CONTRIBUTI DI RICERCA CRENOS

ON THE RELATIONSHIP BETWEEN MARKOV SWITCHING INFERENCE AND FUZZY CLUSTERING: A MONTE CARLO EVIDENCE

Edoardo Otranto
Luca Scaffidi Domianello

WORKING PAPERS

2023/04

arkadia

CENTRO RICERCHE ECONOMICHE NORD SUD
(CRENOS)
UNIVERSITÀ DI CAGLIARI
UNIVERSITÀ DI SASSARI

CRENOS was set up in 1993 with the purpose of organising the joint research effort of economists from the two Sardinian universities (Cagliari and Sassari) investigating dualism at the international and regional level. CRENoS' primary aim is to improve knowledge on the economic gap between areas and to provide useful information for policy intervention. Particular attention is paid to the role of institutions, technological progress and diffusion of innovation in the process of convergence or divergence between economic areas. To carry out its research, CRENoS collaborates with research centres and universities at both national and international level. The centre is also active in the field of scientific dissemination, organizing conferences and workshops along with other activities such as seminars and summer schools.
CRENoS creates and manages several databases of various socio-economic variables on Italy and Sardinia. At the local level, CRENoS promotes and participates to projects impacting on the most relevant issues in the Sardinian economy, such as tourism, environment, transports and macroeconomic forecasts.

www.crenos.unica.it
crenos@unica.it

CRENOS – CAGLIARI
VIA SAN GIORGIO 12, I-09124 CAGLIARI, ITALIA
TEL. +39-070-6756397; FAX +39-070- 6756402

CRENOS - SASSARI
VIA MURONI 25, I-07100 SASSARI, ITALIA
TEL. +39-079-213511

Title: ON THE RELATIONSHIP BETWEEN MARKOV SWITCHING INFERENCE AND FUZZY CLUSTERING: A MONTE CARLO EVIDENCE

Prima Edizione: Aprile 2023

ISBN: 978 88 68514 655

Viale Bonaria 98 - 09125 Cagliari
Tel. 070/6848663 - info@arkadiaeditore.it
www.arkadiaeditore.it

On the relationship between Markov Switching inference and Fuzzy Clustering: A Monte Carlo evidence

Edoardo Otranto[∝]
University of Messina & CRENoS
Luca Scaffidi Domianello[*]
University of Florence

Abstract

Markov Switching models have had increasing success in time series analysis due to their ability to capture the existence of unobserved discrete states in the dynamics of the variables under study. This result is generally obtained thanks to the inference on states derived from the so–called Hamilton filter. One of the open problems in this framework is the identification of the number of states, generally fixed a priori; it is in fact impossible to apply classical tests due to the problem of the nuisance parameters present only under the alternative hypothesis. In this work we show, by Monte Carlo simulations, that fuzzy clustering is able to reproduce the parametric state inference derived from the Hamilton filter and that the typical indices used in clustering to determine the number of groups can be used to determine the number of states in this framework. The procedure is very simple to apply, considering that it is performed (in a nonparametric way) independently of the data generation process and that the indicators we use are present in most statistical packages.

Keywords: Nuisance parameters, Groups identification, Number of states, Simulations, Markov chains
Jel Classification: C15, C22, C24, C38.

[∝] Corresponding author - Department of Economics; E-mail: eotranto@unime.it
[*] Department of Statistics, Computer Science, Applications; E-mail: luca.scaffididomianello@unifi.it

1 Introduction

Markov Switching (MS) models (Hamilton, 1990) have received increasing attention in time series analysis with several applications in economics (see, for example Hamilton, 2016), finance (Gallo and Otranto, 2015), neuroscience (Degras, Ting, and Ombao, 2022), just to name a few fields. Its main advantage consists in the possibility to consider the existence of several states, interpreted as particular regimes (for example, growth and recession in business cycle, quiet and turmoil periods in the financial markets), which are not observed, but whose dynamics can be represented by an ergodic Markov chain. Thanks to the properties of ergodic Markov chains, it is possible to make inference on the unobserved state, assigning to each state at each time a filtered or smoothed probability obtained from the so–called Hamilton filter (Hamilton, 1990). An open problem is the identification of the number of states, which is not feasible with classical tests for the problem of nuisance parameters present only under the alternative hypothesis. In practice, when the model under the null hypothesis is a MS model with k states and the model under the alternative is a model with $k+1$ states, the two models are not nested due to the presence of transition probabilities referring to the state $k+1$, not identified under the null hypothesis (see, for example, Hansen, 1992). As known, in this framework the classical tests, such as Likelihood Ratio (LR), do not follow the standard distributions and the true distribution is unknown. There are several proposals to bypass this problem; the first approaches were based on the supremum of a LR test (Davies, 1977), trying to derive its asymptotic distribution over a range of nuisance parameters via simulations (Hansen, 1992), but with a high computational cost. Alternatively, Garcia (1998) proposes a similar approach by reducing the range of nuisance parameters, with the exclusion of some important particular cases. Most recent approaches avoid the development of tests, trying to directly identify the number of states by penalized likelihood criteria (Psadarakis and Spagnolo, 2003, 2006) or Kullback–Leibler divergence (Smith, Naikb, and Tsai, 2006).

A very practical idea is to identify the number of states before the estimation step, emphasizing the fact that in an MS models each observation is generated by a mixture density, with a number of components equal to the number of states of the Markov chain. This approach was developed in a nonparametric Bayesian framework by Otranto and Gallo (2002), where the posterior distribution of the number of states is derived by means of a Gibbs sampler. This intuition is the basis of our proposal.

First, we note a similarity between the inference on the regime in MS models and the grouping derived from a fuzzy clustering (D'Urso, 2015). This last approach is able to provide a clustering of statistical units in k groups, with a probability of belonging to each of the k clusters. Our first analysis consists in verifying whether the fuzzy approach provides a grouping similar to the one derived from the inference on regimes of the MS models, using the same assignment criterion (the unit is assigned to the group corresponding to the mode). This exercise is performed by means of Monte Carlo experiments. After verifying the similarity of the results in terms of state inference, we verify whether the typical indices used to select the number of clusters are able to identify the true number of states. The resulting approach is very simple, it is applied in the identification phase, before estimating the model, as in Otranto and Gallo (2002), but unlike these, it is very fast and uses tools implemented in the main statistical routines.

The structure of the paper is as follows: in the next section the main features of the MS model and of fuzzy clustering are briefly recalled; in Section 3 we present a large set of Monte Carlo experiments to verify the ability of fuzzy clustering to reproduce the inference on the states of MS models (subsection 3.1) and the performance of several indices in detecting the number of states (subsection 3.2). Some final remarks will conclude the paper.

2 MS Models and Fuzzy Clustering

Let us consider a time series y_t with $t = 1, \dots, T$. We say that its Data Generating Process (DGP) follows an MS process with k states if:

$$y_t = f(\boldsymbol{x}; \boldsymbol{\theta}_{s_t}) + \varepsilon_t \tag{1}$$

where $\boldsymbol{x}$ is a vector of exogenous variables, possibly including lagged values of y_t, ε_t ($t = 1, \dots, T$) are independent disturbances with zero mean and possibly changing variance (this will be clarified shortly). The vector of unknown parameters $\boldsymbol{\theta}_{s_t}$ depends on an unobservable discrete random variable s_t, which can assume values $1, 2, \dots k$, and whose dynamics is driven by an ergodic Markov chain, with elements of the transition probability matrix:

$$p_{ij} = Pr(s_t = j | s_{t-1} = i), \qquad i, j = 1, \dots, k \tag{2}$$

with $\sum_{j=1}^{k} p_{ij} = 1$ for each $i = 1, \dots, k$. Also the variance of ε_t could depend on the state variable s_t, so they could be not identically distributed. A typical model used in an MS framework is the $MS - AR(p)$ process:

$$y_t = \mu_{s_t} + \sum_{i=1}^{p} \phi_i \left(y_{t-i} - \mu_{s_{t-i}} \right) + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2) \tag{3}$$

Under stationarity constraints, specification (3) involves a time–varying unconditional mean of the process, μ_{s_t}, depending on the state at time t. The transition probabilities p_{ij} are estimated with other parameters $\boldsymbol{\theta}_{s_t} = (\mu_1, \dots, \mu_k)'$, $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)'$ and σ^2 by Maximum Likelihood Estimator (MLE), deriving the likelihood function by means of the so–called Hamilton filter(Hamilton, 1990). The same Hamilton filter provides the possibility to specify, at each time t, the conditional probability to fall in a certain state j; given the information set $I_\tau = (y_\tau, y_{\tau-1}, \dots)$, it is possible to obtain the so–called filtered probabilities $Pr(s_t = j | I_t)$, the predicted probabilities $Pr(s_t = j | I_{t-1})$ and the smoothed probabilities $Pr(s_t = j | I_T)$. The predicted probabilities enter the likelihood function; in fact the conditional density of each y_t is expressed as:

$$f(y_t | I_{t-1}; \boldsymbol{\theta}_{s_t}, \boldsymbol{\phi}, \sigma^2) = \sum_{j=1}^{k} f(y_t | s_t = j, I_{t-1}; \boldsymbol{\theta}_{s_t}, \boldsymbol{\phi}, \sigma^2) Pr(s_t = j | I_{t-1}) \tag{4}$$

which is a mixture of distributions with weights represented by the predicted probabilities.

Moreover, the Hamilton filter provides also an intuitive inference on the regime, assigning to state j the observation at time t with mode in j, using as mass distribution the filtered or, more frequently, the smoothed probabilities.

As said in Section 1, the identification of the number of states k is not achievable through classical statistical tests for the problem of nuisance parameters present only under the alternative hypothesis. As an example, let us suppose to verify the null hypothesis of a linear model (no states or $k = 1$):

$$y_t = \mu + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2) \tag{5}$$

against the alternative of a MS model with $k = 2$ states:

$$\begin{aligned} &y_t = \mu_{s_t} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2), \qquad s_t = 1, 2 \\ &\boldsymbol{P} = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix} \end{aligned} \tag{6}$$

The linear model can not be obtained from the MS model simply imposing that the switching parameters are equal in the two states ($\mu_1 = \mu_2$), because the transition probabilities p_{11} and p_{22} are not identified under the null hypothesis and the two models are not nested. Similarly if the 2–state model (6) is assumed under the null hypothesis, whereas the alternative refers to a MS model with $k = 3$:

$$\begin{aligned} &y_t = \mu_{s_t} + \varepsilon_t, \qquad \varepsilon \sim N(0, \sigma^2), \qquad s_t = 1, 2, 3 \\ &\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}; \qquad \sum_{j=1}^{3} p_{ij} = 1, \quad i = 1, 2, 3 \end{aligned} \tag{7}$$

In this case 2 of the 3 probabilities in each row of $\boldsymbol{P}$ are not identified under the null hypothesis.

There is some similarity between the MS models and the fuzzy clustering methods. The latter detect the belonging of each statistical unit to a cluster with a certain probability (the *membership grade*), unlike the classical hard clustering, where each unit can belong to exactly one cluster only. Clustering based on the degree of membership in each group can be seen as similar to state inference done using the smoothed probabilities in an MS framework. For example, considering the most popular fuzzy clustering algorithm, the *fuzzy k–means* (Bezdek, 1981), the fuzzy partition in k groups of the observed y_t's ($t = 1, \ldots, T$) is obtained by minimizing:

$$\min_{\boldsymbol{U}, \boldsymbol{c}} \sum_{t=1}^{T} \sum_{j=1}^{k} u_{tj}^m d^2 \left(y_t, c_j \right) \tag{8}$$

where $\boldsymbol{U} = \{u_{t,j}\}$ ($t = 1, \ldots, T$; $j = 1, \ldots, k$) is the membership grade matrix, $\boldsymbol{c} = (c_1, \ldots, c_k)'$ is the vector of centroids, m is the fuzziness parameter, which tunes the degree of fuzziness, $d(\cdot, \cdot)$ is a distance measure.

Each column of $\boldsymbol{U}$ can be interpreted similarly to the smoothed probabilities of an MS model. Based on this insight, we ask whether nonparametric methods used to detect the number of groups in clustering can be used as a method to identify the number of states in an MS model. The detection of the number of clusters is generally carried out by

means of indicators implemented in the main statistical packages, making this approach very simple and easily usable even by non–experts.

The verification of the previous idea can be performed in two steps through Monte Carlo experiments where data are generated from several MS processes: first, fixing the *true* k, we compare the smoothed probabilities derived from the estimated MS with the corresponding grade of membership matrix derived from (8);[1] then we use the main indices to detect the number of clusters verifying if they are able to identify the correct number of states of the MS DGP. More specifically, we rely on the following cluster validation indices: Partition Coefficient (PC), Partition Entropy (PE), Modified Partition Coefficient, (MPC), Average Silhouette Width (ASW), Average Silhouette Width Fuzzy (ASWF), and Xie-Beni (XB).

The Partition Coefficient (Bezdek, 1981) is given by:

$$PC = \sum_{t=1}^{T} \sum_{j=1}^{k} u_{tj}^2 / T.$$

It can assume values between $1/k$ and 1, with its maximum value, as function of k, giving us the optimal number of clusters.

The Partition Entropy (Bezdek, 1981):

$$PE = -\sum_{j=1}^{k} u_{tj} \ln(u_{tj}) / T,$$

ranges in $[0, \ln k]$ and its lowest value provides the best number of clusters.

The Modified Partition Coefficient (Dave, 1996)

$$MPC = 1 - \frac{k}{k-1}(1 - PC),$$

normalizes the PC index, so that it ranges in $[0, 1]$, thus eliminating dependency on k.

The Average Silhouette Width (Rousseeuw, 1987) is calculated as follows:

$$ASW = (1/T) \sum_{t=1}^{T} \frac{b_t - a_t}{max\{a_t, b_t\}},$$

with a_t the average distance among y_t (belonging to the cluster τ) and the other observations belonging to the same group τ, while b_t is the minimum average distance among y_t and the observations belonging to another cluster $j \neq \tau$. It can assume values between -1 and 1, with the maximum value providing us with the best number of clusters.

The Average Silhouette Width Fuzzy (Campello and Hruschka, 2006),

$$ASWF = \frac{\sum_{t=1}^{T} (u_{tj_1} - u_{tj_2})^{\lambda} ASW_t}{\sum_{t=1}^{T} (u_{tj_1} - u_{tj_2})^{\lambda}},$$

[1] Classifying the European Central Bank announcements, Gallo, Lacava, and Otranto (2021) find very similar results between the classification derived from smoothed probabilities of their MS model and the *k–means* clustering procedure.

is a weighted average of the Silhouette, where u_{tj_1} and u_{tj_2} are the elements of the t–th row of $\mathbf{U}$ with first and second largest values, respectively, ASW_t the Silhouette of the $t-th$ observation, and $\lambda \geq 0$. Notice that, as opposed to ASW, it takes into account the membership grade matrix.

Finally, the Xie-Beni index (Xie and Beni, 1991),

$$XB = \frac{\sum_{t=1}^{T}\sum_{j=1}^{k} u_{tj}^2 d^2(y_t, c_j)}{T \min_{i,j} d^2(c_i, c_j)},$$

is minimized to obtain the best partition. Notice that the XB index, takes into account both membership grade matrix and the observations.

3 Monte Carlo Evidence

The DGPs used for the Monte Carlo experiments are the models in equation (6) (call it MS(2)), equation (7) (MS(3)) and MS–AR(1) models like (3), with $p = 1$, and with 2 (MS(2)-AR(1)) or 3 (MS(3)-AR(1)) states, adopting the Normal distribution for ε_t. We cover several scenarios, combining a set of parameters that yields the 32 models (8 of each model type) shown in Table 1 (labeled to identify them).

As the distance between the μ_i coefficients increases, the existence of states becomes clearer. This can be better appreciated by looking at Figure 1, where we show the mixture of Normal distributions obtained using, as mixture weights, the ergodic probability of each state.[2] The parameters of the Normal distributions are given by the unconditional means of y_t, expressed by μ_{s_t} in the MS(2) and MS(3) DGPs, and $(\mu_{s_t} - \phi\mu_{s_{t-1}})/(1-\phi)$ in the case of MS(2)–AR(1) and MS(3)–AR(1).[3] The number of mixtures is indistinguishable when the μ_i coefficients are close and the variance is larger (as in MS2–1, MS2AR–1, MS3–1, MS3AR–1); the presence of states is more evident by increasing the distance between μ_i parameters and decreasing the variance. The presence of the AR parameters increases the number of components of the mixture, but the number of highest peaks are equal to the number of states.

For each DGP we generate 1000 time series of length $T = 100$. For each series we perform two types of analyses: first, considering the number of states known, we verify whether fuzzy clustering provides a similar inference on the states obtained from the estimated MS model; then we verify through the validation indices listed at the end of Section 2 if the clustering algorithm is able to detect the right number of states.

[2] As shown in equation (4), each density has different mixture weights, given by the predicted probabilities $Pr(s_t|I_{t-1})$; the vector of ergodic probabilities is the (normalized) eigenvector associated to the unit eigenvalue of $\boldsymbol{P}'$, which can be interpreted as the vector of unconditional probabilities of the state s_t (for details, see Hamilton, 1994).

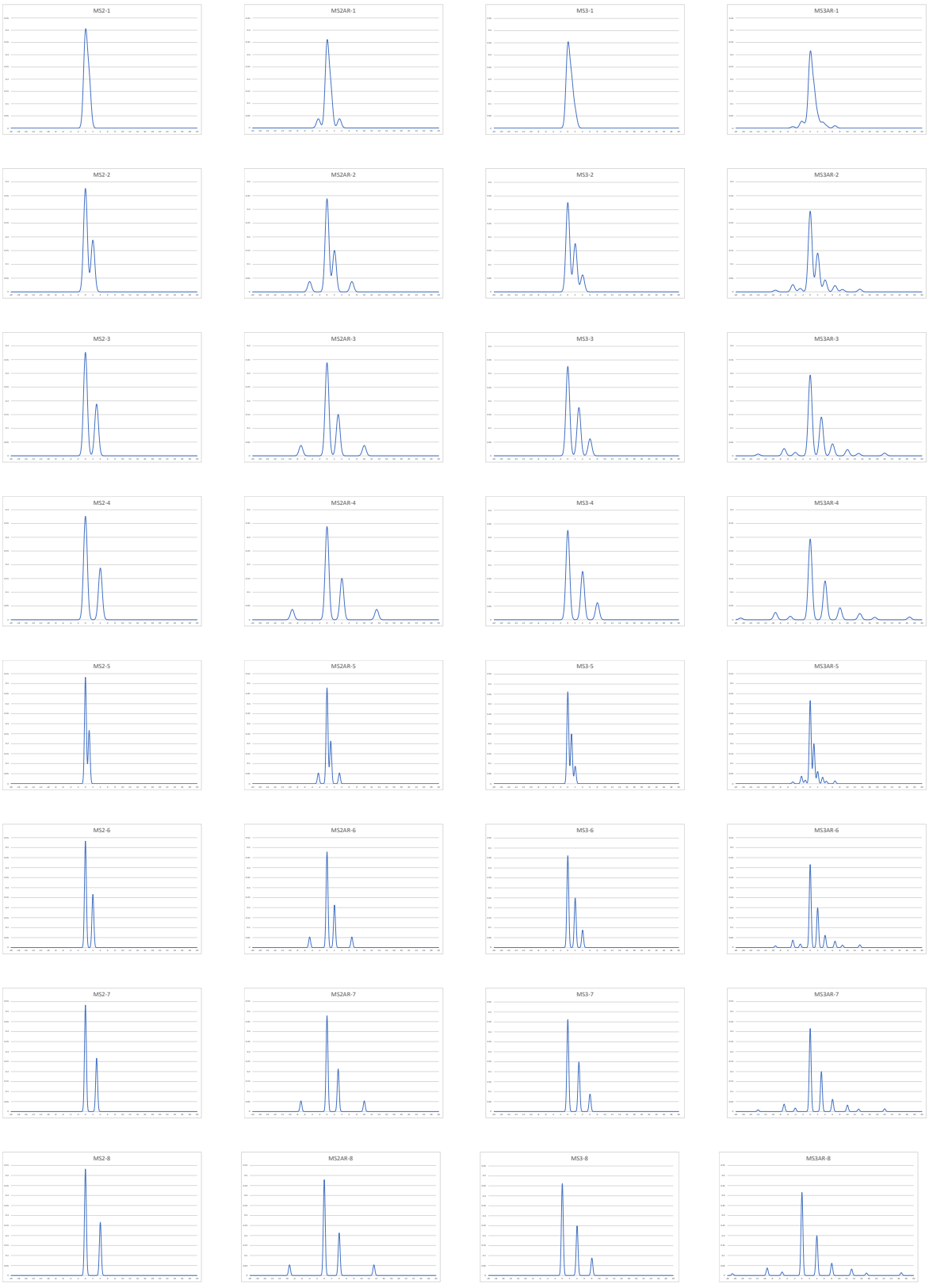
[3] This implies that, in the case of AR(1) model, the number of states can be seen as 2^k (for details, see Hamilton, 1994).

Table 1: Data Generating Processes (DGPs) used in Monte Carlo experiments.

Label	Parameters	Label	Parameters
	DGP: MS(2)		DGP: MS(2)–AR(1)
MS2–1	$\mu_1 = 0$; $\mu_2 = 1$; $\sigma = 0.5$	MS2AR–1	$\mu_1 = 0$; $\mu_2 = 1$; $\sigma = 0.5$
MS2–2	$\mu_1 = 0$; $\mu_2 = 2$; $\sigma = 0.5$	MS2AR–2	$\mu_1 = 0$; $\mu_2 = 2$; $\sigma = 0.5$
MS2–3	$\mu_1 = 0$; $\mu_2 = 3$; $\sigma = 0.5$	MS2AR–3	$\mu_1 = 0$; $\mu_2 = 3$; $\sigma = 0.5$
MS2–4	$\mu_1 = 0$; $\mu_2 = 4$; $\sigma = 0.5$	MS2AR–4	$\mu_1 = 0$; $\mu_2 = 4$; $\sigma = 0.5$
MS2–5	$\mu_1 = 0$; $\mu_2 = 1$; $\sigma = 0.25$	MS2AR–5	$\mu_1 = 0$; $\mu_2 = 1$; $\sigma = 0.25$
MS2–6	$\mu_1 = 0$; $\mu_2 = 2$; $\sigma = 0.25$	MS2AR–6	$\mu_1 = 0$; $\mu_2 = 2$; $\sigma = 0.25$
MS2–7	$\mu_1 = 0$; $\mu_2 = 3$; $\sigma = 0.25$	MS2AR–7	$\mu_1 = 0$; $\mu_2 = 3$; $\sigma = 0.25$
MS2–8	$\mu_1 = 0$; $\mu_2 = 4$; $\sigma = 0.25$	MS2AR–8	$\mu_1 = 0$; $\mu_2 = 4$; $\sigma = 0.25$
	DGP: MS(3)		DGP: MS(3)–AR(1)
MS3–1	$\mu_1 = 0$; $\mu_2 = 1$; $\mu_3 = 2$; $\sigma = 0.5$	MS3AR–1	$\mu_1 = 0$; $\mu_2 = 1$; $\mu_3 = 2$; $\sigma = 0.5$
MS3–2	$\mu_1 = 0$; $\mu_2 = 2$; $\mu_3 = 4$; $\sigma = 0.5$	MS3AR–2	$\mu_1 = 0$; $\mu_2 = 2$; $\mu_3 = 4$; $\sigma = 0.5$
MS3–3	$\mu_1 = 0$; $\mu_2 = 3$; $\mu_3 = 6$; $\sigma = 0.5$	MS3AR–3	$\mu_1 = 0$; $\mu_2 = 3$; $\mu_3 = 6$; $\sigma = 0.5$
MS3–4	$\mu_1 = 0$; $\mu_2 = 4$; $\mu_3 = 8$; $\sigma = 0.5$	MS3AR–4	$\mu_1 = 0$; $\mu_2 = 4$; $\mu_3 = 8$; $\sigma = 0.5$
MS3–5	$\mu_1 = 0$; $\mu_2 = 1$; $\mu_3 = 2$; $\sigma = 0.25$	MS3AR–5	$\mu_1 = 0$; $\mu_2 = 1$; $\mu_3 = 2$; $\sigma = 0.25$
MS3–6	$\mu_1 = 0$; $\mu_2 = 1$; $\mu_3 = 2$; $\sigma = 0.25$	MS3AR–6	$\mu_1 = 0$; $\mu_2 = 2$; $\mu_3 = 4$; $\sigma = 0.25$
MS3–7	$\mu_1 = 0$; $\mu_2 = 1$; $\mu_3 = 2$; $\sigma = 0.25$	MS3AR–7	$\mu_1 = 0$; $\mu_2 = 3$; $\mu_3 = 6$; $\sigma = 0.25$
MS3–8	$\mu_1 = 0$; $\mu_2 = 1$; $\mu_3 = 2$; $\sigma = 0.25$	MS3AR–8	$\mu_1 = 0$; $\mu_2 = 4$; $\mu_3 = 8$; $\sigma = 0.25$

The AR(1) coefficient, when present, is equal to 0.7 in all DGPs. The transition probabilities in the MS(2) and MS(2)–AR(1) DGPs are $p_{11} = 0.9$ and $p_{22} = 0.8$; in the MS(3) and MS(3)–AR(1) DGPs are $p_{11} = 0.9$, $p_{12} = 0.07$, $p_{22} = 0.8$, $p_{21} = 0.15$, $p_{33} = 0.7$, $p_{32} = 0.2$.

Figure 1: Mixture density functions relative to 32 DGP described in Table 1, with weights equal to the corresponding ergodic probabilities.



3.1 Inference on the state

As said, the inference on the state of MS models is performed by assigning the observation at time t to the state with the highest smoothed probability. In our first experiment we generate data from the 32 DGPs shown in Table 1, estimate the corresponding MS model, derive the smoothed probabilities, assign each observation to a state, and finally compare the clustering to the true one using the Adjusted Rand index (Rand, 1971; Hubert and Arabie, 1985):

$$R_a = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}\binom{T_{ij}}{2} - [\sum_{i=1}^{k}\binom{T_i}{2}][\sum_{j=1}^{k}\binom{T_j}{2}]/\binom{T}{2}}{[\sum_{i=1}^{k}\binom{T_i}{2} + \sum_{j=1}^{k}\binom{T_j}{2}]/2 - [\sum_{i=1}^{k}\binom{T_i}{2}][\sum_{j=1}^{k}\binom{T_j}{2}]/\binom{T}{2}} \tag{9}$$

where T_i and T_j represent the number of series belonging to the true group i and group j derived from the MS inference respectively; T_{ij} is the number of series belonging to the group i in the true clustering and assigned to the group j in the MS clustering. The adjusted Rand index R_a ranges in $[0, 1]$, with maximum value in the case of perfect match between true and MS clustering, while it is 0 when the differences between them are at their maximum. We calculated R_a for each simulated time series of each DGP; in the first columns of Table 2 we show the five–number summary (minimum, first quartile, median, third quartile, maximum)[4] of the empirical distribution of R_a for each DGP.

It is interesting to note that, despite the fact that the estimated model is correctly specified, when the means are close to each other–see also Figure 1– it is not trivial to obtain the correct classification of the observations (models MS2–1 , MS3–1), and the problem is more evident when there is an autoregressive term (MS2AR–1, MS2AR–2, MS3AR–1, MS3AR–2). The Rand index increases when the variance of each state is smaller (models labeled with numbers from 4 to 8). However, excluding the MS3AR–1 and the MS2AR–1 cases, in each DGP more than 50% of the simulations provide a Rand index greater than 0.8. The 100% correct classifications are achieved when the distance among the means of each state is larger and the variance is smaller (models MS3–7, MS3–8, MS2AR–7, MS2AR–8, MS3AR–7, MS3AR–8).

Considering the fuzzy classification, obtained using the squared Euclidean distance in (8), and comparing this classification with the true one (second block of columns in Table 2), we note a certain difference with respect to the previous comparison only in the minimum value of the Rand index and a better behavior for MS2–7 and MS2–8. In practice, the classification obtained in a nonparametric way and not considering the model that generates the data, shows very similar results to the case using the right model (not known in practical cases), which requires the estimation step. This is confirmed by observing the third block of columns of Table 2, where the MS and the fuzzy approach are compared (in equation (9) the index i now refers to the MS classification and j to the fuzzy classification); classification differences are reduced when the distance between means is larger and the variability is smaller.

[4]We prefer to show these values in a table and not in box–plots because the latter is difficult to be visualize when the variability of results is close to zero.

Table 2: Summary Statistics of Rand Index for MS Inference on the state and Fuzzy k-means clustering in 1000 Monte Carlo experiments.

DGP	Rand Index														
	MS-True State					Fuzzy-True State					Fuzzy-MS				
	Min.	Q1	Q2	Q3	Max.	Min.	Q1	Q2	Q3	Max.	Min.	Q1	Q2	Q3	Max.
MS2–1	0.50	0.76	0.82	0.87	1.00	0.50	0.64	0.69	0.74	0.89	0.50	0.68	0.74	0.80	1.00
MS2–2	0.79	0.94	0.96	0.98	1.00	0.52	0.92	0.96	0.98	1.00	0.54	0.92	0.94	0.96	1.00
MS2–3	0.94	0.98	0.98	0.98	0.98	0.94	1.00	1.00	1.00	1.00	0.92	0.98	0.98	0.98	0.98
MS2–4	0.92	0.98	0.98	0.98	0.98	0.74	1.00	1.00	1.00	1.00	0.80	0.98	0.98	0.98	0.98
MS2–5	0.82	0.97	0.98	1.00	1.00	0.52	0.90	0.94	0.97	1.00	0.50	0.90	0.95	0.98	1.00
MS2–6	0.98	0.98	0.98	0.98	0.98	0.55	1.00	1.00	1.00	1.00	0.54	0.98	0.98	0.98	0.98
MS2–7	0.98	0.98	0.98	0.98	0.98	1.00	1.00	1.00	1.00	1.00	0.98	0.98	0.98	0.98	0.98
MS2–8	0.98	0.98	0.98	0.98	0.98	1.00	1.00	1.00	1.00	1.00	0.98	0.98	0.98	0.98	0.98
MS2AR–1	0.49	0.56	0.64	0.73	0.98	0.49	0.56	0.60	0.66	0.90	0.49	0.54	0.66	0.79	0.96
MS2AR–2	0.52	0.83	0.90	0.96	1.00	0.50	0.79	0.85	0.90	1.00	0.49	0.76	0.85	0.92	1.00
MS2AR–3	0.63	0.98	1.00	1.00	1.00	0.57	0.94	0.98	1.00	1.00	0.58	0.94	0.98	1.00	1.00
MS2AR–4	0.82	1.00	1.00	1.00	1.00	0.51	1.00	1.00	1.00	1.00	0.51	1.00	1.00	1.00	1.00
MS2AR–5	0.51	0.82	0.90	0.96	1.00	0.50	0.78	0.85	0.90	1.00	0.50	0.76	0.85	0.90	1.00
MS2AR–6	0.89	1.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00	1.00
MS2AR–7	1.00	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00
MS2AR–8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MS3–1	0.53	0.79	0.84	0.89	0.98	0.47	0.63	0.67	0.71	0.88	0.48	0.65	0.69	0.76	0.98
MS3–2	0.85	0.97	0.98	1.00	1.00	0.58	0.90	0.94	0.96	1.00	0.56	0.91	0.95	0.97	1.00
MS3–3	0.96	1.00	1.00	1.00	1.00	0.56	0.99	1.00	1.00	1.00	0.56	1.00	1.00	1.00	1.00
MS3–4	0.98	1.00	1.00	1.00	1.00	0.57	1.00	1.00	1.00	1.00	0.57	1.00	1.00	1.00	1.00
MS3–5	0.82	0.97	0.98	1.00	1.00	0.52	0.90	0.94	0.97	1.00	0.50	0.90	0.95	0.98	1.00
MS3–6	0.98	1.00	1.00	1.00	1.00	0.56	1.00	1.00	1.00	1.00	0.56	1.00	1.00	1.00	1.00
MS3–7	1.00	1.00	1.00	1.00	1.00	0.57	1.00	1.00	1.00	1.00	0.57	1.00	1.00	1.00	1.00
MS3–8	1.00	1.00	1.00	1.00	1.00	0.62	1.00	1.00	1.00	1.00	0.62	1.00	1.00	1.00	1.00
MS3AR–1	0.38	0.58	0.64	0.71	0.96	0.41	0.58	0.62	0.66	0.88	0.37	0.61	0.67	0.74	0.94
MS3AR–2	0.50	0.81	0.89	0.95	1.00	0.44	0.70	0.78	0.86	1.00	0.39	0.69	0.77	0.87	0.99
MS3AR–3	0.62	0.97	1.00	1.00	1.00	0.52	0.92	0.97	0.99	1.00	0.52	0.91	0.96	0.98	1.00
MS3AR–4	0.68	1.00	1.00	1.00	1.00	0.55	0.99	1.00	1.00	1.00	0.55	0.99	1.00	1.00	1.00
MS3AR–5	0.52	0.81	0.89	0.95	1.00	0.50	0.71	0.80	0.87	1.00	0.43	0.70	0.78	0.88	1.00
MS3AR–6	0.70	1.00	1.00	1.00	1.00	0.53	0.99	1.00	1.00	1.00	0.53	0.99	1.00	1.00	1.00
MS3AR–7	1.00	1.00	1.00	1.00	1.00	0.55	1.00	1.00	1.00	1.00	0.55	1.00	1.00	1.00	1.00
MS3AR–8	1.00	1.00	1.00	1.00	1.00	0.63	1.00	1.00	1.00	1.00	0.63	1.00	1.00	1.00	1.00

The summary statistics are the minimum (Min), the first quartile (Q1), the median (Q2), the third quartile (Q3), the maximum (Max) of the Rand index for each set of Monte Carlo experiments. MS refers to the classification obtained by the MS inference on the state, Fuzzy to the classification obtained by the Fuzzy k–means procedure of clustering, True State is the correct classification of data.

3.2 Identification of the number of states

Having empirical evidence of consistent classifications between the typical parametric MS procedure and the fuzzy nonparametric clustering, the next step is to verify if it is possible to use the usual criteria for detecting the number of clusters to identify the number of states in MS models. We use the same datasets generated for the experiment illustrated in subsection 3.1 to apply the criteria for the detection of the number of clusters based on the indices PC, PE, MPC, ASW, ASWF, XB. In Table 3 we show the percentage of correct identification of the number of states for each index.

Considering DGPs with 2 states, in general all indices are able to detect the correct number of states when the differences between the two means is greater than 1 whatever the variance, that is excluding the cases MS2–1, MS2–5, MS2AR–1, MS2AR–5. PE shows a good performance also in the last four cases, with a success rate between 60.1% (MS2–5) and 90.7% (MS2AR-5); second best seems to be the PC index. The indices are often able to detect the correct number of states when the differences between means are larger (DGPs with suffix 3, 4, 7 and 8), equal or very close to 100%.

The cases with 3 states have more uncertainty in detecting the correct number of groups for small differences in means (MS3–1, MS3–2, MS3AR–1, MS3AR–2), but in this case the lower variance seems to promote better performance. The presence of the autoregressive term in DGP seems to worsen the success rate.

In general, the PE index shows the best performance for the 2–state case, presenting the highest percentage of correct detection in the 16 corresponding DGPs (in 6 cases 100%). In the 3–state case there is no a clear preferred index: XB and MPC seem better in half of cases, but the differences in success rate are very small. By summing the success rate in each column of Table 3, the PE index shows the highest score with ASW being second best.

Table 3: Identification of the number of states in 1000 Monte Carlo experiments and 32 Data Generating Processes (DGPs) with several indices

	Indices					
DGP	PC	PE	MPC	ASW	ASWF	XB
MS2–1	64.3	80.2	2.1	47.0	42.9	18.2
MS2–2	99.2	99.6	86.1	99.2	99.3	98.9
MS2–3	100.0	100.0	100.0	100.0	100.0	100.0
MS2–4	100.0	100.0	99.9	99.9	99.9	99.9
MS2–5	54.7	60.1	26.0	39.7	40.7	38.4
MS2–6	99.9	99.9	99.9	99.9	99.9	99.9
MS2–7	100.0	100.0	100.0	100.0	100.0	100.0
MS2–8	100.0	100.0	100.0	100.0	100.0	100.0
MS2AR–1	58.4	74.0	1.6	37.6	34.1	13.7
MS2AR–2	84.3	91.8	28.9	80.5	80.9	67.9
MS2AR–3	98.6	99.4	89.9	98.9	99.2	98.9
MS2AR–4	99.9	99.9	99.4	99.8	99.9	99.9
MS2AR–5	84.2	90.7	27.9	79.3	80.1	68.1
MS2AR–6	99.8	99.9	99.4	99.8	99.8	99.8
MS2AR–7	100.0	100.0	100.0	100.0	100.0	100.0
MS2AR–8	100.0	100.0	100.0	100.0	100.0	100.0
MS3–1	9.6	8.6	3.0	9.5	9.6	10.7
MS3–2	39.5	34.7	60.9	59.0	61.0	62.8
MS3–3	83.5	83.9	90.3	91.1	90.8	89.8
MS3–4	95.3	95.2	95.8	95.5	95.0	95.6
MS3–5	43.9	39.6	60.6	59.4	58.9	60.9
MS3–6	95.2	95.1	96.1	95.8	95.8	95.8
MS3–7	97.1	97.1	97.1	96.1	96.1	97.1
MS3–8	98.7	98.7	98.7	97.9	97.9	98.7
MS3AR–1	12.2	10.9	3.3	13.4	12.7	15.5
MS3AR–2	15.7	12.6	17.8	20.4	21.3	25.3
MS3AR–3	53.9	51.0	67.4	66.5	67.7	63.6
MS3AR–4	77.4	78.3	86.9	88.8	87.9	84.2
MS3AR–5	18.5	15.8	18.1	23.0	23.2	25.7
MS3AR–6	79.1	78.7	86.0	88.6	88.5	84.7
MS3AR–7	95.4	94.9	95.7	95.4	95.4	95.5
MS3AR–8	97.6	97.6	97.6	96.6	96.4	97.5

The AR(1) coefficient, when present, is equal to 0.7 in all DGP. The transition probabilities in the MS(2) and MS(2)–AR(1) DGPs are $p_{11} = 0.9$ and $p_{22} = 0.8$; in the MS(3) and MS(3)–AR(1) DGPs are $p_{11} = 0.9$, $p_{12} = 0.07$, $p_{22} = 0.8$, $p_{21} = 0.15$, $p_{33} = 0.7$, $p_{32} = 0.2$. The table shows the percentage of cases, for each index, on 1000 Monte Carlo replications, where the correct number of states is identified.

4 Final Remarks

Detecting the number of states is a crucial task in estimating Markov Switching models and they are usually fixed a priori due to the impossibility of applying classical statistical tests. Then, in this work we propose to identify the number of regimes before the estimation step through a fuzzy clustering, a non parametric approach available in several statistical packages . To this end, we compare throgh Monte Carlo simulations the regime inference of Markov Switching models and a fuzzy clustering algorithm. We show that they provide very similar classifications as the differences between the simulated mean

regimes increases and the variability decreases. Furthermore, we verify through the most common cluster validation indices the capability of fuzzy clustering approach to identify the correct number of groups: it performs very well except for the cases of two regimes with close mean coefficients. As future research it could be interesting to consider other distance measures, such as the Mahalanobis one, or medoid-based approaches opposed to the centroid-based one proposed here. Then, it is possible to compare the different fuzzy clustering algorithms to evaluate which one performs the best.

References

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press.

Campello, R. J. and E. R. Hruschka (2006). A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets and Systems 157*(21), 2858–2875.

Dave, R. N. (1996). Validating fuzzy partitions obtained through c-shells clustering. *Pattern recognition letters 17*(6), 613–623.

Davies, R. (1977). Hypotheses testing when a nuisance parameter is present only under the alternative. *Biometrika 174*, 247–254.

Degras, D., C.-M. Ting, and H. Ombao (2022). Markov-switching state-space models with applications to neuroimaging. *Computational Statistics and Data Analysis 174*, 107525.

D'Urso, P. (2015). *Fuzzy Clustering*, pp. 545–566. Handbook of Macroeconomics (C. Hennig, M. Meila, F. Murtagh and R. Rocci, eds.). Chapman and Hall.

Gallo, G. and E. Otranto (2015). Forecasting realized volatility with changing average volatility levels. *International Journal of Forecasting 31*, 620–634.

Gallo, G. M., D. Lacava, and E. Otranto (2021). On classifying the effects of policy announcements on volatility. *International Journal of Approximate Reasoning 134*, 23–33.

Garcia, R. (1998). Asymptotic null distribution of the likelihood ratio test in markow-switching models. *International Economic Review 39*, 763–788.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*, 357–384.

Hamilton, J. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics 45*, 39–70.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Hamilton, J. D. (2016). *Macroeconomic Regimes and Regime Shifts*, Volume 2 of *Handbook of Macroeconomics (J. B. Taylor and H. Uhlig,eds.)*, pp. 163–201. Elsevier.

Hansen, B. (1992). The likelihood ratio test under nonstandard conditions: Testing the markov switching model of gnp. *Journal of Applied Econometrics 7*, S61–S82.

Hennig, C. and C.-J. Lin (2015). Flexible parametric bootstrap for testing homogeneity against clustering and assessing the number of clusters. *Statistics and Computing 25*, 821–833.

Hubert, L. and P. Arabie (1985). Comparing partitions. *Journal of Classification 2*, 193–218.

Otranto, E. and G. M. Gallo (2002). A nonparametric bayesian approach to detect the number of regimes in markov switching models. *Econometric Reviews 21*, 477–496.

Psadarakis, Z. and N. Spagnolo (2003). On the determination of the number of regimes in markov–switching autoregressive models. *Journal of Time Series Analysis 24*, 237–252.

Psadarakis, Z. and N. Spagnolo (2006). Joint determination of the state dimension and autoregressive order for models with markov regime switching. *Journal of Time Series Analysis 27*, 753–766.

Rand, W. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association 66*, 846–850.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics 20*, 53–65.

Smith, A., P. A. Naikb, and C.-L. Tsai (2006). Markov-switching model selection using kullback–leibler divergence. *Journal of Econometrics 134*, 553–577.

Xie, X. L. and G. Beni (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis & Machine Intelligence 13*(08), 841–847.

Ultimi Contributi di Ricerca CRENoS

I Paper sono disponibili in: http://www.crenos.unica.it

23/03 *Mario Agostino Maggioni, Emanuela Marrocu, Teodora Erika Uberti, Stefano Usai,* "The role of localised, recombinant and exogenous technological change in European regions"

23/02 *Bianca Biagi, Laura Ciucci, Claudio Detotto, Manuela Pulina,* "University study programmes and students dynamics"

23/01 *Giovanni B. Concu, Claudio Detotto, Marco Vannini,* "Drivers of intentions and drivers of actions: willingness to participate versus actual participation in fire management in Sardinia, Italy"

22/06 *Alberto Tidu, Stefano Usai, Frederick Guy,* Measuring spatial dispersion: an experimental test on the M-index

22/05 *Luca Scaffidi Domianello, Giampiero M. Gallo, Edoardo Otranto,* "Smooth and Abrupt Dynamics in Financial Volatility: the MS-MEM-MIDAS"

22/04 *Claudio Detotto, Riccardo Marselli, Bryan C. McCannon, Marco Vannini,* "Experts and Arbitration Outcomes: Insights from Public Procurement Contract Disputes"

22/03 *Gerardo Marletto, Cécile Sillig,* "Global influence and national diversity in socio-technical transitions: a sectoral taxonomy"

22/02 *Federico Aresu, Emanuela Marrocu, Raffaele Paci,* "Public capital and institutions' quality in the Italian regions"

22/01 *Gabriele Cardullo, Maurizio Conti, Andrea Ricci, Sergio Scicchitano, Giovanni Sulis,* "On the Emergence of Cooperative Industrial and Labor Relations"

21/11 *Gianfranco Atzeni, Luca G. Deidda, Marco Delogu, Dimitri Paolini,* "Effects of migration with endogenous labor supply and heterogeneous skills"

21/10 *Alberto Tidu, Stefano Usai, Frederick Guy,* "Agglomeration in manufacturing and services: an experimental application of a distance-based measure to Sardinia"

21/09 *Andrea Caria, Fabio Cerina, Marco Nieddu,* "Political Selection and Monetary Incentives in Local Parliamentary Systems"

21/08 *Giorgio Garau, Alessio Tola, Maria Veronica Camerada, Salvatore Lampreu, Silvia Carrus,* "Economic and social polarization dynamics in the EU"

21/07 *Emanuela Marrocu, Raffaele Paci, Stefano Usai,* "Direct and indirect effects of universities on European regional productivity"

21/06 *Pasqualina Arca, Gianfranco Atzeni, Luca Deidda,* "The Signalling Role of Trade Credit on Loan Contracts: Evidence from a Counterfactual Analysis"

21/05 *Grazia Sveva Ascione, Laura Ciucci, Claudio Detotto, Valerio Sterzi,* Do universities look like patent trolls? An Empirical Study of University Patent Infringement Litigation in the United States

21/04 *Michele Battisti, Massimo Del Gatto, Antonio Francesco Gravina, Christopher F. Parmeter,* "Robots versus labor skills: a complementarity / substitutability analysis"

21/03 *William Addessi, Marco Delogu,* "Infrastructure Accumulation in Developing Countries: the Role of the Informal Sector"

21/02 *Luca De Benedictis, Vania Licio, Anna Maria Pinna,* "From the historical Roman road network to modern infrastructure in Italy"

21/01 *Silvia Balia, Rinaldo Brau, Marco G. Nieddu,* "Depowering Risk: Vehicle Power Restriction and Teen Driver Accidents in Italy"

20/08 *Giampiero M. Gallo, Demetrio Lacava, Edoardo Otranto,* "On Classifying the Effects of Policy Announcements on Volatility"

20/07 *Luc Bauwens, Edoardo Otranto,* "Modelling Realized Covariance Matrices: a Class of Hadamard Exponential Models"

www.crenos.unica.it

ISBN 9788868514655