REM WORKING PAPER SERIES

Time Inhomogeneous Multivariate Markov Chains: Detecting and Testing Multiple Structural Breaks Occurring at Unknown

Bruno Damásio, João Nicolau

REM Working Paper 0136-2020

June 2020

REM – Research in Economics and Mathematics
Rua Miguel Lúpi 20,
1249-078 Lisboa,
Portugal

ISSN 2184-108X

REM – Research in Economics and Mathematics

Rua Miguel Lupi, 20
1249-078 LISBOA
Portugal

Telephone: +351 - 213 925 912
E-mail: rem@iseg.ulisboa.pt

https://rem.rc.iseg.ulisboa.pt/

https://twitter.com/ResearchRem

https://www.linkedin.com/company/researchrem/

https://www.facebook.com/researchrem/

Time Inhomogeneous Multivariate Markov Chains: Detecting and Testing Multiple Structural Breaks Occurring at Unknown Dates

Bruno Damásio*†‡ João Nicolau ‡

June, 2020

Abstract

Markov chains models are used in several applications and different areas of study. Usually a Markov chain model is assumed to be homogeneous in the sense that the transition probabilities are time invariant. Yet, ignoring the inhomogeneous nature of a stochastic process by disregarding the presence of structural breaks can lead to misleading conclusions. Several methodologies are currently proposed for detecting structural breaks in a Markov chain, however, these methods have some limitations, namely they can only test directly for the presence of a single structural break. This paper proposes a new methodology for detecting and testing the presence multiple structural breaks in a Markov chain occurring at unknown dates.

Keywords: Inhomogeneous Markov chain, structural breaks, time-varying probabilities.

*bdamasio@novaims.unl.pt
†NOVA Information Management School (NOVA IMS)
‡ISEG, Universidade de Lisboa, REM - Research in Economics and Mathematics, CEMAPRE.

1 Introduction

Let $\{S_t, t = 0, 1, 2 \cdots, \infty\}$, hereinafter $\{S_t\}$, be a stochastic process that involves a sequence of discrete random variables with domain $E = \{1, \cdots, q\}$. Furthermore, $\{S_t\}$ is a first order Markov chain in the sense that:

$$P\left(S_t = i_0 \mid \mathcal{F}_{t-1}\right) = P\left(S_t = i_0 \mid S_{t-1} = i_1\right) \equiv P_{i_1 i_0}, \tag{1.1}$$

where $\mathcal{F}_{t-1}$ is the σ-field generated by all available information until the period $t-1$.

Given an initial condition and once the domain E is known, S_t can be fully characterized with the associated transition probability matrix (TPM) P. This matrix contains all possible one step ahead transitions generated in the space $E = \{1, \cdots, q\}$, such that, for the generic period t, may be written as:

$$\boldsymbol{P_t} = \begin{pmatrix} P_{t,11} & P_{t,12} & \cdots & P_{t,1q} \\ P_{t,21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ P_{t,q1} & \cdots & \cdots & P_{t,qq} \end{pmatrix} \tag{1.2}$$

Markov chain models have shown to be proficiently and interdisciplinary used. Notably in Economics (Mehran, 1989), Finance (Siu et al., 2005; Fung and Siu, 2012), Financial Markets (Maskawa, 2003; Nicolau, 2014; Nicolau and Riedlinger, 2015), Biology (Gottschau, 1992; Raftery and Tavaré, 1994; Berchtold, 2001), Environmental Sciences (Turchin, 1986; Sahin and Sen, 2001; Shamshad et al., 2005), Physics (Gómez et al., 2010; Boccaletti et al., 2014), Linguistics[1](Markov, 1913), Medicine (Li et al., 2014), Forecasting (Damásio and Nicolau, 2013), Management (Horvath et al., 2005), Sports (Bukiet et al., 1997), the estimation of expected hitting times (Nicolau, 2017; Damásio et al., 2018), Operational Research (Asadabadi, 2017; Tsiliyannis, 2018; Cabello, 2017), Economic History (Damásio and Mendonça, 2018), among others; see, e.g. (Ching and Ng, 2006; Sericola, 2013).

Notwithstanding the common denominator of this research depicts the assumption about the homogeneous nature of the Markov chain. In fact, in all of the aforementioned studies the Markov chain assumed to be homogeneous in the sense that $P_{t,i_1 i_0} = P_{i_1 i_0}$ and $\boldsymbol{P_t} = \boldsymbol{P}$. In other words, the transition probability matrix does not depend on time.

Ignoring the nonhomogeneous nature of a stochastic process by disregarding the presence of structural breaks can lead to misleading conclusions.[2] As concluded by Hansen (2001) *Structural change is pervasive in economic time series relationships, and it can be quite perilous to ignore. Inferences about economic relationships can go astray, forecasts can be inaccurate, and policy recommendations can be misleading or worse* (Hansen, 2001, p. 237). Structural breaks are, then again, a common issue in economical environments. In fact, the economic dynamics is characterised by deep complex, and mutating, patterns of interdependence between variables and aggregates. For this reason, it is of great relevance to investigate whether the quantities $P_{t,i_1 i_0}$ for $i_k \in E, k = 0, \cdots, q$ are time invariant or, in contrast, time dependent.

The change detection involving dependent observations such as Markov chains has been studied by Lai (1995, 1998); Mei (2006); Yakir (1994); Tan and Yılmaz (2002); Polansky (2007); Höhle (2010); Darkhovsky (2011); Xian et al. (2016), among others. Most of these approaches are essentially based on the calculation of a cumulative sum (CUSUM) and are designed to detect a single structural break.

This paper proposes new methodology for detecting and testing the presence of multiple structural

[1]In 1913 A. A. Markov illustrated his chains for the first time with an example taken from literature. He investigated a sequence of 20,000 letters in Pushkin's text *Eugeny Onegin* to model probability transitions between consonants and vowels, (Markov, 1913)

[2]The consequences of ignoring a structural break are widely documented in the literature, namely in the unit root tests, see, e.g. Rappoport and Reichlin (1989); Perron (1989)

breaks in a Markov chain occurring at unknown dates. It takes advantages of Qu and Perron (2007) procedure to detected structural breaks in econometric type of models. The main advantage of our approach is related to the power gains that can result by explicitly considering the possibility of multiple breaks, as opposite to standard CUSUM tests where the alternative hypothesis allows just a single break. Therefore, although a single break test may be consistent against multiple breaks, substantial power gains can result from using tests for multiple structural changes (see Perron (2006), for a detailed discussion of this topic).

The rest of this article is organized as follows. Section 2 exposes our the theoretical framework, enumerating the assumptions that allow a Vector Autoregressive (VAR) representation of a Markov chain and discussing the main results of inhomogeneity detection in a Markov chain. Section 3 presents a Monte Carlo simulation where the power and the size of the proposed method is analysed, whereas Section 4 illustrates the methodology with an economic application using the NASDAQ stock index. Section 5 discusses some possible extensions of this methodology. Finally, Section 6 elaborates on the summary of the main results and concludes.

2 The model and assumptions

In this section, we present the basic econometric context upon which our analysis will be elaborated. Furthermore a new method for detecting and testing multiple structural breaks occurring at unknown dates in non homogeneous Markov chains is proposed. Our strategy, which consists of representing a Markov chain in the form of a VAR model[3], comprises three distinct phases, namely:

1. To identify the conditions under which a Markov chain admits a VAR representation;
2. To represent a Markov chain into a VAR form;
3. To propose a theoretical econometric framework to detect and test for multiple structural breaks occurring at unknown dates.

2.1 A vectorial autoregressive representation of a stationary Markov chain

Consider the following random q-dimensional vector

$$\boldsymbol{y}_t = \begin{pmatrix} y_{1t} & \cdots & y_{kt} & \cdots & y_{qt} \end{pmatrix}' \tag{2.1}$$

whose k-th element y_{kt} equals $\mathbb{1}\{S_t = k\}$, (here $\mathbb{1}\{\cdot\}$ denotes the indicator function, such that $y_{kt} = 1$ if $S_t = k$ and 0 otherwise).

Moreover, when $S_t = i$ then k-th element of $\boldsymbol{y}_{t+1}$, $y_{k,t+1}$, is a r.v. such that

$$P\left(S_{t+1} = k \mid S_t = i\right) = P\left(y_{k,t+1} = 1 \mid y_{it} = 1\right) = P_{ik} \tag{2.2}$$

and, by Markovian property and, without any loss of generality, assuming a first order MC it follows that $\mathbf{E}\left[\boldsymbol{y}_{t+1}\middle| S_t = i\right] = \boldsymbol{P}_{\boldsymbol{i}\bullet}$, the i-th row of $\boldsymbol{P}$.

Given this result it follows that an ergodic MC with domain $E = \{1, \cdots, q\}$ admits the following system representation:

[3] Markov chains have also proven to be a valuable tool when it comes to the approximation of VAR models, see e.g. Tauchen (1986). The estimated transition probabilities might be relevant to find numerical solutions to integral equations in the absence of integration.

$$\begin{cases} y_{1t} = & P_{11}y_{1,t-1} + P_{21}y_{2,t-1} + \cdots + P_{q1}y_{q,t-1} + \varepsilon_{1t} \\ y_{2t} = & P_{12}y_{1,t-1} + P_{22}y_{2,t-1} + \cdots + P_{q2}y_{q,t-1} + \varepsilon_{2t} \\ \vdots & \vdots \\ y_{q-1,t} = & P_{1,q-1}y_{1,t-1} + P_{2,q-1}y_{2,t-1} + \cdots + P_{q,q-1}y_{q,t-1} + \varepsilon_{q-1,t}. \end{cases} \tag{2.3}$$

Or, equivalently,

$$\begin{cases} y_{1t} = & \boldsymbol{z}_t'\boldsymbol{P}_{\bullet 1} + \varepsilon_{1t} \\ \vdots & \vdots \\ y_{q-1,t} = & \boldsymbol{z}_t'\boldsymbol{P}_{\bullet q-1} + \varepsilon_{q-1,t} \end{cases}, \tag{2.4}$$

where $\boldsymbol{z}_t' = \begin{pmatrix} y_{1,t-1}, & \cdots & , y_{q,t-1} \end{pmatrix}$; $\varepsilon_{it} \equiv y_{it} - \mathbf{E}\left[y_{it} \mid \mathcal{F}_{t-1}\right]$ for $i = 1, \cdots, q-1$; and $\boldsymbol{P}_{\bullet k}$ is the k-th column of $\boldsymbol{P}$, $k = 1, \cdots q$.

Furthermore let the sufficient statistics for $\boldsymbol{P}$, n_{ik}, $\forall i, k \in E$, denote the number of transition frequencies of the type $i \to k$ in the sample, i.e.

$$n_{ik} = \sum_{t=1}^{n} \mathbb{1}\left\{S_t = k, S_{t-1} = i\right\}, \tag{2.5}$$

it can be shown that the likelihood function (the distribution is multinomial) is

$$l\left(P_{ik}\right) \propto \sum_i \sum_k n_{ik} log\left(P_{ik}\right) \tag{2.6}$$

and the maximum likelihood estimator for P_{ik} is

$$\hat{P}_{ik} = \frac{\sum_{t=1}^{n} \mathbb{1}\left\{S_t = k, S_{t-1} = i\right\}}{\sum_{k=1}^{q}\sum_{t=1}^{n} \mathbb{1}\left\{S_t = k, S_{t-1} = i\right\}} = \frac{n_{ik}}{n_i} \tag{2.7}$$

see Basawa (2014); Billingsley (1961).

It is useful to represent models (2.3) and (2.4) in the following matrix form:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{q-1t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{z}_t' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{z}_t' & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{z}_t' \end{pmatrix} \begin{pmatrix} \boldsymbol{P}_{\bullet \mathbf{1}} \\ \boldsymbol{P}_{\bullet \mathbf{2}} \\ \vdots \\ \boldsymbol{P}_{\bullet \boldsymbol{q-1}} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{q-1t} \end{pmatrix} \tag{2.8}$$

or equivalently,

$$\begin{aligned} \underset{(q-1)\times 1}{\boldsymbol{y}_t} &= \left(\underset{(q-1)\times(q-1)}{\boldsymbol{I}} \otimes \underset{1\times q}{\boldsymbol{z}'}\right) \underset{q(q-1)\times q(q-1)}{\mathrm{vec}\left(\boldsymbol{P}^*\right)} + \underset{(q-1)\times 1}{\boldsymbol{\varepsilon}_t} \\ &= \underset{(q-1)\times q(q-1)}{\boldsymbol{x}_t'} \times \underset{q(q-1)\times 1}{\boldsymbol{p}} + \underset{(q-1)\times 1}{\boldsymbol{\varepsilon}_t} \end{aligned} \tag{2.9}$$

where

$$
\begin{aligned}
\boldsymbol{P}^{*} &= \begin{pmatrix} P_{11} & P_{12} & P_{1q-1} \\ P_{21} & P_{22} & P_{2q-1} \\ \vdots & \vdots & \vdots \\ P_{q1} & P_{q2} & P_{qq-1} \end{pmatrix} && (2.10) \\
&= \begin{pmatrix} \boldsymbol{P_{\bullet 1}} & \boldsymbol{P_{\bullet 2}} & \cdots & \boldsymbol{P_{\bullet q-1}} \end{pmatrix}, && (2.11)
\end{aligned}
$$

$$
\boldsymbol{p} \equiv \operatorname{vec}\left(\boldsymbol{P}^{*}\right) = \begin{pmatrix} \boldsymbol{P_{\bullet 1}} \\ \boldsymbol{P_{\bullet 2}} \\ \vdots \\ \boldsymbol{P_{\bullet q-1}} \end{pmatrix}, \tag{2.12}
$$

and $\boldsymbol{\varepsilon}_t$ is a martingale difference sequence with covariance matrix $\boldsymbol{\Sigma} \equiv \mathbf{E}\left[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'\right]$ given by

$$
\boldsymbol{\Sigma} = \begin{pmatrix} \pi_1 - \sum_{k=1}^{q} \pi_k P_{k1}^2 & -\sum_{k=1}^{q} P_{k1} P_{k2} \pi_k & \cdots & -\sum_{k=1}^{q} P_{k1} P_{kq} \pi_k \\ -\sum_{k=1}^{q} P_{k1} P_{k2} \pi_k & \pi_2 - \sum_{k=1}^{q} \pi_k P_{k2}^2 & \cdots & -\sum_{k=1}^{q} P_{k2} P_{kq} \pi_k \\ \vdots & \vdots & \ddots & \vdots \\ -\sum_{k=1}^{q} P_{k1} P_{kq} \pi_k & -\sum_{k=1}^{q} P_{k2} P_{kq} \pi_k & & \pi_q - \sum_{k=1}^{q} \pi_k P_{kq}^2 \end{pmatrix}, \tag{2.13}
$$

see Lemma (1), Mathematical Appendix.

Expression (2.9) suggests that, under certain conditions, a Markov chain may assume a VAR representation.[4] A clear implication of this circumstance is that detecting, and testing, non homogeneities in a Markov chain stochastic process can be treated as a problem of testing for structural breaks in linear systems of equations.

In this sense, the final model may be written as a VAR model (subject to m breaks) with $m+1$ distinct regimes, or segments, such that:

$$
\boldsymbol{y}_t = \begin{cases} \boldsymbol{x}_t' \boldsymbol{p}_1 + \boldsymbol{\varepsilon_1}, & \text{for } t = T_0 + 1, \cdots, T_1 \\ \boldsymbol{x}_t' \boldsymbol{p}_2 + \boldsymbol{\varepsilon_2}, & \text{for } t = T_1 + 1, \cdots, T_2 \\ \vdots & \vdots \\ \boldsymbol{x}_t' \boldsymbol{p}_m + \boldsymbol{\varepsilon_m}, & \text{for } t = T_{m-1} + 1, \cdots, T_m \\ \boldsymbol{x}_t' \boldsymbol{p}_{m+1} + \boldsymbol{\varepsilon_{m+1}}, & \text{for } t = T_m + 1, \cdots, T_{m+1} \end{cases} \tag{2.14}
$$

or

$$
\boldsymbol{y}_t = \boldsymbol{x}_t' \boldsymbol{p_j} + \boldsymbol{\varepsilon_j} \tag{2.15}
$$

where $\boldsymbol{\varepsilon_j}$ is a martingale difference sequence with covariance matrix $\boldsymbol{\Sigma}_j$; $T_0 = 0$ and $T_{m+1} = T$; for $T_{j-1} + 1 \leq t \leq T_j, j = 1, \cdots, m+1$.

The main objective of this article is to consistently estimate the stacked vectors of parameters $\boldsymbol{\theta} \equiv \left(\boldsymbol{p}_1, \cdots, \boldsymbol{p}_{m+1}; \boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_{m+1}\right)$, and the $m-$ *dimensional* break dates vector $\boldsymbol{\mathcal{T}} = (T_1, \cdots, T_m)$.

With regard to estimating and testing for multiple structural breaks in systems of linear equations,

[4] We will explain such conditions later

Bai et al. (1998) proposed a method for testing one single break and Hansen (2003) considered multiple breaks occurring at known dates in a cointegrated system. Bai (2000) considered a problem of testing multiple breaks occurring at unknown dates but, as far as we know, the most general theoretical framework for testing the presence of structural breaks occurring at unknown dates is the one proposed by Qu and Perron (2007). Within this approach it is possible to test within- and cross-equation restrictions of the type $\boldsymbol{g}\left(\boldsymbol{p}^{\star}, \operatorname{vec}\left(\boldsymbol{\Sigma}^{\star}\right)\right)=\mathbf{0}$, where $\boldsymbol{p}^{\star} \equiv\left(\boldsymbol{p}_1^{\prime}, \boldsymbol{p}_2^{\prime}, \cdots, \boldsymbol{p}_{m+1}^{\prime}\right)^{\prime}$ and $\boldsymbol{\Sigma}^{\star} \equiv\left(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \cdots, \boldsymbol{\Sigma}_{m+1}\right)$. This is relevant in the sense that several interesting special cases can be approached: i) partial structural change models (only a subset of the parameters are subject to change), ii) block partial structural change models (only a subset of the equations are subject to change); iii) ordered break models (the breaks can occur in a particular order across subsets of equations); among others. Additionally, the problem of structural breaks in Markov chain models for panel data can be addressed. See Perron (2006) for a discussion of some methodological issues related to estimation and testing of structural changes in the linear models.

2.2 The model: main assumptions

In this subsection we present the main econometric theory that supports our strategy. For this purpose, the following assumptions on the Markov chain are imposed.

Assumption A1. *$\{S_t\}$ is a first order, possibly, m-inhomogeneous Markov chain.*

Remark 1. *A Markov chain is said to be m-inhomogeneous if and only if the maximum number of distinctive transition probability matrices is m - the number of segments is $m+1$. Or, in other words, the number of breaks is m.*

Assumption A2. *$\{S_t\}$ is a positive recurrent MC and aperiodic in the sense that its states are positive recurrent aperiodic, in each potential segment $j=1, \cdots, m+1$.*

Remark 2. *A positive recurrent and aperiodic MC is said to be ergodic or irreducible. In these circumstances the process $\{S_t\}$ admits a unique stationary distribution in each segment given by $\boldsymbol{\Pi}_j$, by the Perron-Frobenius theorem and given that each $\boldsymbol{P}_t$, for $T_{j-1}+1 \leq t \leq T_j, j=1, \cdots, m+1$ matrix has an eigenvalue λ_{jt} equal to one and all of them have roots outside the unity circle, see Suhov and Kelbert (2008).*

Assumption A3. $\exists \ell_0$: *the minimum eigenvalues of*

$$(1/\ell) \sum_{j=T_j^0+1}^{T_j^0+\ell} \boldsymbol{x}_t \boldsymbol{x}_t^{\prime}$$

and of

$$(1/\ell) \sum_{j=T_j^0-\ell}^{T_j^0} \boldsymbol{x}_t \boldsymbol{x}_t^{\prime}$$

are bounded away from zero, for $j=1, \cdots, m ; \forall \ell>\ell_0$.

Assumption A4. *The matrices $\sum_{t=k}^{\ell} \boldsymbol{x}_t \boldsymbol{x}_t^{\prime}$ are invertible $\forall\ \ell-k \geq k_0$, for some $0<k_0<\infty$.*

Assumption A3 requires that there is no local perfect collinearity in the regressors near the break dates. This ensures that the break dates are identifiable. Assumption A4 is a standard invertibility condition.

These assumptions are plausible given that the Markov chain is positive recurrent and aperiodic for each segment.

Within our theoretical framework the following propositions arise.

Propositon 1. *Under Assumptions A1 and A2, the OLS estimator for (2.9) is, regardless of the sample size, numerically equal to the one obtained through the ML that assumes a multinomial distribution in (2.7)*

Proof. See the Mathematical Appendix. ■

Henceforth we will use the superscript 0 to denote the true values of model parameters, such that $\left(\boldsymbol{p}_1^0, \cdots, \boldsymbol{p}_{m+1}^0\right)$, $\left(\boldsymbol{\Sigma}_1^0, \cdots, \boldsymbol{\Sigma}_{m+1}^0\right)$, and $\boldsymbol{\mathcal{T}}^0 = (T_1^0, \cdots, T_m^0)$ denotes, respectively, the true value of the mean equations parameters, the true values of the error covariance matrix and the true break dates. Furthermore let

$$\boldsymbol{\theta}^0 \equiv \left(\boldsymbol{p}_1^0, \cdots, \boldsymbol{p}_{m+1}^0; \boldsymbol{\Sigma}_1^0, \cdots, \boldsymbol{\Sigma}_{m+1}^0\right) \tag{2.16}$$

represent the stacked value of the true parameters of the model.

Propositon 2. *Under Assumptions A1 and A2 we have*

$$\ell_j^{-1} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+\ell_j} \boldsymbol{x}_t \boldsymbol{x}_t' \overset{a.s.}{\to} Q_j, \tag{2.17}$$

a non random positive definite matrix, as $\ell_j \to \infty$, for each $j = 1, \cdots, m+1$ and $\ell_j \leq T_j^0 - T_{j-1}^0 + 1$.

Proof. See the Mathematical Appendix ■

This proposition ensures the verification of all necessary conditions for the application of the central limit theorem.

Propositon 3. *Under Assumptions A1 and A2:*

1. $\{\boldsymbol{x_t \varepsilon_t}, \mathcal{F}_t\}$ *forms a martingale difference sequence;*
2. $\mathbf{E}\left[\boldsymbol{x_t \varepsilon_t}\right] = 0$;

Proof. See the Mathematical Appendix ■

Proposition 3 naturally holds if $\boldsymbol{x}_t$ is replaced by $\boldsymbol{\varepsilon}_t$ (or by $\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' - \boldsymbol{\Sigma}_j$).

2.3 The estimation procedure

The main objective of this section is to discuss the estimation procedure, as well as to expose the limiting distribution of parameter estimators. Let us focus now on the estimation of the parameters $\boldsymbol{\theta}$, we will elaborate on the estimation and inference of the break fractions latter on. Assuming that the transition probabilities change in known periods[5], conditioning on a partition of the sample $\boldsymbol{\mathcal{T}}$, the parameters of the model (2.15) can be consistently estimated through the quasi-maximum likelihood method. The quasi-likelihood function is:

$$L\left(\boldsymbol{p}, \boldsymbol{\Sigma}\right) = \prod_{j=1}^{m+1} \prod_{t=T_{j-1}+1}^{T_j} f\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t; \boldsymbol{p}_j, \boldsymbol{\Sigma}_j\right) \tag{2.18}$$

[5]Or that we have consistent estimators for $\boldsymbol{\mathcal{T}}$.

and the the quasi-likelihood ratio may be written as:

$$LR = \frac{\prod_{j=1}^{m+1} \prod_{t=T_{j-1}+1}^{T_j} f\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t; \boldsymbol{p}_j, \boldsymbol{\Sigma}_j\right)}{\prod_{j=1}^{m+1} \prod_{t=T_{j-1}^0+1}^{T_j^0} f\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t; \boldsymbol{p}_j^0, \boldsymbol{\Sigma}_j^0\right)} \tag{2.19}$$

with

$$f\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t; \boldsymbol{p}_j, \boldsymbol{\Sigma}_j\right) = \frac{1}{(2\pi)^{(q-1)/2} \left|\Sigma_j\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{y}_t - \boldsymbol{x}_t'\boldsymbol{p}_j\right)' \boldsymbol{\Sigma}_j^{-1} \left(\boldsymbol{y}_t - \boldsymbol{x}_t'\boldsymbol{p}_j\right)\right\} \tag{2.20}$$

The estimators for $\boldsymbol{p}_j$ and $\boldsymbol{\Sigma}_j$ are obtained as

$$\left(\hat{\boldsymbol{p}}_j, \hat{\boldsymbol{\Sigma}}_j\right) \equiv \underset{\left(\boldsymbol{\mathcal{T}}, \boldsymbol{p}_j, \boldsymbol{\Sigma}_j\right)}{argmax} \log\left(LR\left(\boldsymbol{p}, \boldsymbol{\Sigma}\right)\right) \tag{2.21}$$

resulting in the following joint closed solutions

$$\hat{\boldsymbol{p}}_j = \left(\sum_{t=T_{j-1}+1}^{T_j} \boldsymbol{x}_t\boldsymbol{x}_t'\right)^{-1} \sum_{t=T_{j-1}+1}^{T_j} \boldsymbol{x}_t\boldsymbol{y}_t \tag{2.22}$$

$$\hat{\boldsymbol{\Sigma}}_j = \frac{1}{T_j - T_{j-1}} \sum_{t=T_{j-1}+1}^{T_j} \left(\boldsymbol{y}_t - \boldsymbol{x}_t'\hat{\boldsymbol{p}}_j\right)\left(\boldsymbol{y}_t - \boldsymbol{x}_t'\hat{\boldsymbol{p}}_j\right)' \tag{2.23}$$

with this maximization being taken over some set of admissible partitions $\boldsymbol{\mathcal{T}}$ in the set:

$$\boldsymbol{\Lambda}_\varepsilon = \left\{(T\lambda_1, ..., T\lambda_m)); \left|\lambda_{j+1} - \lambda_j\right| \geq \varepsilon, \lambda_1 \geq \varepsilon, \lambda_m \leq 1 - \varepsilon\right\}, \tag{2.24}$$

where ε is a trimming parameter that imposes a minimal length for each segment and λ_j denotes the break fractions in such a way that $T_j = T\lambda_j$. It should be noticed that, although the distribution of epsilons are not normal, the use of the log likelihood function based on the Gaussian distribution lead to a consistent estimators. Actually, they coincide with the maximum likelihood estimators (see Proposition 1). To prove consistency and asymptotic normality of $\hat{\boldsymbol{p}}_j$ we treat $\hat{\boldsymbol{p}}_j$ as an M-estimator. In fact, the associated objective function might be written as a sample average $T^{-1}\sum_{t=1}^{T} \boldsymbol{m}\left(\boldsymbol{w}, \boldsymbol{\theta}\right)$.

Furthermore, the log likelihood ratio (2.19) can be decomposed as follows:

$$\begin{aligned}
\log LR &= \log\left(\frac{\prod_{j=1}^{m+1}\prod_{t=T_{j-1}+1}^{T_j} f\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t;\boldsymbol{p}_j,\boldsymbol{\Sigma}_j\right)}{\prod_{j=1}^{m+1}\prod_{t=T_{j-1}^0+1}^{T_j^0} f\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t;\boldsymbol{p}_j^0,\boldsymbol{\Sigma}_j^0\right)}\right)\\
&= \sum_{j=1}^{m+1}\sum_{t=T_{j-1}+1}^{T_j} \log f\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t;\boldsymbol{p}_j,\boldsymbol{\Sigma}_j\right) - \\
& \quad \sum_{j=1}^{m+1}\sum_{t=T_{j-1}^0+1}^{T_j^0} \log f\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t;\boldsymbol{p}_j^0,\boldsymbol{\Sigma}_j^0\right)\\
&= \sum_{t=1}^{T}\left[\left\{\mathbb{1}\left\{t\in seg_1\right\} l\left(\boldsymbol{w}_t;\boldsymbol{\theta}_1\right) - \mathbb{1}\left\{t\in seg_1^0\right\} l\left(\boldsymbol{w}_t;\boldsymbol{\theta}_1^0\right)\right\} + \cdots +\right.\\
&+ \left.\left\{\mathbb{1}\left\{t\in seg_{m+1}\right\} l\left(\boldsymbol{w}_t;\boldsymbol{\theta}_{m+1}\right) - \mathbb{1}\left\{t\in seg_{m+1}^0\right\} l\left(\boldsymbol{w}_t;\boldsymbol{\theta}_{m+1}^0\right)\right\}\right]\\
&= \sum_{t=1}^{T}\left[\sum_{j=1}^{m+1}\left\{\mathbb{1}\left\{t\in seg_j\right\} l\left(\boldsymbol{w}_t;\boldsymbol{\theta}_j\right) - \mathbb{1}\left\{t\in seg_j^0\right\} l\left(\boldsymbol{w}_t;\boldsymbol{\theta}_j^0\right)\right\}\right]\\
&= \sum_{t=1}^{T}\boldsymbol{m}\left(\boldsymbol{w}_t,\boldsymbol{\theta}\right) \qquad (2.25)
\end{aligned}$$

where $\mathbb{1}\left\{t\in seg_j\right\}$ means that we are in the j-th segment or, in other words, that $T_{j-1}+1\leq t\leq T_j$ for $j=1,\cdots,m$, and maximizing the objective function $\sum_{t=1}^{T}\boldsymbol{m}\left(\boldsymbol{w},\boldsymbol{\theta}\right)$ is equivalent to maximizing $T^{-1}\sum_{t=1}^{T}\boldsymbol{m}\left(\boldsymbol{w},\boldsymbol{\theta}\right)$.

As the parameter set is compact ($\boldsymbol{\theta}$ involves only the transition probabilities P_{ij} which are obviously bounded between 0 and 1) and the identification condition automatically holds by Proposition 2, we just need to assume the standard dominance condition

$$\mathbf{E}\left[sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|l\left(\boldsymbol{w}_t,\boldsymbol{\theta}_j\right)\right|\right]<\infty, j=1,\cdots,m+1$$

and Newey and McFadden (1994) Theorem 2.5 is verified, hence we have $\hat{\boldsymbol{p}}_j \xrightarrow{p} \boldsymbol{p}_j^0$.

Moreover, assuming that for each $j=1,\cdots,m$, the standard mild conditions for the asymptotic normality of an M-Estimator[6] and Newey and McFadden (1994) Theorem 3.3 is automatically verified and we have, in these circumstances,

$$\sqrt{T_j-T_{j-1}}\left(\hat{\boldsymbol{p}}_j-\boldsymbol{p}^0\right)\xrightarrow{d} N\left(\mathbf{0},Avar\left(\hat{\boldsymbol{p}}_j\right)\right), \qquad (2.26)$$

where $Avar\left(\hat{\boldsymbol{p}}_j\right)=\left[plim\left(\frac{1}{T_j-T_{j-1}}\sum_{t=T_{j-1}+1}^{T_j}\boldsymbol{z}_t\boldsymbol{z}_t'\right)^{-1}\otimes\boldsymbol{\Sigma}_j\right]$, for $j=1,\cdots,m$.

Let us know focus on the estimation of the break dates

$$(\hat{T}_1,\cdots,\hat{T}_m)=(T\hat{\lambda}_1,...,T\hat{\lambda}_m)$$

elaborating on the estimation of the parameters in a very general setup such that (within and cross-

[6]More precisely, that $\boldsymbol{\theta}_0$ is in the interior of $\boldsymbol{\Theta}$; $f\left(\boldsymbol{y}_t\mid\boldsymbol{x}_t;\boldsymbol{\theta}_j\right)$ is twice continuously differentiable in $\boldsymbol{\theta}_j$ for all $(\boldsymbol{y}_t,\boldsymbol{x}_t)$; $\mathbf{E}\left[\boldsymbol{s}\left(\boldsymbol{w}_t,\boldsymbol{\theta}_0\right)\right]=\mathbf{0}$ and $-\mathbf{E}\left[\boldsymbol{H}\left(\boldsymbol{w}_t,\boldsymbol{\theta}_0\right)\right]=\mathbf{E}\left[\boldsymbol{s}\left(\boldsymbol{w}_t,\boldsymbol{\theta}_0\right)\boldsymbol{s}\left(\boldsymbol{w}_t,\boldsymbol{\theta}_0\right)'\right]$ is non singular; and local dominance condition for the Hessian; where $\boldsymbol{s}\left(\boldsymbol{w}_t,\boldsymbol{\theta}\right)=\frac{\partial\boldsymbol{m}\left(\boldsymbol{w}_t,\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}$ and $\boldsymbol{H}\left(\boldsymbol{w}_t,\boldsymbol{\theta}_0\right)=\frac{\partial\boldsymbol{s}\left(\boldsymbol{w}_t,\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}'}$

equation) restrictions of the type $\boldsymbol{g}(\boldsymbol{p}^{\star}, \mathrm{vec}(\boldsymbol{\Sigma}^{\star})) = \mathbf{0}$ are allowed[7]. We will follow the Qu and Perron (2007) strategy adapted to our model.

To establish theoretical results about the consistency and limit distribution of the estimates of the break dates, some standard conditions on the asymptotic framework and on the break dates must be adopted. We also assume here that the break dates are asymptotically distinct (A5); and some conditions under which the breaks are asymptotic nonnegletable (A6). More precisely we consider the following Assumptions.

Assumption A5. *The following inequalities hold* $0 < \lambda_1^0 < \cdots < \lambda_m^0 < 1$ *and* $T_i^0 = [T\lambda_i^0]$, *where* $[\cdot]$ *denotes the greatest integer for* $i = 0, \cdots, m+1$.

Assumption A6. *The magnitudes of the shifts satisfy* $\Delta \boldsymbol{p}_j^0 = \boldsymbol{p}_j^0 - \boldsymbol{p}_{j-1}^0 = v\boldsymbol{\delta}_j$, $\Delta\boldsymbol{\Sigma}_j^0 = \boldsymbol{\Sigma}_j^0 - \boldsymbol{\Sigma}_{j-1}^0 = v\boldsymbol{\Phi}_j$, *for* $(\boldsymbol{\delta}_j, \boldsymbol{\Phi}_j) \neq 0$, *and do not depend on* T. v *is a positive quantity independent of* T*; or we have* $v \to 0$ *but* $\sqrt{T}v/(logT)^2 \to \infty$, *being* v *a sequence of positive numbers.*

Assumption A6 states that the magnitudes of the shifts can be either fixed (v is a positive number independent of T) capturing large shifts asymptotically or shrinking (v shrinks and $\sqrt{T}v/(logT)^2 \to \infty$) corresponding to small shifts in finite samples. It is worth noting that the assumption on the nature of the magnitudes does not impact the test itself or its asymptotic distribution, as we will show later.[8] Research by Perron (2006) discusses in length the implications of considering fixed shifts or shrinking shifts.

As noted by Qu and Perron (2007) these Assumptions ensure four important results regarding the estimation process of the break dates. First, T_j being unknown does not change the distribution of the estimators. Formally, under Assumptions A1 to A6 the limiting distribution of $\sqrt{T}\left(\hat{\boldsymbol{p}}_j - \boldsymbol{p}_j^0\right)$ is the same regardless of whether the breaks occur at known or at unknown dates, hence the asymptotic normality and consistency of the estimator of the probabilities is established.

Second, the convergence rates of the estimators are as follows:

$$\begin{aligned} Tv^2\left(\hat{T}_j - T^0\right) &= O_p(1), \text{ for } j = 1, \cdots, m && (2.27)\\ \sqrt{T}\left(\hat{\boldsymbol{p}}_j - \boldsymbol{p}_j^0\right) &= O_p(1), \text{ for } j = 1, \cdots, m+1 && (2.28)\\ \sqrt{T}\left(\hat{\boldsymbol{\Sigma}}_j - \boldsymbol{\Sigma}_j^0\right) &= O_p(1), \text{ for } j = 1, \cdots, m+1 && (2.29) \end{aligned}$$

which means that the break dates (or the break fractions) converge faster than $\hat{\boldsymbol{p}}_j$, such that the asymptotic distribution of the latter is not affected by the former.

Third. As a consequence of the last result, that the maximazation of the likelihood function might be done in a subset of the parameter set C_M and in a neighbourhood of the respective true values, such that:

$$\begin{aligned} C_M = &\Big\{\left(\boldsymbol{\mathcal{T}}, \boldsymbol{p}_j, \boldsymbol{\Sigma}_j\right) : v^2\left|T_j - T_j^0\right| \leq M \text{ for } j = 1, \cdots, m, \\ &\left|\sqrt{T}\left(\boldsymbol{p}_j - \boldsymbol{p}_j^0\right)\right| \leq M, \left|\sqrt{T}\left(\boldsymbol{\Sigma}_j - \boldsymbol{\Sigma}_j^0\right)\right| \leq M \text{ for } j = 1, \cdots, m+1\Big\} \end{aligned} \quad (2.30)$$

where M, which can be set to be arbitrarily large, denotes the maximum number of breaks allowed.

[7] Those restrictions might play an important role because, among other cases, they permit the generation of partial change models (only a subset of coefficients are allowed to change). This class of models allows inhomogeneities to be tested in just a few lines of the matrix of transition probabilities, while the rest remain homogeneous. Furthermore, the imposition of relevant restrictions in the parameters of the model might also contribute to power increments.

[8] On the contrary, this assumption only influences some minor technical issues related to the distribution of the break dates.

Finally, the log-likelihood ratio might be decomposed into two asymptotically independent components - one that concerns the estimation of the break dates and another one that refers to the estimation of the stacked vector of parameters $\boldsymbol{\theta}$. Let lr and rlr denote, respectively, the log likelihood ratio and the restricted likelihood ratio, such that the objective function is:

$$rlr = lr + \boldsymbol{\lambda}' \boldsymbol{g} \left(\boldsymbol{p}^{\star}, \text{vec} \left(\boldsymbol{\Sigma}^{\star} \right) \right) \tag{2.31}$$

Under Assumptions A1 to A6, rlr may be decomposed as follows:

$$\begin{aligned} \max_{(\boldsymbol{\mathcal{T}}, \boldsymbol{p}, \boldsymbol{\Sigma}) \in C_M} rlr &= \max_{\boldsymbol{\mathcal{T}} \in C_M, \boldsymbol{p}^0, \boldsymbol{\Sigma}^0} \sum_{j=1}^{m} lr_j^{(1)} \left(T_j - T_j^0 \right) && (2.32) \\ &+ \max_{(\boldsymbol{p}, \boldsymbol{\Sigma}) \in C_M, \boldsymbol{\mathcal{T}}^0} \left[\sum_{j=1}^{m+1} lr_j^{(2)} + \boldsymbol{\lambda}' \boldsymbol{g} \left(\boldsymbol{p}, \boldsymbol{\Sigma} \right) \right] && (2.33) \\ &+ o_p(1) \end{aligned}$$

where,

$$\begin{aligned} lr_j^{(1)}(0) &= 0, \\ lr_j^{(1)}(r) &= \frac{1}{2} \sum_{t=T_j^0+r}^{T_j^0} \boldsymbol{\varepsilon}_t' \left[\left(\boldsymbol{\Sigma}_j^0 \right)^{-1} - \left(\boldsymbol{\Sigma}_{j+1}^0 \right)^{-1} \right] \boldsymbol{\varepsilon}_t - \frac{r}{2} \left(\log \left| \boldsymbol{\Sigma}_j^0 \right| - \log \left| \boldsymbol{\Sigma}_{j+1}^0 \right| \right) \\ &- \frac{1}{2} \sum_{t=T_j^0+r}^{T_j^0} \left(\boldsymbol{p}_j^0 - \boldsymbol{p}_{j+1}^0 \right)' \boldsymbol{x}_t \left(\boldsymbol{\Sigma}_{j+1}^0 \right)^{-1} \boldsymbol{x}_t' \left(\boldsymbol{p}_j^0 - \boldsymbol{p}_{j+1}^0 \right) \\ &- \sum_{t=T_j^0+r}^{T_j^0} \left(\boldsymbol{p}_j^0 - \boldsymbol{p}_{j+1}^0 \right)' \boldsymbol{x}_t \left(\boldsymbol{\Sigma}_{j+1}^0 \right)^{-1} \boldsymbol{\varepsilon}_t \end{aligned}$$

for $r = -1, -2, \cdots,$

$$\begin{aligned} lr_j^{(1)}(r) &= - \frac{1}{2} \sum_{t=T_j^0+1}^{T_j^0+r} \boldsymbol{\varepsilon}_t' \left[\left(\boldsymbol{\Sigma}_j^0 \right)^{-1} - \left(\boldsymbol{\Sigma}_{j+1}^0 \right)^{-1} \right] \boldsymbol{\varepsilon}_t - \frac{r}{2} \left(\log \left| \boldsymbol{\Sigma}_j^0 \right| - \log \left| \boldsymbol{\Sigma}_{j+1}^0 \right| \right) \\ &- \frac{1}{2} \sum_{t=T_j^0+1}^{T_j^0+r} \left(\boldsymbol{p}_j^0 - \boldsymbol{p}_{j+1}^0 \right)' \boldsymbol{x}_t \left(\boldsymbol{\Sigma}_{j+1}^0 \right)^{-1} \boldsymbol{x}_t' \left(\boldsymbol{p}_j^0 - \boldsymbol{p}_{j+1}^0 \right) \\ &- \sum_{t=T_j^0+1}^{T_j^0+r} \left(\boldsymbol{p}_j^0 - \boldsymbol{p}_{j+1}^0 \right)' \boldsymbol{x}_t \left(\boldsymbol{\Sigma}_{j+1}^0 \right)^{-1} \boldsymbol{\varepsilon}_t \end{aligned}$$

for $r = 1, 2, \cdots$, and

$$lr_j^{(2)} = \frac{1}{2}\sum_{t=T_j^0+1}^{T_j^0} \left(\boldsymbol{y}_t - \boldsymbol{x}_t'\boldsymbol{p}_j\right)' \boldsymbol{\Sigma}_j^{-1} \left(\boldsymbol{y}_t - \boldsymbol{x}_t'\boldsymbol{p}_j\right) - \frac{T_j^0 - T_{j-1}^0}{2} \log |\boldsymbol{\Sigma}_j| + \frac{1}{2}\sum_{t=T_j^0+1}^{T_j^0} \left(\boldsymbol{y}_t - \boldsymbol{x}_t'\boldsymbol{p}_j\right)' \left(\boldsymbol{\Sigma}_j^0\right)^{-1} \left(\boldsymbol{y}_t - \boldsymbol{x}_t'\boldsymbol{p}_j\right) + \frac{T_j^0 - T_{j-1}^0}{2} \log \left|\boldsymbol{\Sigma}_j^0\right|.$$

This result implies that the estimator of the break dates are asymptotic independent of the estimator of $\boldsymbol{\theta}$. Additionally, eventual restrictions on the parameters do not affect the distribution of the break dates. Moreover, the estimation procedure is as follows. Firstly, the break dates are consistently estimated assuming that we know the true values of the parameters $\boldsymbol{\theta}^0$, then the mean parameters are estimated, possibly subject to restrictions of the type $\boldsymbol{g}\left(\boldsymbol{p}^\star, \text{vec}\left(\boldsymbol{\Sigma}^\star\right)\right) = \boldsymbol{0}$, keeping the break dates fixed at their true values $\boldsymbol{\mathcal{T}}^0$. Thus, under fixed magnitudes of shifts, it is straightforward to derive the asymptotic distribution of the estimates on the break dates.[9]

Propositon 4. *Under Assumptions A1 to A6, assuming a fixed v, we have:*

$$\hat{T}_j - T_j^0 \xrightarrow{d} argmax_r lr_j^{(1)}(r) \;\; for \; j = 1, \cdots, m \tag{2.34}$$

Proof. See Qu and Perron (2007). ■

It is worth noting some considerations about the computational procedure underlying the estimation of the model parameters. A standard grid search procedure would require the computation of a number of QMLE of an order of magnitude of T^m, which would be virtually impossible with $m > 2$. However, the maximum number of possible segments is actually $\sum_{t=1}^{T} t = 1 + 2 + \cdots + T = \frac{T(1+T)}{2}$ and therefore of an order of magnitude of T^2, no matter the number of breaks m. Thus, a method is required to select which combination of segments leads to a minimum value of the objective function. Here, we follow the dynamic programming algorithm, based on a iterative GLS approach to evaluate the likelihood functions for all segments, proposed by Bai and Perron (2003) and extended by Qu and Perron (2007, pp 476-478).

2.4 Testing for multiple and endogenous inhomogeneities in Markov chains

In this section we consider the problem of testing for inhomogeneities in a Markov chain. Put otherwise, testing structural changes in the one-step transition probabilities occurring at unknown periods. Without any lost of generality we assume a pure structural change model, such that all parameters are allowed to vary over time. In a first moment we will expose the standard likelihood ratio test. This statistic intends to determine a presence of at least one structural break. Next, we discuss the potential of two possible extensions concerning confirmatory analysis. A sequential test that allows us to select the number of changes, given that, sequentially, we test a null of l breaks against $l+1$ breaks - the $Seq\left(\ell+1|\,\ell\right)$; and the $WDmax$ for testing no breaks against an unknown (up to some pre-specified maximum) number of breaks. These procedures are interesting because they do not require the prespecification of the number of breaks under the alternative hypothesis, unlike the $\sup LR$ test.

Nevertheless, the following two Assumptions, on the regressors and on the errors, must be adopted to obtain theoretical results about limiting distribution of the tests, under the null hypothesis of no breaks, $m = 0$.

[9] For the derivation of the asymptotic distribution of the break dates estimator under shrinking magnitudes of shifts see, e.g. Qu and Perron (2007, pp 471-472) or Bai (2000, p312)

Assumption A7. *We have* $T^{-1}\sum_{t=1}^{[Ts]}\boldsymbol{x}_t\boldsymbol{x}_t' \xrightarrow{p} sQ$ *uniformly in* $s\in[0,1]$ *for some* Q *positive definite.*

Assumption A7 imposes that the limit moment matrix of the regressors is homogeneous over all sample.

Assumption A8. *We have* $\mathbf{E}\left[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right]=\boldsymbol{\Sigma}^0\ \forall t$; $T^{-1/2}\sum_{t=1}^{[Ts]}\boldsymbol{x}_t\boldsymbol{\varepsilon}_t \xrightarrow{d} \boldsymbol{\Phi}^{1/2}\boldsymbol{W}(\boldsymbol{s})$ *where* $\boldsymbol{W}(\boldsymbol{s})$ *is a vector of independent Wiener processes, and*

$$\boldsymbol{\Phi}=plimT^{-1}\sum_{t=1}^{T}\boldsymbol{x}_t\left(I_{(q-1)}\otimes\boldsymbol{\Sigma}^0\right)\boldsymbol{x}_t'.$$

Also we have, $T^{-1/2}\sum_{t=1}^{[Ts]}\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t'-\boldsymbol{I}_{(q-1)}\right)\xrightarrow{d}\boldsymbol{\xi}(\boldsymbol{s})$, *where* $\boldsymbol{\xi}(\boldsymbol{s})$ *is a* $(q-1)$ *square matrix of Brownian motion processes with* $\boldsymbol{\Omega}=var\left(vec\left(\boldsymbol{\xi}(\mathbf{1})\right)\right)$; *and* $\boldsymbol{\eta}_t\equiv\left(\eta_{t1},\cdots,\eta_{t,q-1}\right)'=\left(\boldsymbol{\Sigma}^0\right)^{-1/2}\boldsymbol{\varepsilon}_t$. *In addition we assume that* $\mathbf{E}\left[\eta_{tk}\eta_{tl}\eta_{tm}\right]=0\ \forall k,l,m,t; k\neq m\vee k\neq l$.

Assumption A8 states that the functional central limit theorem might be employed. This is a mild weak condition, given that, by construction, under null the hypothesis of homogeneity, $\boldsymbol{x}_t\boldsymbol{\varepsilon}_t$ is a martingale difference sequence with respect to $\boldsymbol{\mathcal{F}}_t$. Moreover, $\boldsymbol{x}_t$ and $\boldsymbol{\varepsilon}_t$ are bounded which ensures the existence of all moments of $\boldsymbol{x}_t\boldsymbol{\varepsilon}_t$, in particular, the Corollary 29.19 of Davidson (1994) can be immediately applied. Both Assumption A7 and A8 are crucial to the derivation of the limiting distribution of the tests under H_0.

Regarding the limiting distribution of the $\sup LR(m,q,\varepsilon)$ test, let

$$\tilde{\boldsymbol{p}}=\left(\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t'\right)^{-1}\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{y}_t \tag{2.35}$$

denote the transition probabilities estimator under the null hypothesis of homogeneity (absence of structural breaks) and let $\widetilde{L}$ denote the associated likelihood, where

$$\log\widetilde{L}=-\frac{T(q-1)}{2}(\log 2\pi+1)-\frac{T}{2}\log\left|\tilde{\boldsymbol{\Sigma}}\right|, \tag{2.36}$$

being $\tilde{\boldsymbol{\Sigma}}$ the error covariance matrix under homogeneity, verifying

$$\tilde{\boldsymbol{\Sigma}}=\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{y}_t-\boldsymbol{x}_t'\tilde{\boldsymbol{p}}\right)\left(\boldsymbol{y}_t-\boldsymbol{x}_t'\tilde{\boldsymbol{p}}\right)'. \tag{2.37}$$

Let additionaly $\log\hat{L}\left(T_{1,}\cdots,T_{m,}\right)$ denote the log likelihood statistic for a given partition $\boldsymbol{\mathcal{T}}$ associated with the QML (2.18).

The observable value for this test is the supremum of the likelihood ratio:

$$2\left[\log\hat{L}\left(T_{1,}\cdots,T_{m,}\right)-\log\widetilde{L}\right] \tag{2.38}$$

evaluated over all possible partitions Λ_ε (expression 2.24). Formally, we have

$$\begin{aligned}\sup LR(m,q,\varepsilon) &= \sup_{(\lambda_1,\cdots,\lambda_m)\in\boldsymbol{\Lambda}_\varepsilon} 2\left[\log\hat{L}\left(T_1,\cdots,T_m\right)-\log\widetilde{L}\right]\\ &= 2\left[\log\hat{L}\left(\hat{T}_1,\cdots,\hat{T}_m\right)-\log\widetilde{L}\right],\end{aligned} \tag{2.39}$$

where $\hat{T}_1, \cdots, \hat{T}_m$ results from the maximization (2.32). This statistic depends on three parameters: i) the number of breaks allowed, m ; ii) the trimming parameter, ε; and iii) the dimension of the space state of the Markov chain (the number of states). The critical values are presented in Bai and Perron (1998, 2003). Regarding the limiting distribution of the $\sup LR(m, q, \varepsilon)$ statistic under the null hypothesis of homogeneity, Proposition 5 holds.

Propositon 5. *Under Assumptions A1 to A8 the limiting distribution of the* $\sup LR$ *statistic is as follows.*

$$\sup LR \xrightarrow{d} \sup \sum_{j=1}^{m} \frac{\left\|\lambda_j W_q\left(\lambda_{j+1}\right) - \lambda_{j+1} W_q\left(\lambda_{j+1}\right)\right\|^2}{\left(\lambda_{j+1} - \lambda_j\right) \lambda_j \lambda_{j+1}} \tag{2.40}$$

where $W_q(\cdot)$ *denote a q-dimensional vector of independent Wiener processes.*

Proof. See Qu and Perron (2007, p. 487). ■

We extend our analysis to a sequential test procedure. Actually, the $Seq\left(\ell+1 \mid \ell\right)$ statistic proposed by Bai and Perron (1998) to the univariate case and adapted by Qu and Perron (2007) to the multivariate case, can be used to select the number of different segments in an inhomogeneous Markov chain. Let us consider a model with ℓ breaks, whose estimates, $(\hat{T}_1, \cdots, \hat{T}_\ell)$, were obtained by a global maximization procedure. This statistic tests, sequentially, $H_0 : \ell$ breaks against $H_1 : \ell+1$ breaks by performing a single-break test for for each one of the segments $(\hat{T}_1, \cdots, \hat{T}_\ell)$ and then evaluating the significance of the maximum of the tests. Formally we have:

$$Seq\left(\ell+1 \mid \ell\right) = \max_{1 \le j \le \ell+1} \sup_{\tau \in \boldsymbol{\Lambda}_{j,\varepsilon}} lr\left(\hat{T}_1, \cdots, \hat{T}_{j-1}, \tau, \hat{T}_j, \cdots \hat{T}_\ell\right) - lr\left(\hat{T}_1, \cdots, \hat{T}_\ell\right), \tag{2.41}$$

where

$$\boldsymbol{\Lambda}_{j,\varepsilon} = \left\{\hat{T}_{j-1} + \left(\hat{T}_j - \hat{T}_{j-1}\right)\varepsilon \le \tau \le \hat{T}_j - \left(\hat{T}_j - \hat{T}_{j-1}\right)\varepsilon\right\}. \tag{2.42}$$

The limiting distribution of this statistic under the null of ℓ breaks can be found in Qu and Perron (2007) and the critical values were tabulated in Bai and Perron (1998, 2006).

An important feature concerning the $\sup LR(m, q, \varepsilon)$ statistic relates to the need to specify a priori the number of breaks to be tested, m, under the alternative hypothesis. That is frequently not the case and one may still not want to specify the number of breaks under the alternative hypothesis. For this purpose, Bai and Perron (1998) suggested a class of tests called *double maximum tests*. One of these statistics is the $WDmaxLR(M)$, that tests for a null of no breaks against an unknown number of breaks between 1 and some upper limit M. The mechanic of this test consists on the evaluation of the maximum of the supremum of the $\sup LR(m, q, \varepsilon)$ over the number of possible breaks from 1 to M, as follows

$$WDmaxLR(m, q, \varepsilon)\left(M, q\right) = \max_{1 \le m \le M} a_m \times \sup_{(\lambda_1, \cdots, \lambda_m) \in \Lambda_\varepsilon} LR(m, q, \varepsilon). \tag{2.43}$$

The terms a_m are the weights of the test,[10]

$$a_m \equiv \frac{c\left(q, \alpha, 1\right)}{c\left(q, \alpha, m\right)}, \tag{2.44}$$

the ratio between the asymptotic critical values of the $\sup LR(m, q, \varepsilon)$ for $m = 1$ $\left(c\left(q, \alpha, 1\right)\right)$ and for $m = 1, \cdots, M$ $\left(c\left(q, \alpha, m\right)\right)$, so that $a_1 = 1$. Critical values might be found in Bai et al. (1998) and in Bai and Perron (2006).

[10]An unweighted version of this test, the $UDmaxLR(m, q, \varepsilon)$, reported poor power, mainly as m increases (Bai et al., 1998; Bai and Perron, 2006).

3 Monte Carlo Experiments

In this section we evaluate the size of the tests through a Monte Carlo experiment. We consider a simple process with two categories S_t with three states ($q = 3$). The Markov chain is simulated, using the GAUSS package, according to the algorithm:

1. Define $m + 1$ q-dimensional transition probability matrices whose elements are the probabilities

$$P_{j,i_1 i_0} \equiv P_j \left(S_t = i_0 | \, S_{t-1} = i_1 \right) \tag{3.1}$$

   - see the definition of the data generating process (DGP) below;

2. Initialize the process $\{S_t\}$ by assigning values to $S_{T_{j-1}}$, for $j = 1, \cdots, m$ accordingly to the stationary distributions $\mathbf{\Pi}_j$;

3. Simulate a continuous random variable W uniformly distributed, $W \sim U(0, 1)$;

4. Given the initial values $S_{T_{j-1}}$ (step 2), simulate the process $\{S_t\}$, $t = T_{j-1} + 1, .., T_j$, for $j = 1, \cdots, m + 1$ as follows:

   (a) Let $P_{i,j} \equiv P_j \left(S_t = i | \, S_{1t-1} = i_1 \right)$, $t = T_{j-1} + 1, .., T_j; j = 1, \cdots, m + 1$;

   (b) $S_t = \begin{cases} 1 & \text{if} \quad 0 \leq W < P_{1,j} \\ 2 & \text{if} \quad P_{1,j} \leq W < P_{1,j} + P_{2,j} \\ 3 & \text{if} \quad P_{1,j} + P_{2,j} \leq W < 1; \end{cases}$

5. Repeat the steps 1-4 until $t = T_{m+1}$.

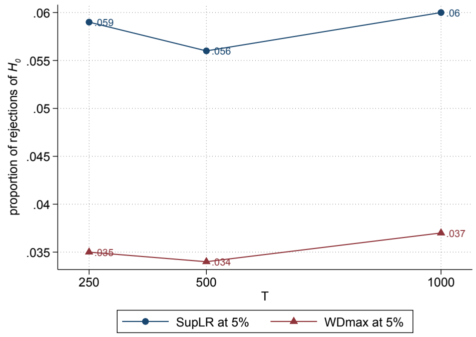
In all simulations, computed with 5000 replications, a trimming parameter is set to be $\varepsilon = 0.15$. Sample sizes of 250, 500 and 1000 were considered. The proportion of times when the null hypothesis was rejected is reported. Several levels of persistency of the regimes of the Markov chain are analysed. The analysis is also extended to the situation where the regimes are independent.

3.1 Size Analysis

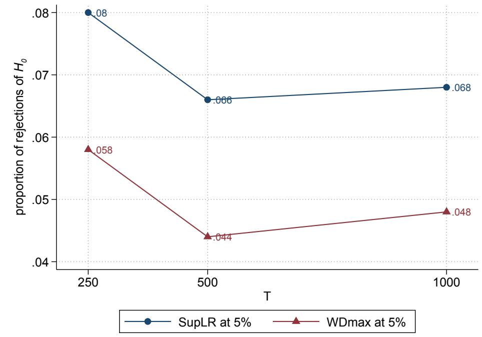
We now examine three different DGP (i.e. three different cases) for homogeneous Markov chains, $m = 1$, so that $T_0 = 1$ and $T_1 = T$ as follows. The regimes are persistent in Case 1, and not persistent in Case 2. Finally, the regimes are independent in Case 3.

Case	P_{11}	P_{12}	P_{13}	P_{21}	P_{22}	P_{32}	P_{31}	P_{32}	P_{33}
1	0.4	0.3	0.3	0.2	0.4	0.4	0.6	0.2	0.2
2	0.55	0.25	0.2	0.25	0.55	0.2	0.2	0.25	0.55
3	0.5	0.3	0.2	0.5	0.3	0.2	0.5	0.3	0.2

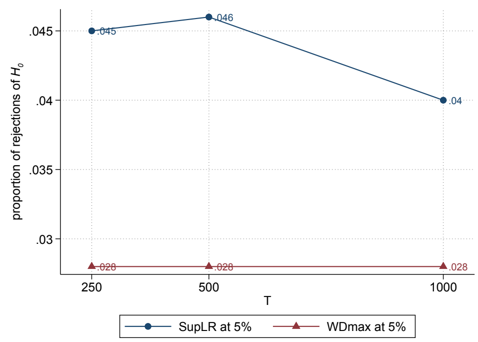
Next, we consider two cases: (1) one break is allowed, (2) two breaks are allowed. In the first case both the $\sup LR$ and the $WDmax$ tests 0 breaks against 1 break. To simplify the notation we consider $\sup LR(0 \mid 1)$ and $WDmax(0 \mid 1)$. In the second situation the $\sup LR$ tests 0 breaks against 2, $\sup LR(0 \mid 1)$, and the $WDmax$ 0 against an undetermined number of breaks up to 2, $WDmax(0 \mid 1, 2)$. Figures 1 and 2 display the results. In general, there appear to be no size distortions of the tests, as expected. The exception is the $\sup LR(0 \mid 2)$ for the Case 2, where the regimes are persistent, that tends to be slightly oversized (Figure 2b). It should also be noted that $WDmax$ tends to present an undersize behaviour in all situations, in that it tends to under-reject the null hypothesis.



(a) Case 1

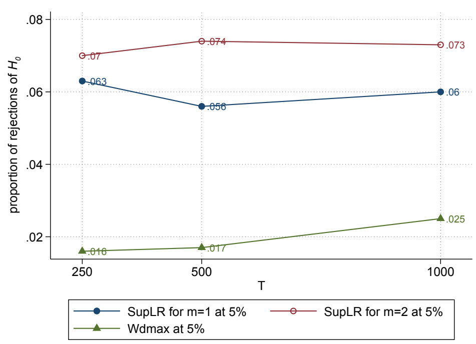


(b) Case 2

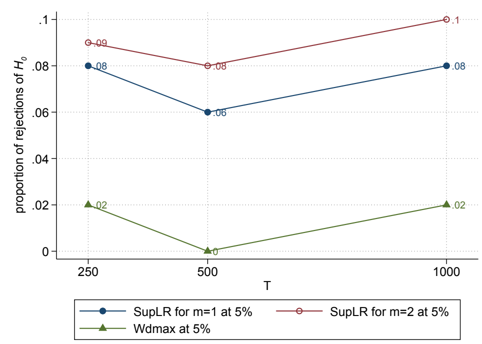


(c) Case 3

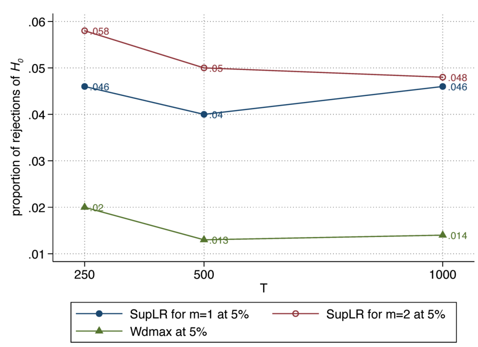
Figure 1: Homogeneous Markov chain, one break is allowed (nominal size: 5%)



(a) Case 1



(b) Case 2



(c) Case 3

Figure 2: Homogeneous Markov chain, up to two breaks are allowed (nominal size: 5%)

3.2 Power Analysis

To access the power of the tests either in finite samples and asymptotically, our analysis includes six cases for inhomogeneous Markov chains. Cases 4, 5, 6 correspond to Markov chains generated with one structural break $m = 1$, so that $T_0 = 1$, $T_1 = [T/2]$ and $T_2 = T$. Cases 7, 8, 9 concern chains generated subject to three breaks ($T_0 = 1$, $T_1 = [T/3]$, $T_2 = [2T/3]$ and $T_3 = T$). As follows:

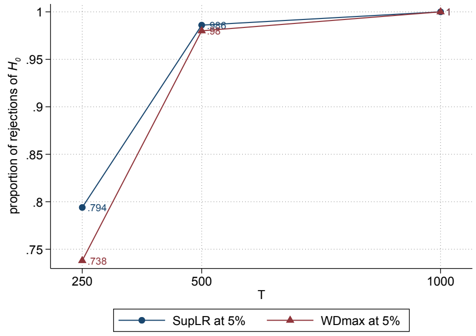
Case	Regime	P_{11}	P_{12}	P_{13}	P_{21}	P_{22}	P_{32}	P_{31}	P_{32}	P_{33}
4	**1**	0.4	0.3	0.3	0.2	0.4	0.4	0.6	0.2	0.2
4	**2**	0.2	0.4	0.4	0.2	0.6	0.2	0.6	0.2	0.2
5	**1**	0.55	0.25	0.2	0.25	0.55	0.2	0.2	0.25	0.55
5	**2**	0.45	0.2	0.35	0.3	0.45	0.25	0.15	0.40	0.45
6	**1**	0.5	0.3	0.2	0.5	0.3	0.2	0.5	0.3	0.2
6	**2**	0.3	0.4	0.3	0.3	0.4	0.3	0.3	0.4	0.3
7	**1**	0.4	0.3	0.3	0.2	0.4	0.4	0.6	0.2	0.2
7	**2**	0.2	0.4	0.4	0.2	0.6	0.2	0.3	0.4	0.4
7	**3**	0.4	0.3	0.3	0.2	0.4	0.4	0.6	0.2	0.2
8	**1**	0.55	0.25	0.2	0.25	0.55	0.2	0.2	0.25	0.55
8	**2**	0.45	0.2	0.35	0.3	0.45	0.25	0.15	0.4	0.45
8	**3**	0.55	0.25	0.2	0.25	0.55	0.2	0.2	0.25	0.55
9	**1**	0.5	0.3	0.2	0.5	0.3	0.2	0.5	0.3	0.2
9	**2**	0.3	0.4	0.3	0.3	0.4	0.3	0.3	0.4	0.3
9	**3**	0.5	0.3	0.2	0.5	0.3	0.2	0.5	0.3	0.2

In general, the tests are consistent in the sense that the respective power tends to 1. The exception is Case 8, where the regimes exhibit some persistency and there are three different regimes, in that the $\sup LR(0 \mid 1)$ and the $WDmax(0 \mid 1, 2)$ presents low asymptotic power. Considering only two regimes, both the power of the $\sup LR(0 \mid 1)$ and of the $WDmax(0 \mid 1, 2)$ is satisfactory, even with high persistency.

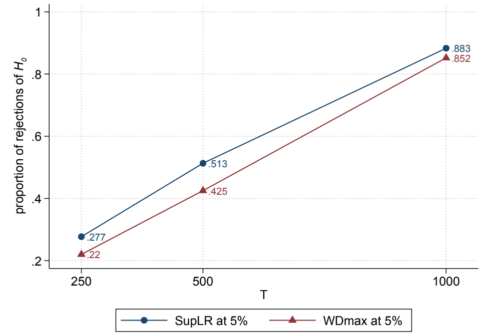
As expected, with one break (Cases 4, 5, 6) the $\sup LR(0 \mid 1)$ always performed better than the $WDmax(0 \mid 1)$. When two breaks are present, the $\sup LR(0 \mid 2)$ outperformed the $WDmax(0 \mid 1, 2)$, however the $WDmax(0 \mid 1, 2)$ surpassed the $\sup LR(0 \mid 1)$. This suggests that if we are not sure about the number of breaks the $WDmax$ might be a good option.

Furthermore, the $Seq(1 \mid 2)$ behaves as expected. It tends to correctly reject 1 break against 2 in all situations, both in small and in large samples, except for the case of highly persistent regimes (Figure 3b).

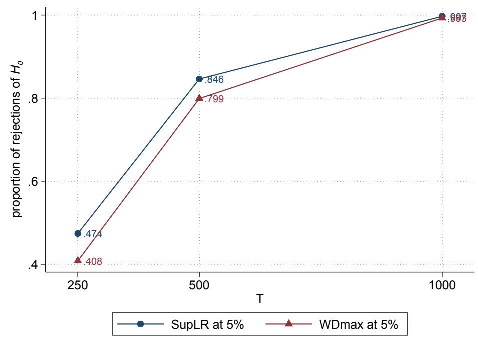
In small samples, with persistent regimes and in the presence of two breaks the tests tend to evidence lack of power, notably the $\sup LR(0 \mid 1)$ and the $WDmax(0 \mid 1, 2)$, Figure 4b. In turn, in finite samples with two breaks, against an alternative of one single break the null hypothesis of no break is only rejected about 18.3% of the time by the $\sup LR(0 \mid 2)$ and around 20% by the $WDmax(0 \mid 1, 2)$, which is far from being positive. However, the DGP was quite extreme, as the induced magnitude of the shifts is relatively small, the regimes are persistent in all of the three segments, and DGP of the first segment equals the third one. When two breaks are present, the $\sup LR(0 \mid 2)$ always present a good power, larger than the $WDmax(0 \mid 1, 2)$ in all circumstances. This is the expected result, given that H_1 fully specified $\sup LR(0 \mid 2)$. In a nutshell, the higher the persistency of the regimes the lower the power of the tests. It must, however, be pointed out that the multiple breaks have been induced in such a



(a) Case 4



(b) Case 5

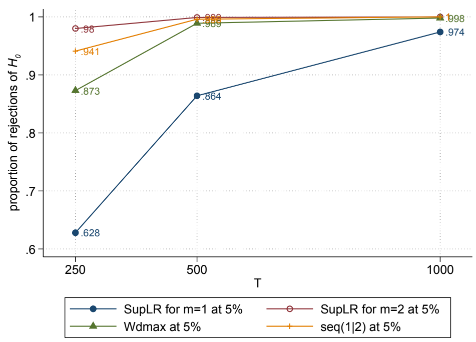


(c) Case 6

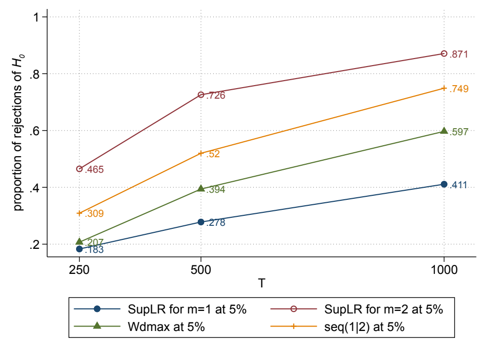
Figure 3: Inhomogeneous Markov chain with two segments, one break is allowed (nominal size: 5%)

way that they are difficult to detect: two breaks with the first and third regimes the same and small magnitudes (Perron, 2006, p 32).

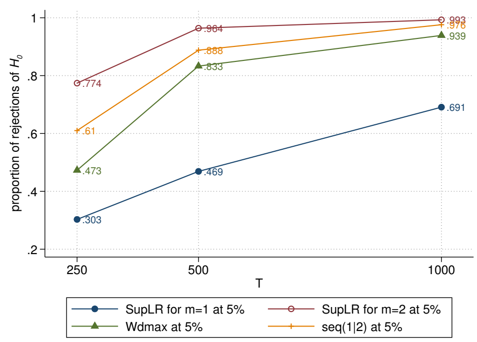
A practical recommendation of this exercise may involve the joint use of the various tests to detect and test inhomogeneities in a Markov chain. In fact, one can use the $\sup LR(0 \mid m)$ for several levels of m, in conjunction with a confirmatory analysis with the other tests. The $WDmax$ may be used with an arbitrary large M to confirm if there is at least one structural break, then the $Seq(\ell \mid \ell + 1)$ can be employed to corroborate the number of breaks of the process.



(a) Case 7



(b) Case 8



(c) Case 9

Figure 4: Inhomogeneous Markov chain with three segments, two breaks are allowed (nominal size: 5%)

4 Example

In this section, which is far from being the main focus of this article, we provide a brief illustration of the method with an application on the NASDAQ stock index. Weekly data from January 1990 to January 2019 has been considered, with 1517 observations. The purpose of We start by considering a simple AR(1) on NASDAQ returns ($r_t = 100 \times \log(P_t/P_{t-1})$, where P_t denotes the closing prices), as follows:

$$r_t = \alpha + \phi r_{t-1} + \varepsilon_t.$$

Although the process is univariate, for the sake of comparison we apply Qu and Perron (2007) approach for detecting structural breaks. We set $M = 3$ so that three breaks and four regimes are allowed. There is no trace of structural breaks (Table 1). A trimming parameter $\varepsilon = 0.15$ was considered.

This lack of statistical success in detecting structural breaks may be due to three reasons. First, no breaks exist and hence they are not correctly detected. Second, the breaks exist somewhere in the higher moments of the probabilistic structure of the process, but not at its conditional mean. Third,

Test	Test obs	1% Critical Value	5% Critical Value	10% Critical Value
$\sup LR(1)$	4.005	15.711	12.092	10.434
$\sup LR(2)$	5.047	23.953	19.258	17.086
$\sup LR(3)$	6.273	31.007	25.280	22.620
$WDmaxLR(3)$	4.005	16.743	12.715	10.892

Table 1: Structural Break Tests for the Original Process (AR(1) approach)

Test	Test obs	1% Critical Value	5% Critical Value	10% Critical Value
$\sup LR(1)$	**50.883**	42.687	35.878	33.025
$\sup LR(2)$	**151.469**	68.324	60.155	57.066
$\sup LR(3)$	**187.179**	93.365	83.612	80.193
$WDmaxLR(3)$	**90.338**	45.375	38.618	36.098
$Seq\,(2\|\,1)$	**63.912**	135.772	31.734	17.811
$Seq\,(3\|\,2)$	**36.239**	129.198	27.787	13.765

Table 2: Structural Break Tests for the Reconstructed Process (Markov Chain Approach)

there are non-linearities in the conditional mean of the process that are not captured by a simple AR(1) but may be captured by the Markov Chain, i.e., the conditional mean was misspecified by the AR(1).

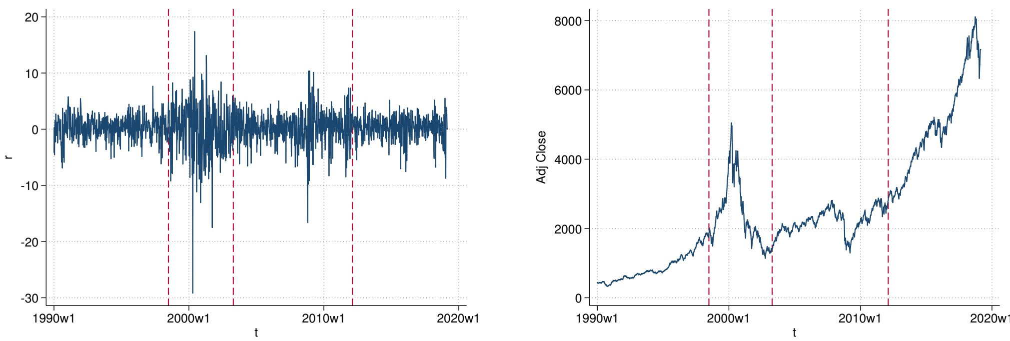
To investigate this, the original process was reconstructed into a 5-state Markov chain, in accordance with the following rule:

$$S_t = \begin{cases} 1 \text{ if } & r_t < \hat{q}_{0.2} \\ 2 \text{ if } & \hat{q}_{0.2} < r_t < \hat{q}_{0.4} \\ 3 \text{ if } & \hat{q}_{0.4} < r_t < \hat{q}_{0.6} \\ 4 \text{ if } & \hat{q}_{0.6} < r_t < \hat{q}_{0.8} \\ 5 \text{ if } & r_t > \hat{q}_{0.8} \end{cases} \tag{4.1}$$

where $\hat{q}_\alpha$ is the estimated quantile of order α of the marginal distribution of r_t. Then, the Markov chain was represented in a VAR form, see equation 2.9, and the proposed method was applied.

Unlike the AR(1) we have evidence of 3 structural breaks: 22/06/1998 (1998w25); 07/04/2003 (2003w14); and 23/01/2012 (2013w4). The results are summarised in Table 2. Figure 5 shows the breaks, both in the Prices (Figure 5b) and in the Returns (Figure 5a), represented by a vertical dash line. The identified regimes are consistent with NASDAQ's historical dynamics. On the one hand, with regard to the first segment, 1990-1998, it was a calm period, with increasing prices and a slightly volatile returns. This regime has some similarities with the Regime 4 - 2012 onwards, after the end of the subprime crisis. On the other hand, Regime 2 (1998-2003) captures the rise and fall of the NASDAQ that featured the *Dotcom era*. This period is turbulent, where very fast and asymmetric movements in Prices - in 1999 NASDAQ became, for a while, the world's largest stock market by dollar volume - are coupled with extremely volatile returns. Similarly, Regime 3 is also turbulent and unstable as it coincides with the financial and subprime crisis.

In each segment, the Markov chain state space is (obviously) finite and irreducible (all states communicate). Hence, the Markov chain in each segment is positive recurrent and thus has a stationary distribution, see for example, Karlin and Taylor (1981). Figure 6 displays the four transition matrices. Each figure represents a regime. For example, in figure 6a, Panel P(1|i0) represents the first row of P_1, Panel P(2|i0) represents the second row of P_1, and so on. In general, panel P(j|•) represents the j-th row of each P_i, $i = 1, 2, 3, 4$. The regimes can be compared vertically, i.e., Panel by Panel. For example, Panels P(1|i0) of Figures 6a and 6b report the probabilities of the first and second second regime, respectively, when $S_{t-1} = 1$. The main conclusion from Figures 6a to 6d is that the regimes behave quite differently, especially if we compare the regimes related to the bull periods (regimes 1 and 4) with those of bear periods (regimes 2 and 3). Some examples of this fact: (i) Panel P(1|i0) of Figure 6a (first regime) shows that the probabilities $P\left(S_t = i \middle| S_{t-1} = 1\right)$ for $i = 1, 2, ..., 5$, are approximately equal; in contrast, Panel P(1|i0) of Figure 6b (second regime) shows that these same probabilities present a U



(a) NASDAQ Returns

(b) NASDAQ Prices

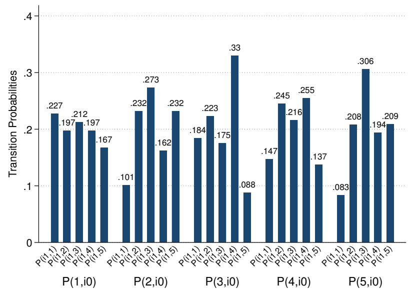
Figure 5: Structural Breaks from the Markov chain

shape. This U shape implies, for example, that if prices dropped in the previous period ($S_{t-1} = 1$, that is, if the return is below the 0.10 quantile), then the probability of the prices falling in the same magnitude ($S_t = 1$), or skyrocketing in the next period, is relatively high. This result was expected given that regime 2 corresponds to the period of financial crises and high market volatility. (ii) To reinforce the previous point, we note that all the panels associated with the second regime present the U shape. (iii) in the regime 4 (Figure 6d), which can be identified as the bull period, the market quickly changes from a price fall to a price increase (see panel Panel P(1|i0) of Figure 6d).

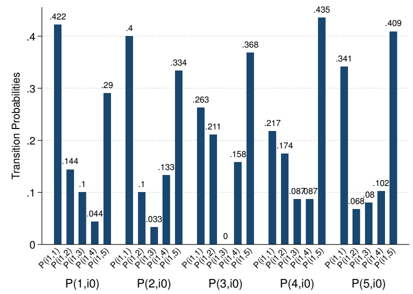
A way to summarize some of the main regime dynamic features is to compute the first two conditional moments. In general, the Markov chain does not allow us to calculate such moments, if the states are (nominal) categories. However, in our case, the 5 states represent a class interval of type $\left[\hat{q}_{(k-1)0.2}, \hat{q}_{k0.2}\right]$, where $\hat{q}_\alpha$ is the quantile of order α, and $k = 1, 2, ..., 5$. Hence, we may consider a representative value for each interval, e.g. the corresponding midpoint, say m_k, and then estimate the approximate conditional mean and variance on state $S_{t-1} = j$ as follows:

$$
\begin{aligned}
\hat{\mu}\left(r, j\right) &= \sum_{k=1}^{5} m_k \times \hat{P}_r\left(k \middle| S_{t-1} = j\right) \\
\widehat{\sigma}^2\left(r, j\right) &= \sum_{k=1}^{5} m_k^2 \times \hat{P}_r\left(k \middle| S_{t-1} = j\right) - \hat{\mu}\left(j\right)^2
\end{aligned}
$$

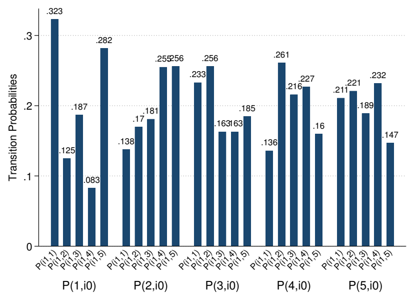
This procedure is also suggested in Nicolau (2014). Figure 7 shows the conditional mean and variances across the various regimes. The conditional moments' differences across regimes are quite obvious. For example, regime 4 has a conditional mean on $S_{t-1} = 1$ quite high (this moment is represented by $\hat{\mu}\left(4, 1\right)$); it means that a significant drop in the previous period tends to lead to a sharp increase in the following period. In other words, the market tends to recover rapidly after a significant drop. The opposite occurs with regime 1 and to some extend with regime 2 and 3. The conditional variance also displays quite different behavior. For example, the conditional variance of regime 2 is the highest among the four, regardless of the state of the process in the previous period (i.e., $\widehat{\sigma}^2\left(r, j\right)\left(2, \bullet\right) > \widehat{\sigma}^2\left(r, j\right)\left(r, \bullet\right)$, $r = 1, 3, 4$). In regime 2 (financial crisis) the conditional mean is generally lower and volatility is generally higher than in the other regimes, as shown in figure 2. This is expected and has been extensively described in the extant literature. What is more subtle, but also clear from Figures 6 and 7, is that the market in regime 2 does not recover easily after a fall, since $P\left(S_t = 1 \middle| S_{t-1} = 1\right)$ is relatively high (see for example Panel P(1|i0) of Figure 6b and Figure 7a); furthermore, it is also interesting to observe that $P\left(S_t = 3 \middle| S_{t-1} = 3\right)$ is very low when compared to other regimes, which means that the process



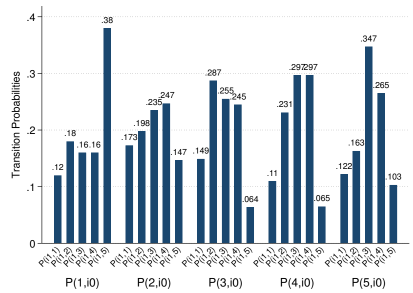
(a) Regime 1



(b) Regime 2

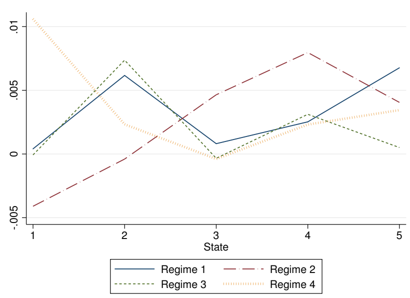


(c) Regime 3

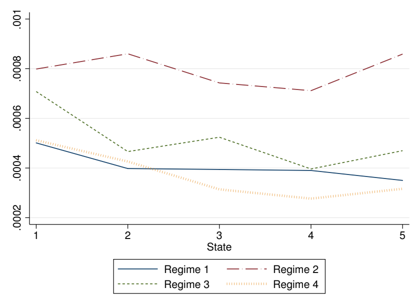


(d) Regime 4

Figure 6: Transition Matrices



(a) Conditional Mean



(b) Conditional Variance

Figure 7: Conditional Mean and Variance

in regime 2 does not spend much time in the "quiet" state $S_t = 3$ (this state includes all the values of returns that lie between the quantiles 0.4 and 0.6).

There are more features that emerge from figures 1 and 2, but the point is the following: the dissimilarities between the transition probability are clear, which confirm the structural breaks uncovered by the proposed methodology. By contrast, the standard autoregressive model is unable to detect structural changes. This may be attributed to the fact that returns usually display very weak linear autocorrelations in any regime, and nonlinearities of any form cannot be detected by a vanilla AR process. On the contrary, Markov processes (with a large number of states) can potentially detect any form of nonlinearity of a Markov process of order one.

5 Conclusions

This article proposes a new approach for detecting and testing inhomogeneities in Markov chains occurring at unknown periods. The usual methods described in the literature for testing inhomogeneities in Markov chains have some limitations. Namely that they can only test for the presence of a single structural break; and either the break is assumed to occur at a known date or the limiting distribution of the test is unknown. Our strategy relies on the fact that, under certain conditions, an ergodic Markov chain admits a stationary vectorial autoregressive representation. The numerical equivalence between the MLE estimator for the one-step transition probabilities and the VAR mean parameter estimators is proved. Taking advantage of the possibility of representing a Markov chain in VAR form, the methods that usually apply to a VAR model remain valid for the Markov chains, namely, the $\sup LR$, the $WDmax$, and the $Seq\,(\ell+1|\,\ell)$. These procedures are applied for the first time to Markov chains. A Monte Carlo simulation study points to a higher power of $\sup LR$ tests compared to $WDmax$ tests when the alternative hypothesis is correctly specified. This evidence occurs with one and with two structural breaks. As for size analysis, with one and with two breaks, there were no size distortions in either small or large samples, for all tests. An illustration of the proposed method using the NASDAQ stock index was also provided. The advantages of the proposed methodology were demonstrated for the reason that, unlike the *traditional* AR(1), it is possible to detect breaks that make sense in light of the historical dynamics of the NASDAQ. Namely, the rise and fall of NASDAQ during the *dotcom crash*, as well as the financial crisis of 1998-2012. With regard to power analysis, in general the tests were asymptotically consistent. Since the power increased substantially with T converging to 1. However, as DGP implied an increase in the persistence of regimes, the asymptotic power of the tests worsened slightly, in particular in the $\sup LR(0 \mid 1)$ and in the $WDmax$ when two breaks are present. Regarding further research we believe that inhomogeneity in Markov chains will continue to deserve analytical work as well as close care in accounting for its consequences in practical forecasting exercises.

References

Asadabadi, M. R., 2017. A customer based supplier selection process that combines quality function deployment, the analytic network process and a markov chain. European Journal of Operational Research 263 (3), 1049–1062.

Bai, J., 2000. Vector autoregressive models with structural changes in regression coefficients and in variance-covariance matrices. Annals of Economics and Finance 1 (2), 303–339.

Bai, J., Lumsdaine, R. L., Stock, J. H., 1998. Testing for and dating common breaks in multivariate time series. The Review of Economic Studies 65 (3), 395–432.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica, 47–78.

Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. Journal of applied econometrics 18 (1), 1–22.

Bai, J., Perron, P., 2006. Multiple structural change models: a simulation analysis. Econometric theory and practice: Frontiers of analysis and applied research, 212–237.

Basawa, I. V., 2014. Statistical Inferences for Stochasic Processes: Theory and Methods. Elsevier.

Berchtold, A., 2001. Estimation in the mixture transition distribution model. Journal of Time Series Analysis 22 (4), 379–397.

Billingsley, P., 1961. Statistical methods in markov chains. The Annals of Mathematical Statistics, 12–40.

Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardenes, J., Romance, M., Sendina-Nadal, I., Wang, Z., Zanin, M., 2014. The structure and dynamics of multilayer networks. Physics Reports 544 (1), 1–122.

Bukiet, B., Harold, E. R., Palacios, J. L., 1997. A markov chain approach to baseball. Operations Research 45 (1), 14–23.

Cabello, J. G., 2017. The future of branch cash holdings management is here: New markov chains. European Journal of Operational Research 259 (2), 789–799.

Ching, W.-K., Ng, M. K., 2006. Markov chains. Models, algorithms and applications.

Damásio, B., Louçã, F., Nicolau, J., 2018. The changing economic regimes and expected time to recover of the peripheral countries under the euro: A nonparametric approach. Physica A: Statistical Mechanics and its Applications 507, 524–533.

Damásio, B., Mendonça, S., 2018. Modelling insurgent-incumbent dynamics: Vector autoregressions, multivariate markov chains, and the nature of technological competition. Applied Economics Letters, 1–7.

Damásio, B., Nicolau, J., 2013. Combining a regression model with a multivariate markov chain in a forecasting problem. Statistics & Probability Letters 90, 108–113.

Darkhovsky, B., 2011. Change-point problem for high-order markov chain. Sequential Analysis 30 (1), 41–51.

Davidson, J., 1994. Stochastic limit theory: An introduction for econometricians. OUP Oxford.

Fung, E. S., Siu, T. K., 2012. A flexible markov chain approach for multivariate credit ratings. Computational Economics 39 (2), 135–143.

Gómez, S., Arenas, A., Borge-Holthoefer, J., Meloni, S., Moreno, Y., 2010. Discrete-time markov chain approach to contact-based disease spreading in complex networks. EPL (Europhysics Letters) 89 (3), 38009.

Gottschau, A., 1992. Exchangeability in multivariate markov chain models. Biometrics, 751–763.

Hansen, B. E., 2001. The new econometrics of structural change: dating breaks in us labour productivity. Journal of Economic perspectives 15 (4), 117–128.

Hansen, P. R., 2003. Structural changes in the cointegrated vector autoregressive model. Journal of Econometrics 114 (2), 261–295.

Höhle, M., 2010. Online change-point detection in categorical time series. In: Statistical modelling and regression structures. Springer, pp. 377–397.

Horvath, P. A., Autry, C. W., Wilcox, W. E., 2005. Liquidity implications of reverse logistics for retailers: A markov chain approach. Journal of retailing 81 (3), 191–203.

Karlin, S., Taylor, H. E., 1981. A second course in stochastic processes. Elsevier.

Lai, T. L., 1995. Sequential changepoint detection in quality control and dynamical systems. Journal of the Royal Statistical Society: Series B (Methodological) 57 (4), 613–644.

Lai, T. L., 1998. Information bounds and quick detection of parameter changes in stochastic systems. IEEE Transactions on Information Theory 44 (7), 2917–2929.

Li, Y., Zhu, M., Klein, R., Kong, N., 2014. Using a partially observable markov chain model to assess colonoscopy screening strategies–a cohort study. European Journal of Operational Research 238 (1), 313–326.

Markov, A. A., 1913. Primer statisticheskogo issledovanija nad tekstomevgenija onegina'illjustrirujuschij svjaz'ispytanij v tsep (an example of statistical study on the text ofeugene onegin'illustrating the linking of events to a chain). Izvestija Imp. Akad. nauk.

Maskawa, J.-i., 2003. Multivariate markov chain modeling for stock markets. Physica A: Statistical Mechanics and its Applications 324 (1-2), 317–322.

Mehran, F., 1989. Longitudinal analysis of employment and unemployment based on matched rotation samples. Labour 3 (1), 3–20.

Mei, Y., 2006. Suboptimal properties of page's cusum and shiryayev-roberts procedures in change-point problems with dependent observations. Statistica Sinica 16 (3), 883.

Newey, W. K., McFadden, D., 1994. Large sample estimation and hypothesis testing. Handbook of econometrics 4, 2111–2245.

Nicolau, J., 2014. A new model for multivariate markov chains. Scandinavian Journal of Statistics 41 (4), 1124–1135.

Nicolau, J., 2017. A simple nonparametric method to estimate the expected time to cross a threshold. Statistics & Probability Letters 123, 146–152.

Nicolau, J., Riedlinger, F. I., 2015. Estimation and inference in multivariate markov chains. Statistical Papers 56 (4), 1163–1173.

Perron, P., 1989. The great crash, the oil price shock, and the unit root hypothesis. Econometrica: Journal of the Econometric Society, 1361–1401.

Perron, P., 2006. Dealing with structural breaks. Palgrave handbook of econometrics 1 (2), 278–352.

Polansky, A. M., 2007. Detecting change-points in markov chains. Computational Statistics & Data Analysis 51 (12), 6013–6026.

Qu, Z., Perron, P., 2007. Estimating and testing structural changes in multivariate regressions. Econometrica 75 (2), 459–502.

Raftery, A., Tavaré, S., 1994. Estimation and modelling repeated patterns in high order markov chains with the mixture transition distribution model. Applied Statistics, 179–199.

Rappoport, P., Reichlin, L., 1989. Segmented trends and non-stationary time series. The Economic Journal 99 (395), 168–177.

Sahin, A. D., Sen, Z., 2001. First-order markov chain approach to wind speed modelling. Journal of Wind Engineering and Industrial Aerodynamics 89 (3-4), 263–269.

Sericola, B., 2013. Markov chains: theory and applications. John Wiley & Sons.

Shamshad, A., Bawadi, M., Hussin, W. W., Majid, T., Sanusi, S., 2005. First and second order markov chain models for synthetic generation of wind speed time series. Energy 30 (5), 693–708.

Siu, T.-K., Ching, W.-K., Fung, S. E., Ng, M. K., 2005. On a multivariate markov chain model for credit risk measurement. Quantitative Finance 5 (6), 543–556.

Suhov, Y., Kelbert, M., 2008. Markov chains: A primer in random processes and their applications, cambrigde.

Tan, B., Yılmaz, K., 2002. Markov chain test for time dependence and homogeneity: an analytical and empirical evaluation. European Journal of Operational Research 137 (3), 524–543.

Tauchen, G., 1986. Finite state markov-chain approximations to univariate and vector autoregressions. Economics letters 20 (2), 177–181.

Tsiliyannis, C. A., 2018. Markov chain modeling and forecasting of product returns in remanufacturing based on stock mean-age. European Journal of Operational Research.

Turchin, P. B., 1986. Modelling the effect of host patch size on mexican bean beetle emigration. Ecology 67 (1), 124–132.

Xian, J.-G., Han, D., Yu, J.-Q., 2016. Online change detection of markov chains with unknown post-change transition probabilities. Communications in Statistics-Theory and Methods 45 (3), 597–611.

Yakir, B., 1994. Optimal detection of a change in distribution when the observations form a markov chain with a finite state space. Lecture Notes-Monograph Series, 346–358.

Appendix 1: Mathematical Appendix

Propositon 1. *Under Assumptions A1 and A2, the OLS estimator for (2.9) is, regardless of the sample size, numerically equal to the one obtained through the ML that assumes a multinomial distribution in (2.7)*

Proof. The proof is immediate by the Frisch-Waugh-Lovell theorem, given that the variables of the right-hand side of the equation (5) are, by construction, orthogonal.

In fact, the stacked vector $\hat{\boldsymbol{p}}$ of the estimators $\hat{\boldsymbol{P}}_{\bullet i}$ is given by

$$\begin{aligned}
\hat{\boldsymbol{p}} &= \sum_{t=1}^{T} (\boldsymbol{x}_t \boldsymbol{x}_t')^{-1} \sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{y}_t \\
&= \begin{pmatrix} \sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t' & \cdots & \vdots \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t' \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T} \boldsymbol{z}_t y_{1t} \\ \sum_{t=1}^{T} \boldsymbol{z}_t y_{2t} \\ \vdots \\ \sum_{t=1}^{T} \boldsymbol{z}_t y_{q-1t} \end{pmatrix},
\end{aligned}$$

thus, the i-th estimator is given by

$$\hat{\boldsymbol{P}}_{\bullet i} = \left(\sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t' \right)^{-1} \sum_{t=1}^{T} \boldsymbol{z}_t y_{it}.$$

In view of the fact that

$$\begin{aligned}
\boldsymbol{z}_t\boldsymbol{z}_t' &= \begin{pmatrix} y_{1t-1} \\ \vdots \\ y_{qt-1} \end{pmatrix} \begin{pmatrix} y_{1t-1} & \cdots & y_{qt-1} \end{pmatrix} \\
&= \begin{pmatrix} y_{1t-1}^2 & 0 & \cdots & 0 \\ 0 & y_{2t-1}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & y_{qt-1}^2 \end{pmatrix} = \begin{pmatrix} y_{1t-1} & 0 & \cdots & 0 \\ 0 & y_{2t-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & y_{qt-1} \end{pmatrix}
\end{aligned}$$

then

$$\left(\sum_{t=1}^{T}\boldsymbol{z}_t\boldsymbol{z}_t'\right)^{-1} = \begin{pmatrix} \frac{1}{\sum_{t=1}^{T} y_{1t-1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sum_{t=1}^{T} y_{2t-1}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{1}{\sum_{t=1}^{T} y_{qt-1}} \end{pmatrix}.$$

Therefore, the k-th entry of $\hat{\boldsymbol{P}}_{\bullet i}$ is

$$\begin{aligned}
\hat{P}_{ki} &= \frac{1}{\sum_{t=1}^{T} y_{kt-1}} \sum_{t=1}^{T} z_{kt} y_{it} \\
&= \frac{\sum_{t=1}^{T} y_{kt-1} y_{it}}{\sum_{t=1}^{T} y_{kt-1}} = \\
&= \frac{\sum \mathbb{1}\left(y_{kt-1} = 1, y_{it} = 1\right)}{\sum \mathbb{1}\left(y_{kt-1} = 1\right)} \\
&= \frac{\sum \mathbb{1}\left(S_{t-1} = k, S_t = i\right)}{\sum \mathbb{1}\left(S_{t-1} = k\right)} \\
&= \frac{n_{ik}}{n_i}.
\end{aligned}$$

This expression numerically equals expression (9) as we have orthogonal partitioned regressions. ■

Propositon 2. *Under Assumptions A1 and A2 we have*

$$\ell_j^{-1} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+\ell_j} \boldsymbol{x}_t\boldsymbol{x}_t' \overset{a.s.}{\to} Q_j, \tag{2.17}$$

a non random positive definite matrix, as $\ell_j \to \infty$, for each $j = 1, \cdots, m+1$ and $\ell_j \leq T_j^0 - T_{j-1}^0 + 1$.

Proof. $\{S_t\}$ is a stationary and ergodic Markov chain sequence and $\mathbf{E}\left[|S_t|\right]$ is finite, therefore, using, for example, the pointwise ergodic theorem for stationary sequences (Stout, 1974; White, 2014) we have

$$\begin{aligned}
\frac{1}{\ell_j}\sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+\ell_j} \boldsymbol{z}_t\boldsymbol{z}_t' &= \left(\frac{1}{\ell_j}\sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+\ell_j} \mathbb{1}\left\{S_{t-1}=i\right\}\right) \\
&= \left(\frac{1}{\ell_j}\sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+\ell_j} y_{it-1}\right)_{i=1,\cdots,q} \\
&\overset{as}{\to} E_j\left[y_{it}\right]=\pi_i\left(j\right)
\end{aligned} \tag{5.1}$$

Using the continuous mapping theorem

$$\frac{1}{\ell_j}\sum \boldsymbol{x_t x_t}' \overset{as}{\to} \boldsymbol{\Pi_j},$$

where $\boldsymbol{\Pi_j}$ is the vector of stationary probabilities for the j-th segment. ∎

Propositon 3. *Under Assumptions A1 and A2:*

1. $\{\boldsymbol{x_t\varepsilon_t}, \mathcal{F}_t\}$ *forms a martingale difference sequence;*
2. $\mathbf{E}\left[\boldsymbol{x_t\varepsilon_t}\right]=0$*;*

Proof. The vector $\boldsymbol{\varepsilon}_t$ is, by construction, a martingale difference sequence with respect to $\mathcal{F}_t$, given that $\varepsilon_{it}=y_{it}-\mathbf{E}\left[y_{it}\mid\mathcal{F}_t\right]$. Hence, $\{\boldsymbol{x}_t\boldsymbol{\varepsilon}_t\}$ is also a martingale difference sequence because $\mathbf{E}\left[\boldsymbol{x}_t\boldsymbol{\varepsilon}_t\mid\mathcal{F}_t\right]=\boldsymbol{x}_t\mathbf{E}\left[\boldsymbol{\varepsilon}_t\mid\mathcal{F}_t\right]=0$. Therefore, $\mathbf{E}\left[\boldsymbol{x_t\varepsilon_t}\right]=0$ (by the law of iterated expectations), $\{\boldsymbol{z}_t\boldsymbol{\varepsilon}_t\}$ is an uncorrelated sequence and as a direct consequence $\{\boldsymbol{x_t\varepsilon_t}, \mathcal{F}_t\}$ forms a martingale difference sequence and, thus, a strongly mixing sequence.

∎

Lemma 1. *The covariance matrix (15)* $\boldsymbol{\Sigma}\equiv\mathbf{E}\left[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right]$ *given by*

$$\boldsymbol{\Sigma}=\begin{pmatrix}
\pi_1-\sum_{k=1}^q\pi_kP_{k1}^2 & -\sum_{k=1}^qP_{k1}P_{k2}\pi_k & \cdots & -\sum_{k=1}^qP_{k1}P_{kq}\pi_k \\
-\sum_{k=1}^qP_{k1}P_{k2}\pi_k & \pi_2-\sum_{k=1}^q\pi_kP_{k2}^2 & \cdots & -\sum_{k=1}^qP_{k2}P_{kq}\pi_k \\
\vdots & \vdots & \ddots & \vdots \\
-\sum_{k=1}^qP_{k1}P_{kq}\pi_k & -\sum_{k=1}^qP_{k2}P_{kq}\pi_k & & \pi_q-\sum_{k=1}^q\pi_kP_{kq}^2
\end{pmatrix}.$$

Proof. The covariance writes:

$$\begin{aligned}
\mathbf{E}\left[\varepsilon_{it}\varepsilon_{ij}\right] &= \mathbf{E}\left[\left(y_{it}-x_t'P_{\bullet i}\right)\left(y_{jt}-x_t'P_{\bullet j}\right)\right]= \\
&= \mathbf{E}\left[y_{it}y_{jt}-y_{it}x_t'P_{\bullet j}-x_t'P_{\bullet i}y_{jt}+x_t'P_{\bullet i}x_t'P_{\bullet j}\right] \\
&= \mathbf{E}\left[y_{it}y_{jt}\right]-\mathbf{E}\left[y_{it}x_t'P_{\bullet j}\right]-\mathbf{E}\left[x_t'P_{\bullet i}y_{jt}\right]+\mathbf{E}\left[x_t'P_{\bullet i}x_t'P_{\bullet j}\right] \\
&= 0-\sum_{k=1}^qP_{kj}P_{ki}\pi_k-\sum_{k=1}^qP_{ki}P_{kj}\pi_k+\sum_{k=1}^q\pi_kP_{ki}P_{kj} \\
&= -\sum_{k=1}^qP_{ki}P_{kj}\pi_k.
\end{aligned}$$

Since, by construction $y_{it}y_{jt}=0, i\neq j$, then

$$\begin{aligned}
\mathbf{E}\left[y_{it}y_{j,t-1}\right] &= P\left(S_t=i,S_{t-1}=j\right)=P\left(S_t=i\right|S_{t-1}=j\right)P\left(S_{t-1}=j\right) \\
&= P_{ji}\pi_j \\
\mathbf{E}\left[y_{it}x_t'P_{\bullet j}\right] &= \mathbf{E}\left[y_{it}\left(y_{1,t-1}P_{1j}+...+y_{q,t-1}P_{qj}\right)\right]=\sum_{k=1}^{q}P_{kj}\mathbf{E}\left[y_{it}y_{k,t-1}\right] \\
&= \sum_{k=1}^{q}P_{kj}P_{ki}\pi_k \\
\mathbf{E}\left[x_t'P_{\bullet i}y_{jt}\right] &= \mathbf{E}\left[y_{jt}x_t'P_{\bullet i}\right]=\sum_{k=1}^{q}P_{ki}P_{kj}\pi_k \\
\mathbf{E}\left[x_t'P_{\bullet i}x_t'P_{\bullet j}\right] &= \mathbf{E}\left[P_{\bullet i}'x_tx_t'P_{\bullet j}\right]=P_{\bullet i}'\mathbf{E}\left[x_tx_t'\right]P_{\bullet j}= \\
&= P_{\bullet i}'\begin{bmatrix}\pi_1 & 0 & \cdots & 0\\ 0 & \pi_2 & \cdots & 0\\ \vdots & & \ddots & \vdots\\ 0 & 0 & \cdots & \pi_q\end{bmatrix}P_{\bullet j}=\sum_{k=1}^{q}\pi_kP_{ki}P_{kj}.
\end{aligned}$$

On the other hand, $\mathbf{E}\left[\varepsilon_{it}^2\right]$ may be written as

$$\begin{aligned}
\mathbf{E}\left[\varepsilon_{it}^2\right] &= \mathbf{E}\left[\left(y_{it}-x_t'P_{\bullet i}\right)^2\right]= \\
&= \mathbf{E}\left[y_{it}^2-2y_{it}x_t'P_{\bullet i}+\left(x_t'P_{\bullet i}\right)^2\right] \\
&= \mathbf{E}\left[y_{it}^2-2y_{it}x_t'P_{\bullet i}+P_{\bullet i}'x_tx_t'P_{\bullet i}\right]
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{E}\left[y_{it}y_{it}\right] &= \mathbf{E}\left[y_{it}^2\right]=\mathbf{E}\left[y_{it}\right]=\pi_i \\
\mathbf{E}\left[y_{it}x_t'P_{\bullet i}\right] &= \sum_{k=1}^{q}P_{kj}^2\pi_k \\
\mathbf{E}\left[P_{\bullet i}'x_tx_t'P_{\bullet i}\right] &= \sum_{k=1}^{q}\pi_kP_{ki}^2,
\end{aligned}$$

therefore

$$\begin{aligned}
\mathbf{E}\left[\varepsilon_{it}^2\right] &= \pi_i-2\sum_{k=1}^{q}P_{kj}^2\pi_k+\sum_{k=1}^{q}\pi_kP_{ki}^2 \\
&= \pi_i-\sum_{k=1}^{q}\pi_kP_{ki}^2.
\end{aligned}$$

■