normalization, the estimates of the transition probability matrices, i.e., $\hat{P}^{(jk)}$, are obtained.

Regarding the λ_{jk} coefficients, the estimation method proposed by Ching, Fung, and Ng (2002) involves the following optimization problem:

$$\begin{gathered} min_{\lambda} max_i |[\sum_{k=1}^{m} \lambda_{jk} \hat{P}^{(jk)} \hat{\boldsymbol{x}}^{(k)} - \hat{\boldsymbol{x}}^{(j)}]| \\ \text{s.t. } \sum_{k=1}^{s} \lambda_{jk} \text{ and } \lambda_{jk} \geq 0 \end{gathered} \tag{3}$$

Besides this, different models have been proposed for multiple categorical data sequences. Kijima, Komoribayashi, and Suzuki (2002) proposed a parsimonious MMC model to simulate correlated credit risks. Siu et al. (2005) proposed a model easy to implement, however, its applicability was limited by the number of parameters involved. Ching, Ng, and Fung (2008) proposed a simplified model based on an assumption proposed in Zhang, King, and Hyndman (2006). Zhu and Ching (2010) proposed a method of estimation based on minimizing the prediction error with equality and inequality restrictions and Nicolau and Riedlinger (2014) proposed a new approach to estimate MMC which avoids imposing restrictions on the parameters, based on non-linear least squares estimation, facilitating the model estimation and the statistical inference. Lastly, Wang, Huang, and Ching (2014) proposed a new multivariate Markov chain model to reduce the number of parameters. Thus, generally, the models used in the published papers were developed by Ching, Fung, and Ng (2002) or were a consequent generalization of them and addressed the MMC as an end in itself.

In Damásio (2013), a different and innovative concept was proposed: the usage of MMC as regressors in a certain model. Hence, given that the MMC Granger causes a specific dependent variable, and taking advantage of the information about the past state interactions between the MMC categories, it was possible to forecast the current dependent variable more accurately.

In a vast majority of the studies in MMC models, it is assumed a positive correlation between the different data sequences due to the restrictions imposed. This means it is always considered that at moment t, an increase in a state probability for a data sequence has an increasing impact on another data sequence, for time $t+1$. Thereupon, if one has a negative correlation between series, the parameter estimates are forced to be zero. The solution to this problem is very straightforward, one can relax the assumptions and not assume the constraints. However, that means the results produced by the model will no longer be probabilities. Raftery and Tavaré (1994) presented an alternative to this, by dropping the positivity condition and imposing another set of restrictions. Ching, Ng, and Fung (2008) also tackled this issue and proposed a method where one splits the Q matrix into the sum of two other matrices and one represents the positive correlations and another the negative correlations. Also, in Nicolau (2014), a specification completely free from constraints, inspired by the MTD model, was proposed, facilitating the estimation procedure and, at the same time, providing a more accurate specification for $P_j(i_0|i_1,\ldots,i_s)$. The model was:

$$P_j(i_0|i_1,\ldots,i_s) = P_j^{\Phi}(i_0|i_1,\ldots,i_s) := \frac{\Phi(\eta_{j0}+\eta_{j1}P(i_0|i_1)+\cdots+\eta_{js}P(i_0|i_s))}{\sum_{k=1}^{m}\Phi(\eta_{j0}+\eta_{j1}P(k|i_1)+\cdots+\eta_{js}P(k|i_s))} \tag{4}$$

where $n_{ji} \in \mathbb{R}(j=1,\ldots,s; i=1,\ldots,m)$ and Φ is the (cumulative) standard normal distribution function. This specification is denoted as and MTD-Probit model. The log-likelihood is given by:

$$LL = \sum_{i_1,i_2,\ldots,i_{i_s},i_0} n_{i_1,i_2,\ldots,i_{i_s},i_0} log(P_j^{\Phi}(i_0|i_1,\ldots,i_s)) \quad (5)$$

and the maximum likelihood estimator is defined, as usual, as $\hat{\eta}$ = arg max$_{n_{j1},\ldots,n_{js}} LL.$ The parameters $P_{jk}(i_0|i_1)$, k =1, . . . , s can be estimated in advance, through the consistent and unbiased estimators proposed by Ching, Fung, and Ng (2002):

$$\widehat{P_{jk}(i_0|i_1)} = \frac{n_{i_1 i_0}}{\sum_{i_0=1}^{n} n_{i_1 i_0}} \quad (6)$$

This specification can be superior to the MTD because in the absence of constraints, the estimation procedure is easier and the standard numerical optimization routines can be easily applied. However, similarly to the standard MTD, the likelihood is not a strictly concave function on the entire parameter state-space, thus the choice of starting values is still important. Additionally, since the parameters are not constrained, the model describes a wider range of possible dependencies. Moreover, this proposed model is more accurate than the MTD model. For more details on this see Nicolau (2014).

Overall, the published work on MMC models was mostly based on improving the estimation methods and/or make the model more parsimonious. In Damásio (2013), a different approach was used and the work developed focused on the usage of MMC as regressors in a certain model. Particularly, it showed that a MMC can improve the forecast of a dependent variable. In a way, it demonstrated that a MMC can be an end in itself, but also it can be an instrument to reach an end or a purpose. In this work, the opposite will be developed: instead of considering a MMC as regressors, a model in which a vector with pre-determined exogenous variables are part of $\mathcal{F}_{t-1}$ is proposed.

2.2 Covariates in Markov Chain Models

Given the scope of this essay, it is relevant to study the previous work regarding the inclusion of covariates in Markov chains models. Regier (1968) proposed a two-state Markov chain model, where the probabilities of the transition matrix were a function of a parameter, q, that described the tendency of the subject to move from state to state. Kalbfleisch and Lawless (1985) proposed a method of analysis of panel data under a continuous-time Markov model, that could be generalized to handle covariate analysis and the fitting of certain non-homogeneous models. This work overcame the limitations of Bartholomew (1968), Spilerman and Singer (1976) and Wasserman (1980) methodologies, by developing a new algorithm that provided a very efficient way of obtaining maximum likelihood estimates. Also, Muenz and Rubinstein (1985) developed a Markov model for covariates dependence of binary sequences, where the transitions probabilities were estimated through two logistic regressions that depended on a set of covariates. Essentially, Muenz and Rubinstein (1985) modeled a non-homogeneous Markov chain through logistic regression, considering only two states. Islam, Arabia, and Chowdhury (2004) developed an extension of this model considering three states, and Islam and Chowdhury (2006) generalized this approach for HOMC. Additionally, Azzalini (1994) proposed a model to study the influence of time-dependent covariates on the marginal distribution of a binary response in serially correlated binary data, where Markov chains are expressed in terms of transitional probabilities.

More recently, Bolano (2020) proposed a MTD-based approach to handle categorical covariates, that considers each covariate separately and combines the effects of the lags of the MTD and the covariates employing a mixture model. Specifically, the model is given by:

$$P(X_t = k \mid X_{t-1} = i, C_1 = c_1, \dots, C_l = c_l) \approx \theta_0 a_{ik} + \sum_{h=1}^{l} \theta_h d_{c_h k} \tag{7}$$

where a_{ik} is the transition probability from state i to state k, as in a conventional Markov chains and $d_{c_h k}$ is the probability of observing the states k given the modality c_h of the covariate h. Lastly, $\theta_0, \dots, \theta_l$ are the weights of the explanatory elements of the model.

According to the literature presented, several researchers have proposed methodologies or generalizations to include covariates in Markov chain models. Mostly for social sciences and health applications, where the transition probabilities were generally modeled through logistic regression. However, there has been an increased focus on categorical covariates, opposing continuous covariates and a lack of approaches to multivariate Markov chain models. Thus, with this work, we aim to tackle this research gap.

3 THEORETICAL FRAMEWORK

3.1 Homogeneous Markov Chains

A Markov chain is a model used to represent dependencies between successive observations of a random variable. It concerns a sequence of random variables, which corresponds to the states of a certain system, in such a way that the state at one time period depends only on the state of the previous period. Specifically, for a first-order Markov Chain, the present observation at time t is conditionally independent of those up to and including time $(t-2)$ given the immediate past [time $(t-1)$]. Mathematically, this is given by:

$$P(X_t = i_0 \mid \mathcal{F}_{t-1}) = P(X_t = i_0 | X_{t-1}) \tag{8}$$

where $\mathcal{F}_{t-1}$ is the σ - algebra generated by the available information until $t-1$ and X_t is a random variable taking values in the countable set $E = \{1, \ldots, m\}$. Thus, it is possible to write:

$$P(X_t = i_0 | X_0 = i_0, \ldots, X_{t-1} = i_1) = P(X_t = i_0 | X_{t-1} = i_1) = q_{i_1 i_0} \tag{9}$$

where $i_t, \ldots, i_0 \in E$. If $q_{i_1 i_0}$ does not depend on time t, than the the Markov chain is homogeneous.

Considering all combinations of i_1 and i_0, the Transition Probability Matrix (TPM):

$$\begin{bmatrix} P(X_t = 1 | X_{t-1} = 1) & \ldots & P(X_t = m | X_{t-1} = 1) \\ \vdots & \ddots & \vdots \\ P(X_t = 1 | X_{t-1} = m) & \ldots & P(X_t = m | X_{t-1} = m) \end{bmatrix} = \begin{bmatrix} q_{11} & \ldots & q_{1m} \\ \vdots & \ddots & \vdots \\ q_{m1} & \ldots & q_{mm} \end{bmatrix} \tag{10}$$

The i-th row of the transition probability matrix, for i = 0, 1,..., is the probability distribution of the values of X_{t+1} under the condition that $X_t = i$. If the number of states is finite, then the transition probability matrix is a finite square matrix whose order (the number of rows) corresponds to the number of states. A Markov process is completely defined by the transition probability matrix and the initial state X_0 (or, more generally, the probability distribution of X_0).

It is relevant to assess the events' long-term probability. To do so, it is required to assume that all states communicate with each other and are aperiodic, which implies that all states are non-null persistent and, therefore, the chain is ergodic.

Proposition (1)**.** *If X is an aperiodic positive recurrent* [1]*Markov chain with finite state-space, then the row vector of stationary probabilities is $\pi = [\pi_1, \ldots, \pi_m]$, where $\pi_i > 0$ and it satisfies:*

$$\sum_{i=1}^{m} \pi_i = 1 \pi \mathbf{P} = \pi$$

where $\mathbf{P}$ is the transition probability matrix.

[1]A state i is recurrent if and only if, after the process starts from state i, the probability of its returning to state i after some finite length of time is one. (Taylor and Karlin (1984)).

3.2 Non-homogeneous Markov Chains

In non-homogeneous Markov chain, the Markov property maintains, however the transitions probabilities depend on time, such that:

$$P(X_t = i_0|X_0 = i_0, \ldots, X_{t-1} = i_1) = P(X_t = i_0|X_{t-1} = i_1) = q_{i_1 i_0}(t) \tag{11}$$

where $i_t, \ldots, i_0 \in E$ and the quantity $q_{i_1 i_0}(t)$ depends on time. In this case, the matrix

$$P(t) = q_{i_1 i_0}(t) \tag{12}$$

is the transition matrix at time t.

Theorem 3.1. *A non-homogeneous finite Markov chain for which*

a) there exists a sequence of stochastic matrices S_1, S_2, $\ldots$, with identical rows such that

   1) $\sum(S_j P_{j+1} - S_{j+1})$ converges absolutely

   2) The sequence S_j has a limit S

b) $\lim\limits_{n\to\infty} \prod_{j=1}^{n}(1 - {}_jp_{min}) = 0$

is strongly ergodic, and the limiting matrix is S.

See proof in Hajnal and Bartlett (1956), page 69.

Here ${}_jp_{min}$ denotes the smallest element of the j-th transition matrix P_j. Condition (a1) means that the sums of corresponding elements in the matrices $(S_1P_2 - S_2)$, $(S_2P_3 - S_3)$, $\ldots$ converges absolutely. For weak ergodicity, only condition (b) from the Theorem 3.1 needs to hold.

3.3 The Mixture Transition Distribution Model and the First-Order Multivariate Markov Chain Model

The MTD model was introduced by Raftery (1985) for the modeling of HOMC in discrete time. It is defined as

$$P(X_t = i_0|X_0 = i_0, \ldots, X_{t-l} = i_l) = \sum_{g=1}^{l} P(X_t = i_0|X_{t-g} = i_g) = \sum_{g=1}^{l} \lambda_g q_{i_g i_0} \tag{13}$$

where $i_0, \ldots, i_l \in \{1, \ldots, m\}$, $q_{i_g i_0}$ are the probabilities of an $m \times m$ transition matrix and $\lambda = \{\lambda_1, \ldots, \lambda_g\}$ is a vector of lag parameters subject to the following constraints:

$$\sum_{g=1}^{l} \lambda_g = 1 \tag{14}$$

$$\lambda_g \geq 0 \tag{15}$$

This model has only $m(m-1) + (l-1)$ independent parameters and each additional lag adds only one additional parameter. Comparing to the corresponding Markov chain, the MTD model is more parsimonious. The equilibrium distribution of the MTD model is the same as the first-order Markov chain with probability transition matrix $\mathbf{P}$, regardless its order.

The parameters can be estimated through maximum-likelihood, the log-likelihood function is given by:

$$LL = \sum_{i_l,\ldots,i_0}^{m} n_{i_l,\ldots,i_0} log(\sum_{g=1}^{l} \lambda_g q_{i_g i_0}) \tag{16}$$

where $n_{i_l,\ldots,i_0}$ is the number of sequences of the form: $X_{t-l} = i_l, \ldots, X_t = i_0$ in the data.

The log-likelihood ought to be optimized concerning the constraints (4) and (5), to ensure that the model describes a HOMC. Additionally, these constraints are a sufficient but not necessary condition to secure that the probability terms are non-negative and less than one.

Since the introduction of this model, several estimation methods have been proposed. For example, Raftery and Tavaré (1994) showed that the MTD model can be estimated through an iterative procedure in Generalized Linear Interactive Modeling (GLIM), when the number of values taken by the random variable X_t is $m = 2$. Berchtold and Raftery (2002) proposed an iterative algorithm for numerical maximization of the log-likelihood, Lèbre and Bourguignon (2008) proposed an Expectation-Maximization algorithm and Chen and Lio (2009) proposed an approach of MLE, converting the nonlinear embedded constraints into box constraints.

As described in Section 2.1, the MTD model faces drawbacks, which Ching, Fung, and Ng (2002) overcame. Hence, the generalization presented in this work is based on Ching's First-Order Multivariate Markov Chain model. Withal, Ching, Fung, and Ng (2002) estimation procedure does not address the statistical inference problem. Therefore, it might be useful to write Ching's MMC model using Raftery's HOMC model notation and deduce the log-likelihood function.

Considering the multivariate stochastic process $\{(S_{1t}, \ldots, S_{st} =; t = 1, 2, \ldots\}$ where S_{jt} $(j = 1, 2, \ldots, s)$ can take values from the set $\{1, 2, \ldots, m\}$, the model is:

$$P(S_{jt} = k|S_{1,t-1} = i_1, \ldots, S_{s,t-1} = i_s)^{MTD} = \lambda_{j1} P(S_{jt} = k|S_{1,t-1} = i_1) + \cdots + \lambda_{js} P(S_{jt} = k|S_{s,t-1} = i_s) \quad (17)$$

subject to the usual constraints.

The log-likelihood is:

$$LL = \sum_{i_1, i_2, \ldots, i_{i_s}, i_0} n_{i_1, i_2, \ldots, i_{i_s}, i_0} log(P_j^{MTD}(i_0|i_1, \ldots, i_s)) \quad (18)$$

where $n_{i_1, i_2, \ldots, i_{i_s}, i_0}$ is the number s the number of sequences of the form: X_{t-l} = i_l, ..., X_t = i_0 in the data.

The transition probabilities matrices can be estimated as proposed in Ching, Fung, and Ng (2002), however the λ_{jk} coefficients are estimated through Maximum Likelihood Estimator (MLE), addressing the statistical inference problem. In the following section, we present the generalization proposed, based on this model.

4 MULTIVARIATE MARKOV CHAINS WITH COVARIATES

4.1 Theoretical model

In this work, a new generalization of Ching, Fung, and Ng (2002) MMC model is presented: the Generalized Multivariate Markov Chain (GMMC) model, that is, we will consider exogeneous or pre-determined covariates in the σ - algebra generated by the available information until $t-1$ ($\mathcal{F}_{t-1}$). These variables can be deterministic or stochastic and do not need necessarily to be reported at time t. Broadly, the model is given by:

$$P(S_{jt} = k|\mathcal{F}_{t-1}) = P(S_{jt} = k|S_{1t-1} = i_1, S_{2t-1} = i_2, \ldots, S_{st-1} = i_s, \boldsymbol{x}_t) \quad (19)$$

As presented in the previous section, we can specify this model as proposed by Ching, Fung, and Ng (2002) with Raftery's notation:

$$P(S_{jt} = i_0|S_{1t-1} = i_1, \ldots, S_{st-1} = i_s, \boldsymbol{x}_t) \equiv \lambda_{j1} P(S_{jt} = i_0|S_{1t-1} = i_1, \boldsymbol{x}_t) + \cdots + \lambda_{js} P(S_{jt} = i_0|S_{st-1} = i_s, \boldsymbol{x}_t) \quad (20)$$

subject to the usual constraints.

4.2 Estimation and Inference

Similarly to the standard MTD model, this proposed model is estimated through MLE. The log-likelihood is given by:

$$LL = \sum_{t=1}^{n} logP(S_{jt} = i_0 | S_{1t-1} = i_1, \ldots, S_{st-1} = i_s, \boldsymbol{x}_t) \tag{21}$$

Additionally, the probabilities can be estimated through an multinomial logit model. This is a common approach to estimate non-homogeneous Markov chains (Rajarshi (2013)). The consistency of the MLE estimator is held by the following Proposition:

Proposition (2). *Let $\{\boldsymbol{w}_t\}$ be ergodic stationary random variable, with likelihood function $f(\boldsymbol{w}_t|\theta_0)$. Let $\hat{\theta}$ be the MLE estimator. Suppose that:*

(i) $E[log\ f(\boldsymbol{w}_t \mid \theta_0)]$ *is uniquely maximized on* Θ *at* $\theta_0 \in \Theta$*,*

(ii) $\theta_0 \in \Theta$*, which is compact,*

(iii) $log\ f(\boldsymbol{w}_t \mid \theta_0)$ *is continuous at each* $\theta \in \Theta$ *with probability one,*

(iv) $E[sup_{\theta \in \Theta} \mid log\ f(w_t \mid \theta) \mid] < \infty$

Then $\hat{\theta} \xrightarrow{p} \theta_0$

Condition (i) is verified according to Lemma 2.2 of Newey and Mcfadden (1994). Condition (ii) is verified and guaranteed by the restrictions imposed in the model parameters. Knowing that $P(S_{jt} = i_0 | S_{1t-1} = i_1, \ldots, S_{st-1} = i_s, \boldsymbol{x}_t)$ is linear combination of a set of n probabilities and since the logarithm function is a continuous function, condition (iii) is verified. Finally, condition (iv) is verified according to Lemma 2.4 of Newey and Mcfadden (1994).

Regarding inference, MLE will be asymptotically normal if it is consistent and the following Proposition verifies:

Proposition (3). *Let $\{\boldsymbol{w}_t\}$ be ergodic stationary random variable and let $s(\boldsymbol{w}_t;\theta)$ and $H(\boldsymbol{w}_t;\theta)$ be the first and second partial derivatives of the log $f(\boldsymbol{w}_t \mid \theta)$, respectively. Suppose the estimator $\hat{\theta}$ is consistent and suppose, further, that*

(i) θ_0 *is in the interior of* Θ*,*

(ii) *log* $f(\boldsymbol{w}_t \mid \theta_0)$ *is twice continuously differentiable in* θ *for any* $\boldsymbol{w}_t$*,*

(iii) $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} s(\boldsymbol{w}_t; \theta_0) \underset{d}{\rightarrow} N(0, \Sigma)$*, where* Σ *is positive definite,*

(iv) *For some neighborhood of* $\mathcal{N}$ *of* θ_0*,*

$$E[\underset{\theta \in \mathcal{N}}{sup} ||H(\boldsymbol{w}_t; \theta)||] < \infty$$

so that for any consistent estimator $\tilde{\theta}$*,* $\frac{1}{n} \sum_{t=1}^{n} H(\boldsymbol{w}_t; \tilde{\theta}) \underset{p}{\rightarrow} E[H(\boldsymbol{w}_t; \theta)]$

(v) $E[H(\boldsymbol{w}_t, \theta_0)]$ *is nonsingular.*

Then $\hat{\theta}$ *is asymptotically normal with*

$$Avar(\hat{\theta}) = \{E[H(\boldsymbol{w}_t; \theta_0)]\}^{-1} \Sigma \{E[H(\boldsymbol{w}_t; \theta_0)]\}^{-1}$$

Condition (i) is verified as in condition (ii) of Proposition (2). Condition (ii) is also verified as in condition (iii) of Proposition (2), since the logarithm function is twice continuously differentiable. Condition (iii) is verified according to the Ergodic Stationary Martingale Differences CLT (Billingsley (1961)). In this case, Σ = $E[s(\boldsymbol{w}_t; \theta_0) s(\boldsymbol{w}_t; \theta_0)']$ = $-E[H(\boldsymbol{w}_t; \theta_0)]$, which implies that $Avar(\hat{\theta}) = -\{E[H(\boldsymbol{w}_t; \theta_0)]\}^{-1}$. Condition (iv) is verified according to Lemma 2.4 of Newey and Mcfadden (1994). Considering only one equation, let $\boldsymbol{q}_t$ be a $t \times s$ matrix of the probabilities $P(S_{1t} \mid S_{1t}, x_t), \ldots, P(S_{1t} \mid S_{st}, x_t)$ and $\boldsymbol{\lambda}$ a row-vector of $\lambda_{11}, \ldots, \lambda_{1s}$, the hessian matrix is given by $E[\boldsymbol{q}_t' \boldsymbol{q}_t [(\boldsymbol{\lambda} \boldsymbol{q}_t')(\boldsymbol{\lambda} \boldsymbol{q}_t')']^{-1}]$. Condition (v) is verified if $E[\boldsymbol{q}_t' \boldsymbol{q}_t]$ is nonsingular.

4.3 Implementation in R

As it was shown in the previous section, several studies and generalizations of the MMC models were made. However, the availability of packages that allow the estimation and application of these models is scarce and most of these methods use algorithms and software that are not broadly available or can only be applied in special situations.

In the last few years, **R** software has been gaining importance in the field of statistical computing. This might be because it is free and open-source software, which compiles and runs on a wide variety of operating systems.

Specifically, in **R** software, there are some available packages related to Markov chains and Multivariate Markov chains. For example, the package `march` (Maitre and Emery (2020)) allows the computation of various Markovian models for categorical data including homogeneous Markov chains of any order, MTD models, Hidden Markov models, and Double Chain Markov Models. This package was developed by Ogier Maitre, with contributions from Andre Berchtold, Kevin Emery, and Oliver Buschor and it is maintained by Andre Berchtold. All the models computed by this package are for univariate categorical data. The package `markovchain` (Spedicato (2017)) contains functions and methods to create and manage discrete-time Markov chains. In addition, it includes functions to perform statistical and probabilistic analysis (analysis of their structural proprieties). Finally, the package `DTMCPack` (Nicholson (2013)) contains a series of functions that aid in both simulating and determining the properties of finite, discrete-time, discrete-state Markov chains. There are two main functions: `DTMC` and `MultDTMC`, which produce n iterations of a Markov Chain(s) based on transition probabilities and an initial distribution given by the user, for the univariate and multivariate case, respectively. This last package is the only one available in **R** for MMC.

The main goal is not only the development of the Generalized Multivariate Markov chain (GMMC) models, but also the implementation of these methods in an **R** package.

The R package will include three functions: `multimtd`, `multimtd_probit` and `mmcx`. The first two functions estimate the MTD model for multivariate categorical data, with Chings's specification (Ching, Fung, and Ng (2002)) and with the Probit specification (Nicolau (2014))), respectively. The last function allows the estimation of our proposed model, the Generalized Multivariate Markov Chain (GMMC) model presented in the previous section.

Regarding the estimation methods for each function, for the `multimtd` the estimation method was presented in Berchtold (2001) applied to the multivariate case. For the `multimtd_probit`, a package for numerical maximization of the log-likelihood, `maxLik` (Henningsen and Toomet (2011)), was used. This package performs Maximum Likelihood estimation through different optimization methods that can be chosen by the user. The optimization methods available are Newton-Raphson, Broyden - Fletcher - Goldfarb - Shanno, BFGS algorithm, Berndt - Hall - Hall - Hausman, Simulated ANNealing, Conjugate Gradients, and Nelder-Mead. Finally, for the `mmcx` function a different approach was used. Unlike the MTD-Probit, the model proposed has equality and inequality restrictions in the parameters. The `maxLik` package only allows one type of restriction for each Maximum Likelihood estimation, so it was not possible to use this package to estimate the proposed model with exogenous variables. Hence, the algorithm used was the Augmented Lagrangian method, available in package `alabama` through the function `auglag`. This estimation method for the proposed model is not very common, however, it has been applied to Markov chain models (Rajarshi (2013)). Regarding the

GMMC model's probabilities, these were estimated through a Multinomial Logit using `rmultinom` of the package `neet` (Venables and Ripley (2002)).

Additionally, the hessian matrices were also computed, which allowed performing statistical inference. The `maxLik` and `auglag` compute the Hessian matrices with the estimates. For the function `multimtd`, since the optimization procedure of Berchtold (2001) was used, the hessian was computed through the second partial derivatives.

The **R** package, `GenMarkov()`, with these three functions is available in Comprehensive R Archive Network (CRAN) (`https://CRAN.R-project.org/package=GenMarkov`).

5 MONTE CARLO SIMULATION STUDY

A Monte Carlo simulation study was designed to evaluate the dimension and power of the test of the parameters of the model proposed. The **R** statistical environment was used for all computations. This simulation study was comprised of two parts.

5.1 Part I: Detect a non-homogeneous Markov chain

First, we considered two sequences with two and three states. The main goal was to assess if the model detected correctly the presence of a non-homogeneous Markov chain and if the estimate of the parameter would correspond to the expected. So, given two sequences, one generated through a non-homogeneous Markov chain and the other generated through a homogeneous Markov chain, it would be expected that the parameter associated with the transition probabilities of the first sequence would be one and the parameter associated with the transition probabilities of the second sequence would be zero. With this in mind, the transitions probabilities of the first sequence were estimated through a logistic regression, where parameters of this regression were randomly generated in **R**, and the second sequence was generated through a first-order Markov chain.

The Data Generating Process (DGP) for each series for two states was:

Series	DGP
S_{1t}	$\begin{bmatrix} \frac{1}{(1+e^{0.83+0.97x_{t-1}})} & \frac{e^{0.83+0.97x_{t-1}}}{(1+e^{0.83+0.97x_{t-1}})} \\ \frac{1}{(1+e^{0.38+0.77x_{t-1}})} & \frac{e^{0.38+0.77x_{t-1}}}{(1+e^{0.38+0.77x_{t-1}})} \end{bmatrix}$
S_{2t}	$\begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$

For three states, the series S_{1t} was generated through the probability transition matrix:

$$\begin{bmatrix} \frac{1}{(1+e^{(0.92+0.84x_{t-1})+(0.75+0.75x_{t-1})})} & \frac{e^{0.92+0.84x_{t-1}}}{(1+e^{(0.92+0.84x_{t-1})+(0.75+0.75x_{t-1})})} & \frac{e^{0.75+0.75x_{t-1}}}{(1+e^{(0.92+0.84x_{t-1})+(0.75+0.75x_{t-1})})} \\ \frac{1}{(1+e^{(0.06+0.92x_{t-1})+(0.70+0.04x_{t-1})})} & \frac{e^{0.06+0.92x_{t-1}}}{(1+e^{(0.06+0.92x_{t-1})+(0.70+0.04x_{t-1})})} & \frac{e^{0.70+0.04x_{t-1}}}{(1+e^{(0.06+0.92x_{t-1})+(0.70+0.04x_{t-1})})} \\ \frac{1}{(1+e^{(0.09+0.75x_{t-1})+(0.57+0.45x_{t-1})})} & \frac{e^{0.09+0.75x_{t-1}}}{(1+e^{(0.09+0.75x_{t-1})+(0.57+0.45x_{t-1})})} & \frac{e^{0.57+0.45x_{t-1}}}{(1+e^{(0.09+0.75x_{t-1})+(0.57+0.45x_{t-1})})} \end{bmatrix}$$

And the probability transition matrix of the series S_{2t} was given by:

$$\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.5 & 0.2 & 0.3 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

Hence, for both states cases considered, it was expected that the estimated regression would be:

$$P(S_{1t} = i_0 | S_{1t-1} = i_1, S_{2t-1} = i_2, \boldsymbol{x}_{t-1}) = \\ 1 \times P(S_{1t} = i_0 | S_{1t-1} = i_1, \boldsymbol{x}_{t-1}) + 0 \times P(S_{1t} = i_0 | S_{2t-1} = i_2, \boldsymbol{x}_{t-1}) \quad (22)$$

To assess the power and dimension of test, we used the Wald test with the following hypothesis:

	Hypothesis	Test
Power	$H_0 : \lambda_{11} = 0$	$\frac{\hat{\lambda}_{11}^2}{se(\hat{\lambda}_{11})^2} \sim \chi^2_{(1)}$
	$H_0 : \lambda_{12} = 1$	$\frac{(\hat{\lambda}_{12}-1)^2}{se(\hat{\lambda}_{12})^2} \sim \chi^2_{(1)}$
Dimension	$H_0 : \lambda_{11} = 1$	$\frac{(\hat{\lambda}_{11}-1)^2}{se(\hat{\lambda}_{11})^2} \sim \chi^2_{(1)}$
	$H_0 : \lambda_{12} = 0$	$\frac{\hat{\lambda}_{12}^2}{se(\hat{\lambda}_{12})^2} \sim \chi^2_{(1)}$

The simulation procedure was performed as follows:

1. Generate randomly the values of the coefficients for the probability transition matrix of series S_{1t};

2. Generate randomly the probability transition matrix of series S_{2t};

3. Set the initial value of S_{2t} to 1 and simulate the following from the defined probability transition matrix;

4. In each iteration (of 1000 repetitions),

   - Generate $X_t \sim N(2, 25)$;
   - Generate the time-varying probabilities of series S_{1t} through the values of the fixed coefficients and the lagged variable x_t;
   - Set the initial values of the series S_{1t} as 1;
   - For each period t, simulate the next state of S_{1t} from the probabilities simulated for that moment;
   - Estimate the model through the function `mmcx()`;
   - Calculate the Wald test and add to the counter if it is rejected.

Surely, only the first equation of the output was evaluated. The results are available in Tables I and II.

Parameter	Sample size	Power (%)	95% CI	Dimension (%)	95% CI
1	100	8.2	[6.6; 10.1]	5.7	[4.4; 7.4]
	500	25.2	[22.6; 28.0]	7.6	[6; 9.4]
	1000	46.6	[43.5; 49.7]	5.8	[4.5; 7.5]
	5000	99.4	[98.6; 99.8]	6	[4.6; 7.7]
0	100	8.2	[6.6;10.1]	5.7	[4.4; 7.4]
	500	25.2	[22.6; 28.0]	7.6	[6; 9.4]
	1000	46.6	[43.5; 49.7]	5.8	[4.5; 7.5]
	5000	99.4	[98.6; 99.8]	6	[4.6; 7.7]

TABLE I: Simulation study results: Two-states

Considering two states, the dimension of test was at 5.7% with a sample size of 100 observations, sightly increased with 500 observations and returned to the expected values in 1000 and 5000 observations. For a sample size of 100, 500, and 1000 observations, we have low power of test. So, when considering two states, it is necessary that the sample has at least 5000 observations, or, if that is not possible, consider a higher significance level, when testing for individual significance.

Parameter	Sample size	Power (%)	95% CI	Dimension (%)	95% CI
1	100	90.5	[88.5; 92.2]	9.7	[8; 11.7]
	500	100	[99.5; 100]	0.2	[0.03; 0.8]
	1000	100	[99.5; 100]	0.3	[0.08; 0.95]
0	100	90.5	[88.5; 92.2]	9.7	[8; 11.7]
	500	100	[99.5; 100]	0.2	[0.03; 0.8]
	1000	100	[99.5; 100]	0.3	[0.08; 0.95]

TABLE II: Simulation study results: Three-states

Considering three states, the dimension of test was 9.7% for a sample size of 100 observations, 0.2% for a sample size of 500 observations, and 0.3% for a sample size of 1000. Regarding the power of test, we see similar behavior, for a sample of 100 observations, the power of test was 90.5% and from a sample of 500 observations, we reach a power of test of 100%. Thus, when considering three states, one may consider a sample of 500 observations without compromising the power and dimension of test.

5.2 Part II: Detect parameters assigned values

Secondly, we performed a simulation study where we considered two non-homogeneous Markov chains with two states. Here, the main goal was to assess if the model detected correctly the parameters assigned. So, in this case, we started by generating the terms of the model proposed. These terms were estimated through logistic regression, and the parameters of this regression were randomly generated in R. The DGP were:

Model term	DGP
$P(S_{1t}\|S_{1t-1},x_{t-1})$	$\begin{bmatrix} \frac{1}{(1+e^{0.31+0.73x_{t-1}})} & \frac{e^{0.31+0.73x_{t-1}}}{(1+e^{0.31+0.73x_{t-1}})} \\ \frac{1}{(1+e^{0.73+0.54x_{t-1}})} & \frac{e^{0.73+0.54x_{t-1}}}{(1+e^{0.73+0.54x_{t-1}})} \end{bmatrix}$
$P(S_{1t}\|S_{2t-1},x_{t-1})$	$\begin{bmatrix} \frac{1}{(1+e^{0.35+0.05x_{t-1}})} & \frac{e^{0.35+0.05x_{t-1}}}{(1+e^{0.35+0.05x_{t-1}})} \\ \frac{1}{(1+e^{0.85+0.03x_{t-1}})} & \frac{e^{0.85+0.03x_{t-1}}}{(1+e^{0.85+0.03x_{t-1}})} \end{bmatrix}$
$P(S_{2t}\|S_{2t-1},x_{t-1})$	$\begin{bmatrix} \frac{1}{(1+e^{0.32+0.93x_{t-1}})} & \frac{e^{0.32+0.93x_{t-1}}}{(1+e^{0.32+0.93x_{t-1}})} \\ \frac{1}{(1+e^{0.1+0.96x_{t-1}})} & \frac{e^{0.1+0.96x_{t-1}}}{(1+e^{0.1+0.96x_{t-1}})} \end{bmatrix}$
$P(S_{2t}\|S_{1t-1},x_{t-1})$	$\begin{bmatrix} \frac{1}{(1+e^{0.29+0.42x_{t-1}})} & \frac{e^{0.29+0.42x_{t-1}}}{(1+e^{0.29+0.42x_{t-1}})} \\ \frac{1}{(1+e^{0.07+0.32x_{t-1}})} & \frac{e^{0.07+0.32x_{t-1}}}{(1+e^{0.07+0.32x_{t-1}})} \end{bmatrix}$

Similarly to Part I, we considered a Wald test to assess the power and dimension of the test. The simulation procedure was performed as follows:

1. Generate randomly the coefficients values to calculate the probability transition matrices;
2. In each iteration (of 1000 repetitions),

- Generate $X_t \sim N(2, 25)$;
- Generate the probabilities $P(S_{jt}|S_{st-1}, x_{t-1})$, with $j = 1, 2$ and $s = 1, 2$.
- Set the initial values of the series S_{1t} and S_{2t} as 1;
- For each period t, calculate the probabilities $P(S_{1t}|S_{1t-1}, S_{2t-1}, x_{t-1})$ and $P(S_{2t}|S_{1t-1}, S_{2t-1}, x_{t-1})$ through the assigned values of the λ's. Considering the calculated probabilities, simulate the next state for each series, S_{1t} and S_{2t}.
- Estimate the model through the function `mmcx()`;
- Calculate the Wald test and add to the counter if it is rejected.

The probabilities $P(S_{1t}|S_{1t-1}, x_{t-1})$ and $P(S_{1t}|S_{2t-1}, x_{t-1})$ presented some differences regarding its values' distributions. Specifically, $P(S_{1t}|S_{1t-1}, x_{t-1})$ had more extreme probabilities values, with the minimum value being close to 0 and the maximum value being close to 1. And, the probabilities $P(S_{1t}|S_{2t-1}, x_{t-1})$ had more moderate values, with the minimum value being, on average, 0.3 and the maximum value, 0.7. When the probabilities have values close to 1, one says that the states/regimes are persistent.

We calculated the power and dimension of test for each value of λ when the estimated probabilities are moderate and when they are extreme. Hence, considering equation 1:

$$P(S_{1t} = i_0|S_{1t-1} = i_1, \ldots, S_{2t-1} = i_2, \boldsymbol{x}_{t-1}) = \lambda_{11} P(S_{1t} = i_0|S_{1t-1} = i_1, \boldsymbol{x}_{t-1}) + \lambda_{2s} P(S_{1t} = i_0|S_{2t-1} = i_s, \boldsymbol{x}_{t-1}) \quad (23)$$

The parameter λ_{11} will be associated with more extreme probabilities and λ_{12} will be associated with more moderate probabilities.

Value	Sample size	Power (%)	95% CI	Dimension (%)	95% CI	Parameter
0.2	100	7.3	[5.8; 9.1]	1.8	[1.1; 2.9]	λ_{11}
	500	6.5	[5.1; 8.3]	1.4	[0.8; 2.4]	
	1000	4.6	[3.4; 6.1]	0.9	[0.4; 1.8]	
	5000	1.9	[1.2; 3]	0	[0.0; 0.5]	
0.4	100	9.2	[7.5; 11.2]	0.5	[0.2; 1.2]	λ_{11}
	500	9.6	[7.9; 11.6]	0.4	[0.1; 1.1]	
	1000	9.2	[7.5; 11.2]	0.2	[0.03; 0.8]	
	5000	9.7	[8; 11.7]	0.2	[0.03; 0.8]	
0.6	100	12.6	[10.6; 14.9]	0.5	[0.2; 1.2]	λ_{12}
	500	28.2	[25.5; 31.1]	0.4	[0.1; 1.1]	
	1000	26.1	[23.4; 29]	0.2	[0.03; 0.8]	
	5000	27.6	[24.9; 30.5]	0.2	[0.03; 0.8]	
0.8	100	13.9	[11.8; 16.2]	1.8	[1.1; 2.9]	λ_{12}
	500	43.5	[40.4; 46.6]	1.4	[0.8; 2.4]	
	1000	69.5	[66.5; 72.3]	0.9	[0.4; 1.8]	
	5000	99.9	[99.4; 99.9]	0	[0.0; 0.5]	
0.2	100	5.7	[4.3; 7.4]	1.8	[1.1; 2.9]	λ_{12}
	500	7.6	[6.1; 9.5]	2.5	[1.7; 3.7]	
	1000	15	[12.9; 17.4]	3.8	[2.7; 5.2]	
	5000	25.6	[17.1; 37.9]	6.4	[4.9; 8.1]	
0.4	100	8.5	[6.9; 10.4]	0.2	[0.03; 0.8]	λ_{12}
	500	14	[11.9; 16.3]	0.3	[0.08; 0.1]	
	1000	20.9	[18.4; 23.6]	7	[5.5; 8.8]	
	5000	71.5	[68.6; 74.3]	10.3	[8.5; 12.4]	
0.6	100	7.1	[5.6; 8.9]	0.2	[0.03; 0.8]	λ_{11}
	500	8.7	[7.1; 10.7]	0.3	[0.08; 0.1]	
	1000	14.2	[12.1; 16.6]	7	[5.5; 8.8]	
	5000	31.5	[28.6; 34.5]	10.3	[8.5; 12.4]	
0.8	100	5.3	[4; 6.9]	1.8	[1.1; 2.9]	λ_{11}
	500	10.3	[8.5; 12.4]	2.5	[1.7; 3.7]	
	1000	36.2	[33.2; 39.3]	3.8	[2.7; 5.2]	
	5000	59.9	[56.8; 62.9]	6.4	[4.9; 8.1]	

TABLE III: Simulation study results: Part II

By analyzing Table III, we can see that, when the states are persistent and the value of the parameter is low (i.e, 0.2 and 0.4), we have low power of test. By increasing this value, the power of test increases as well. When the states are not persistent, we do not have a clear pattern regarding the power of test, for a value of the parameter of 0.2, the power of test is still low (although not as low as the first scenario), increases when we have a value of 0.4, decreases when the value is 0.6 and increases again when the value is 0.8. Overall, the estimated standard errors seem high, which leads to low power of test.

Regarding the dimension of test, when we have higher weight associated with the non-persistent states, the dimension of test converges to 0. However, when this weight is associated with the persistent states, the dimension of test increases with the sample size, reaching a value of 10% in some cases. Hence, in this situation, one must use a 10% significance level to perform statistical inference on the parameters.

6 ILLUSTRATION

Markov chain models are used in interdisciplinary areas, such as economics, business, biology, and engineering, with applications to predict long-term behavior from traffic flow to stock market movements, among others. Modeling and predicting stock markets returns is particularly relevant for investors and policy makers. Since the stock market is a volatile environment, and the returns are difficult to predict, estimating the set of probabilities that describe these movements, might provide relevant input. Additionally, incorporating the effect of key macroeconomic variables could provide a more accurate picture of this specific environment. The following empirical illustration aims to model stock returns of two indexes as a function of the interest rate spread, specifically the 10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity.

The interest rate spread is a key macroeconomic variable and provides valuable information regarding the economy state. Specifically, it has been used to forecast recessions as in Estrella and Mishkin (1996), Dombrosky and Haubrich (1996), Chauvet and Senyuz (2016), Tian and Shen (2019) and McMillan (2021). Generically, when the economy is in expansion, short-term yields are lower than long-term yields On the other hand, when the economy is in recession, short-term yields are higher than long-term yields. The difference between these yields (or more specifically, the slope of the yield curve) can be use to forecast the state of the economy. Hence, this indicator might provide relevant input for investors.

We considered the 5-week-day daily stock returns ($r_t = 100 \times log(P_t/P_{t-1})$, where P_t is the adjusted close price) of two indexes, S&P500 and Dow Jones Industrial Average (DJIA), from November 11^{th} 2011 to September 1^{st} 2021 (2581 observations). Additionally, we considered the interest rate spread ($spread_t$), the 10-Year Treasury Constant

Maturity Minus 3-Month Treasury Constant Maturity. The data was retrieved from Federal Reserve Economic Data (FRED). Below in table IV, we have the descriptive statistics of these variables.

Variable	Minimum	1st Qu.	Median	Mean	3rd Qu.	Maximum
$spread_t$	-0.520	0.920	1.540	1.454	2.030	2.970
$r_{t;SP500}$	-12.765	-0.3198	0.070	0.054	0.518	8.968
$r_{t;DJIA}$	-13.841	-0.327	0.071	0.046	0.508	10.764

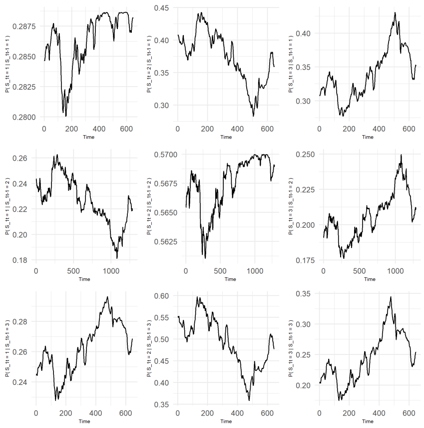
TABLE IV: Descriptive statistics

Moreover, to apply the model proposed, it is necessary to have a categorical time series, thus we applied the following procedure:

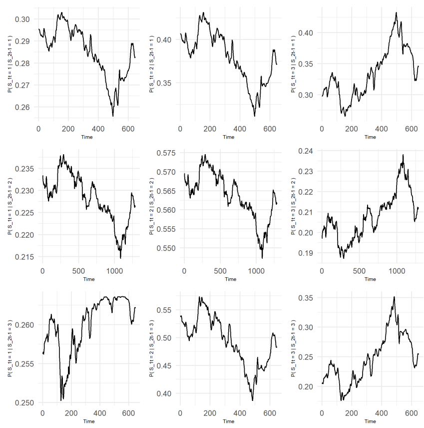
$$S_{st} = \begin{cases} 1, r_t \leq \hat{q}_{s;0.25} \\ 2, \hat{q}_{s;0.25} < r_t < \hat{q}_{s;0.75} \\ 3, r_t \geq \hat{q}_{s;0.75} \end{cases}$$

where $\hat{q}_{s;\alpha}$ is the estimated quantile of order α of the marginal distribution of r_t.

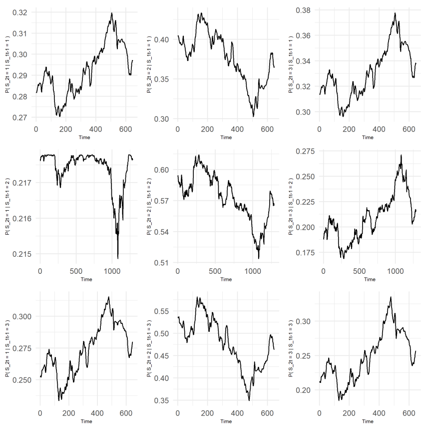
In Figure 5, we have the smoothed conditional probabilities of both series. The number of observations is high, and the probabilities varied abruptly in a small time frame, making the plots hard to read. So, to simplify, a moving average model of order 5 (due to the frequency of the data) was adjusted to these probabilities, to illustrate how they evolve throughout time. Within each series, we see a similar behavior regardless it depends on the previous states of S_{1t} or S_{2t}. Additionally, the scales of the graphs are small, indicating that these probabilities vary around the same set of values.



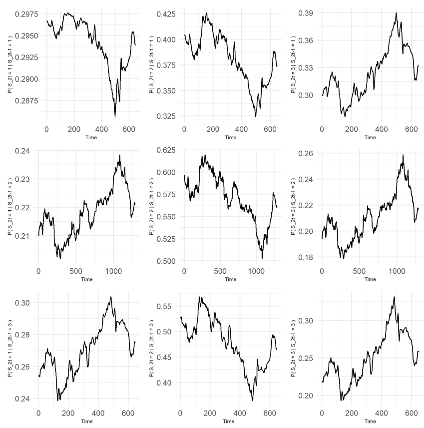
(a) Probabilities of series 1 depending on $spread_{t-1}$ and on series 1 previous state



(b) Probabilities of series 1 depending on $spread_{t-1}$ and on series 2 previous state



(c) Probabilities of series 2 depending on $spread_{t-1}$ and on series 1 previous state



(d) Probabilities of series 2 depending on $spread_{t-1}$ and on series 2 previous state

FIGURE 5: Conditional Probabilities

From this set of probabilities, we can estimate the model. The results are displayed in Table V. Considering the first equation, the effect of the probabilities depending on S&P500's previous state and the interest rate spread has a higher weight on the overall probability. Also, this estimate is highly significant, presenting a p-value close to zero. The effect of DJIA's previous state in S&P500 is lower but it is also significant for a 10% significance level. In the second equation, the effect of S&P500's previous state is higher than DJIA's and both estimates are highly significant.

Equation		Estimate	Std. Error	t-value	p-value	Log-Likelihood
1 (SP500)	$\hat{\lambda}_{11}$	0.685660	0.171241	4.004	0.000	-2636.355
	$\hat{\lambda}_{12}$	0.314340	0.171241	1.836	0.066	
2 (DJIA)	$\hat{\lambda}_{21}$	0.629992	0.176310	3.573	0.000	-2636.622
	$\hat{\lambda}_{22}$	0.370008	0.176309	2.099	0.036	

TABLE V: Results from the function `mmcx()`

One of the advantages of this approach is the possibility to assess the transition probabilities for specific values of x_t, in this case, the interest rate spread. So, for both series, we calculated the transition probabilities for the minimum and maximum value of this variable in the sample, which are -0.52 and 2.97, respectively. In Figure 6, we have the transition probabilities network for S&P500, corresponding to the minimum and maximum value of the spread.

The most noticeable difference between these two networks is regarding the transition probability from the second state to the third state. For the maximum value of $spread_{t-1}$, the transition probability from the second state to the third state is 0.6. So, when the economy is strong, one might expect to have higher returns, when in $t-1$ was in the second state. However, this scenario shifts when considering the minimum value of $spread_{t-1}$. The probability of obtaining higher returns, that is, being in state three, becomes almost evenly distributed, regardless the state in $t-1$. This indicates the instability of the stock market, when the economy is weaker. Another difference in these networks, is regarding the transition probability from the third state to the first state. For the maximum value of $spread_{t-1}$, this probability is 0.27 and for the minimum value increases to 0.44. This is also expected, since when the economy is weaker, the probability of having lower returns is greater.

(a) Conditional probabilities of series 1 for the maximum value of $spread_{t-1}$

(b) Conditional probabilities of series 1 for the minimum value of $spread_{t-1}$

FIGURE 6: Transition Probabilities of Series 1: S&P500

Considering the second equation, corresponding to the DJIA's returns, we see a similar behaviour as in S&P500's networks. The transition probability from the second state to the third state is higher for the maximum value of $spread_{t-1}$ and the transition probability from the third state to the first state is higher when we consider the minimum value of $spread_{t-1}$. Although, the difference of this last probability between the minimum and maximum value of $spread_{t-1}$ is not as big as in S&P500. Overall, the rest of the probabilities structure, remains the same.

(a) Conditional probabilities of series 2 for the maximum value of $spread_{t-1}$

(b) Conditional probabilities of series 2 for the minimum value of $spread_{t-1}$

FIGURE 7: Transition Probabilities of Series 2: DJIA

7 CONCLUSIONS, LIMITATIONS AND FURTHER RESEARCH

Several proposals for inclusion of exogenous variables in MMC models have been presented. The main limitations were associated with the high complexity of the models to be developed and estimated. Additionally, most models considered only categorical exogenous variables, existing a lack of focus on continuous exogenous variables.

This work proposes a new approach to include continuous exogenous variables in Ching, Fung, and Ng (2002) model for multivariate Markov chains. This is relevant because it allows studying the effect of previous series and exogenous variables on the transition probabilities.

The model is based on Ching, Fung, and Ng (2002) MMC model but considers non-homogeneous Markov chains. Thus, the probabilities that compose the model are dependent on exogenous variables. These probabilities are estimated as a usual non-homogeneous Markov chain, through a multinomial logit model. The parameters of the model are then estimated through MLE, as well the standard errors. We developed a package with the estimation function of the model proposed. In this, we considered the Augmented Lagrangian optimization method for estimating the parameters through MLE. Additionally, we designed a Monte Carlo simulation study to assess the power and dimension of test in this model. The results showed that the model detected a non-homogeneous Markov chain. However, when considering a regular case, with two non-homogeneous Markov chains, for low parameter values, the model had lack of power of test. Moreover, an empirical illustration demonstrated the relevance of this new model, by estimating the probability transition matrix, for different values of the exogenous variable. Ignoring the effect of exogenous variables in MMC, means that we would not detect the changes in the probabilities according to the values of the covariates. In this setting, one would have a limited view of the process being study. Hence, this approach allows to understand how a specific variable influences a specific process.

The limitations regarding this work are related to the implementation in **R**, specifically the optimization algorithm applied is not common for MMC models, in that sense, it would be beneficial to study new approaches regarding optimization of the maximum likelihood function as further research. Additionally, it would also be relevant to extend this generalization to the MTD-probit model proposed by Nicolau (2014) (or something similar), that removes the constraints of the model's parameters and allows the model to detect negative effects.

8 REFERENCES

Adke, S.R. and S.R. Deshmukh (1988). "Limit Distribution of a High Order Markov Chain". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 50(1), pp. 105–108. URL: https://www.jstor.org/stable/2345812.

Azzalini, A. (1994). "Logistic regression for autocorrelated data with application to repeated measures". In: *Biometrika* 81(4), pp. 767–775. ISSN: 00063444. DOI: 10.1093/biomet/81.4.767.

Bartholomew, J. (1968). "Stochastic Models for Social Processes". In: *The Australian and New Zealand Journal of Sociology* 4(2), pp. 171–172. DOI: https://doi.org/10.1177/144078336800400215.

Berchtold, A. (1995). "Autoregressive Modelling of Markov Chains". In: *Proc. 10th International Workshop on Statistical Modelling* 104, pp. 19–26. DOI: 10.1007/978-1-4612-0789-4_3.

Berchtold, A. (1996). "Modélisation autorégressive des chaınes de Markov : utilisation d'une matrice différente pour chaque retard". fr. In: *Revue de Statistique Appliquée* 44(3), pp. 5–25. URL: http://www.numdam.org/item/RSA_1996__44_3_5_0/.

Berchtold, A. (2001). "Estimation in the Mixture Transition Distribution Model". In: *Journal of Time Series Analysis* 22(4), pp. 379–397. DOI: https://doi.org/10.1111/1467-9892.00231.

Berchtold, A. (2003). "Mixture transition distribution (MTD) modeling of heteroscedastic time series". In: *Computational Statistics and Data Analysis* 41(3-4), pp. 399–411. ISSN: 01679473. DOI: 10.1016/S0167-9473(02)00191-3.

Berchtold, A., O. Maitre, and K. Emery (2020). "Optimization of the mixture transition distribution model using the march package for R". In: *Symmetry* 12(12), pp. 1–14. ISSN: 20738994. DOI: 10.3390/sym12122031.

Berchtold, A. and A. Raftery (2002). "The mixture transition distribution model for high-order Markov chains and non-Gaussian time series". In: *Statistical Science* 17(3), pp. 328–356. ISSN: 08834237. DOI: 10.1214/ss/1042727943.

Billingsley, P. (1961). “The Lindeberg-Lévy Theorem for Martingales”. In: *Proceedings of the American Mathematical Society* 12(5), pp. 788–792. URL: http://www.jstor.org/stable/2034876.

Bolano, D. (2020). “Handling covariates in markovian models with a mixture transition distribution based approach”. In: *Symmetry* 12(4). ISSN: 20738994. DOI: 10.3390/SYM12040558.

Chauvet, M. and Z. Senyuz (2016). “A dynamic factor model of the yield curve components as a predictor of the economy”. In: *International Journal of Forecasting* 32(2), pp. 324–343. ISSN: 0169-2070. DOI: https://doi.org/10.1016/j.ijforecast.2015.05.007.

Chen, D. G. and Y. L. Lio (2009). “A Novel Estimation Approach for Mixture Transition Distribution Model in High-Order Markov Chains”. In: *Communications in Statistics - Simulation and Computation* 38(5), pp. 990–1003. DOI: 10.1080/03610910802715009.

Ching, W. K., E. S. Fung, and M. K. Ng (2002). “A multivariate Markov chain model for categorical data sequences and its applications in demand predictions”. In: *IMA Journal of Management Mathematics* 13(3), pp. 187–199. DOI: 10.1093/imaman/13.3.187.

Ching, W. K., E. S. Fung, and M. K. Ng (2003). “A Higher-Order Markov Model for the Newsboy’s Problem”. In: *The Journal of the Operational Research Society* 54(3), pp. 291–298.

Ching, W. K., E. S. Fung, and M. K. Ng (2004). “Higher-Order Markov Chain Models for Categorical Data Sequences”. In: *International Naval Research Logistics* 51, pp. 557–574. DOI: 10.1002/nav.20017.

Ching, W. K. and M. K. Ng (2006). *Markov Chains: Models, Algorithms and Applications*. Springer. ISBN: 9780387293370. DOI: 10.1007/0-387-29337-X.

Ching, W. K., M. K. Ng, and E. S. Fung (2008). “Higher-order multivariate Markov chains and their applications”. In: *Linear Algebra and its Applications*(428), pp. 492–507. DOI: 10.1016/j.laa.2007.05.021.

Damásio, B. (2013). “Multivariate Markov Chains - Estimation, Inference and Forecast. A New Approach: What If We Use Them As Stochastic Covariates?” Master dissertation. Universidade de Lisboa, Instituto Superior de Economia e Gestão. URL: http://hdl.handle.net/10400.5/6397.

Damásio, B. (2018). "Essays on Econometrics: Multivariate Markov Chains". PhD dissertation. Universidade de Lisboa, Instituto Superior de Economia e Gestão. URL: https://www.repository.utl.pt/bitstream/10400.5/18128/1/TD-BD-2019.pdf.

Damásio, B. and S. Mendonça (2019). "Modelling insurgent-incumbent dynamics: Vector autoregressions, multivariate Markov chains, and the nature of technological competition". In: *Applied Economics Letters* 26(10), pp. 843–849. DOI: 10.1080/13504851.2018.1502863.

Damásio, B. and S. Mendonça (2020). *Leader-follower dynamics in real historical time: A Markovian test of non-linear causality between sail and steam (co-)development, mimeo*.

Damásio, B. and J. Nicolau (2014). "Combining a regression model with a multivariate Markov chain in a forecasting problem". In: *Statistics & Probability Letters* 90, pp. 108–113. ISSN: 0167-7152. DOI: https://doi.org/10.1016/j.spl.2014.03.026.

Damásio, B. and J. Nicolau (2020). *Time inhomogeneous multivariate Markov chains : detecting and testing multiple structural breaks occurring at unknown*. REM Working Papers 0136–2020. Instituto Superior de Economia e Gestão. URL: http://hdl.handle.net/10400.5/20164.

Dombrosky, A. M. and J. Haubrich (1996). "Predicting real growth using the yield curve". In: *Economic Review*(Q I), pp. 26–35. URL: https://EconPapers.repec.org/RePEc:fip:fedcer:y:1996:i:qi:p:26-35.

Estrella, A. and F. S. Mishkin (1996). "The yield curve as a predictor of U.S. recessions". In: *Current Issues in Economics and Finance* 2(Jun). URL: https://www.newyorkfed.org/research/current_issues/ci2-7.html.

FRED, Federal Reserve Bank of St. Louis (n.d.[a]). *10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity [T10Y3M]*. URL: https://fred.stlouisfed.org/series/T10Y3M.

FRED, Federal Reserve Bank of St. Louis (n.d.[b]). *S&P Dow Jones Indices LLC, Dow Jones Industrial Average [DJIA]*. URL: https://fred.stlouisfed.org/series/DJIA.

FRED, Federal Reserve Bank of St. Louis (n.d.[c]). *S&P Dow Jones Indices LLC, S&P 500 [SP500]*. URL: https://fred.stlouisfed.org/series/SP500.

Hajnal, J. and M. S. Bartlett (1956). "The ergodic properties of non-homogeneous finite Markov chains". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 52(1), pp. 67–77. DOI: 10.1017/S0305004100030991.

Harrell Jr, F. E. (2021). *Hmisc: Harrell Miscellaneous*. R package version 4.5-0. URL: https://CRAN.R-project.org/package=Hmisc.

Hayashi, F. (2000). *Econometrics / Fumio Hayashi.* eng. Princeton University Press: Princeton, N.J. ISBN: 0691010188.

Henningsen, A. and O. Toomet (2011). "maxLik: A package for maximum likelihood estimation in R". In: *Computational Statistics* 26(3), pp. 443–458. DOI: 10.1007/s00180-010-0217-1. URL: http://dx.doi.org/10.1007/s00180-010-0217-1.

Islam, M. A., S. Arabia, and R. I. Chowdhury (2004). "A Three State Markov Model for Analyzing Covariate Dependence". In: *International Journal of Statistical Sciences* 3(i), pp. 241–249. URL: http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/01/P21.V3s.pdf.

Islam, M. A. and R. I. Chowdhury (2006). "A higher order Markov model for analyzing covariate dependence". In: *Applied Mathematical Modelling* 30(6), pp. 477–488. ISSN: 0307904X. DOI: 10.1016/j.apm.2005.05.006.

Jacobs, P.A. and A.W. Lewis (1978). "Discrete Time Series Generated by Mixtures II : Asymptotic Properties". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 40(2), pp. 222–228. URL: https://www.jstor.org/stable/2984759.

Kalbfleisch, J. D. and J. F. Lawless (1985). "The analysis of panel data under a Markov assumption". In: *Journal of the American Statistical Association* 80(392), pp. 863–871. ISSN: 1537274X. DOI: 10.1080/01621459.1985.10478195.

Kaplan, J. (2020). *fastDummies: Fast Creation of Dummy (Binary) Columns and Rows from Categorical Variables*. R package version 1.6.3. URL: https://CRAN.R-project.org/package=fastDummies.

Kijima, M., K. Komoribayashi, and E. Suzuki (July 2002). "A multivariate Markov model for simulating correlated defaults". In: *Journal of Risk* 4. DOI: 10.21314/JOR.2002.066.

Le, N. D., R. D. Martin, and A. Raftery (1996). "Modeling Flat Stretches, Brusts, and Outliers in Time Series Using Mixture Transition Distribution Models". In: *Journal of the American Statistical Association* 91(436), pp. 1504–1515. DOI: 10.1111/j.2517-6161.1985.tb01383.x.

Lèbre, S. and P. Y. Bourguignon (2008). "An EM algorithm for estimation in the mixture transition distribution model". In: *Journal of Statistical Computation and Simulation* 78(8), pp. 713–729. DOI: 10.1080/00949650701266666.

Logan, J. (1981). "A structural model of the higher-order Markov process incorporating reversion effects". In: *The Journal of Mathematical Sociology* 8(1), pp. 75–89. DOI: 10.1080/0022250X.1981.9989916.

Maitre, O. and K. Emery (2020). *march: Markov Chains*. R package version 3.3.2. URL: https://CRAN.R-project.org/package=march.

Martin, R. D. and A. Raftery (1987). "Non-Gaussian State-Space Modeling of Nonstationary Time Series: Comment: Robustness, Computation, and Non-Euclidean Models". In: *Journal of the American Statistical Association* 82(400), pp. 1044–1050. DOI: 10.2307/2289377.

McMillan, D. G. (2021). "Predicting GDP growth with stock and bond markets: Do they contain different information?" In: *International Journal of Finance & Economics* 26(3), pp. 3651–3675. DOI: https://doi.org/10.1002/ijfe.1980.

Mehran, F. (1989). "Analysis of Discrete Longitudinal Data: Infinite-Lag Markov Models". In: *Statistical Data Analysis and Inference*. North-Holland: Amsterdam, pp. 533–541. ISBN: 978-0-444-88029-1. DOI: https://doi.org/10.1016/B978-0-444-88029-1.50053-8.

Muenz, L. R. and L. V. Rubinstein (1985). "Markov Models for Covariate Dependence of Binary Sequences ". In: *Biometrics* 41(1), pp. 91–101. URL: http://www.jstor.org/stable/2530646.

Newey, W.K and D. Mcfadden (1994). "Chapter 36 Large sample estimation and hypothesis testing". In: vol. 4. Handbook of Econometrics. Elsevier, pp. 2111–2245. DOI: https://doi.org/10.1016/S1573-4412(05)80005-4.

Nicholson, W. (2013). *DTMCPack: Suite of functions related to discrete-time discrete-state Markov Chains*. R package version 0.1-2. URL: https://CRAN.R-project.org/package=DTMCPack.

Nicolau, J. (2014). "A new model for multivariate markov chains". In: *Scandinavian Journal of Statistics* 41(4), pp. 1124–1135. ISSN: 14679469. DOI: 10.1111/sjos.12087.

Nicolau, J. and F. I. Riedlinger (2014). "Estimation and inference in multivariate Markov chains". In: *Statistical Papers* 56(4), pp. 1163–1173. ISSN: 09325026. DOI: 10.1007/s00362-014-0630-6.

Novomestky, F. (2012). *matrixcalc: Collection of functions for matrix calculations*. R package version 1.0-3. URL: https://CRAN.R-project.org/package=matrixcalc.

Pegram, G. (1980). "An Autoregressive Model for Multilag Markov Chains". In: *Journal of Applied Probability* 17(2), pp. 350–362. DOI: 10.2307/3213025.

Raftery, A. (1985). "A Model for High-Order Markov Chains". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 47(3), pp. 528–539. ISSN: 0035-9246. DOI: 10.1111/j.2517-6161.1985.tb01383.x.

Raftery, A. and S. Tavaré (1994). "Estimation and Modelling Repeated Patterns in High Order Markov Chains with the Mixture Transition Distribution Model". In: *Applied Statistics* 43(1), pp. 179–199. DOI: 10.2307/2986120.

Rajarshi, M.B. (2013). *Statistical Inference for Discrete Time Stochastic Processes*. SpringerBriefs in Statistics. ISBN: 9783642179792. URL: http://www.springer.com/978-81-322-0762-7.

Regier, M. H. (1968). "A Two-State Markov Model for Behavioral Change". In: *Journal of the American Statistical Association* 63(323), pp. 993–999. DOI: 10.1080/01621459.1968.11009325.

Siu, T. K. et al. (2005). "On a multivariate Markov chain model for credit risk measurement". In: *Quantitative Finance* 5(6), pp. 543–556. ISSN: 14697688. DOI: 10.1080/14697680500383714.

Spedicato, G. A. (July 2017). "Discrete Time Markov Chains with R". In: *The R Journal*. R package version 0.6.9.7. URL: https://journal.r-project.org/archive/2017/RJ-2017-036/index.html.

Spilerman, S. and B. Singer (1976). "The Representation of Social Processes by Markov Models". In: *American Journal of Sociology* 82(1), pp. 1–54. URL: https://www.jstor.org/stable/2777460.

Taylor, H. M. and S. Karlin (1984). *An Introduction to Stochastic Modeling - 3rd ed.* Academic Press. ISBN: 9780126848878. DOI: 10.1016/C2013-0-11589-9.

Tian, R. and G. Shen (2019). "Predictive power of Markovian models: Evidence from US recession forecasting". In: *Journal of Forecasting* 38(6), pp. 525–551. DOI: https://doi.org/10.1002/for.2579.

Varadhan, R. (2015). *alabama: Constrained Nonlinear Optimization*. R package version 2015.3-1. URL: https://CRAN.R-project.org/package=alabama.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S*. Fourth. ISBN 0-387-95457-0. Springer: New York. URL: https://www.stats.ox.ac.uk/pub/MASS4/.

Wang, C., T. Z. Huang, and W. K. Ching (2014). "A new multivariate Markov chain model for adding a new categorical data sequence". In: *Mathematical Problems in Engineering* 2014. DOI: 10.1155/2014/502808.

Wasserman, S. (1980). "Analyzing social networks as stochastic processes". In: *Journal of the American Statistical Association* 75(370), pp. 280–294. DOI: 10.1080/01621459.1980.10477465.

Wong, C. S. and W. K. Li (2001). "On a mixture autoregressive conditional heteroscedastic model". In: *Journal of the American Statistical Association* 96(455), pp. 982–995. DOI: 10.1198/016214501753208645.

Zhang, X., M. L. King, and R. J. Hyndman (2006). "A Bayesian approach to bandwidth selection for multivariate kernel density estimation". In: *Computational Statistics and Data Analysis* 50(11), pp. 3009–3031. DOI: 10.1016/j.csda.2005.06.019.

Zhu, D. M. and W. K. Ching (2010). "A new estimation method for multivariate Markov chain model with application in demand predictions". In: *Proceedings - 3rd International Conference on Business Intelligence and Financial Engineering, BIFE 2010*, pp. 126–130. DOI: 10.1109/BIFE.2010.39.