

Proceeding Paper

An Advanced Markov Switching Approach for the Modelling of Consultation Rate Data [†]

Emmanouil-Nektarios Kalligeris ^{1,‡}, Alex Karagrigoriou ^{1,‡} and Christina Parpoula ^{2,*}

¹ Lab of Statistics and Data Analysis, Department of Statistics and Actuarial-Financial Mathematics, University of the Aegean, 83200 Samos, Greece; ekalligeris@aegean.gr (E.-N.K.); alex.karagrigoriou@aegean.gr (A.K.)

² Department of Psychology, Panteion University of Social and Political Sciences, 17671 Athens, Greece

* Correspondence: chparpoula@panteion.gr

† Presented at the 7th International conference on Time Series and Forecasting, Gran Canaria, Spain, 19–21 July 2021.

‡ These authors contributed equally to this work.

Abstract: Regime switching in conjunction with penalized likelihood techniques could be a robust tool concerning the modelling of dynamic behaviours of consultation rate data. To that end, in this work we propose a methodology that combines the aforementioned techniques, and its performance and capabilities are tested through a real application.

Keywords: elastic net; dynamic system; markov switching; penalized likelihood; regimes; regularization methods; two-state modelling; variable selection



check for
updates

Citation: Kalligeris, E.-N.; Karagrigoriou, A.; Parpoula, C. An Advanced Markov Switching Approach for the Modelling of Consultation Rate Data. *Eng. Proc.* **2021**, *5*, 2. <https://doi.org/10.3390/engproc2021005002>

Academic Editors: Ignacio Rojas, Fernando Rojas, Luis Javier Herrera and Hector Pomare

Published: 24 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Dynamic behaviours in time events are always quite complex, and their modelling is often a challenging task. The level of difficulty is accelerated in cases where the dynamics of a system cannot be satisfactorily described by linear models, but more perplexed non-linear functions are required. Classical time series approaches are not capable of capturing complex functional behaviours. Even advanced models recently proposed are not flexible enough and as a result are not easily adjusted to handle more general non-linear schemes [1–4].

It is also frequently observed that the behaviour changes its general pattern in different regions of the time space. Such changes may affect either the mean or the variation or both. A breakthrough in this field took place 40 years ago with the proposal of Markov regime models and the switching regressions [5]. Such models allow a great degree of flexibility, and as a result they could be implemented to capture complex dynamic behaviours. The unobserved state variable associated with such models is an attractive feature directly related to the switching mechanism of the underlying modelling approach. The resulting advanced models rely on the Markovian property, which is an easily handled issue in terms of inferential statistics. Note finally that censoring [6] or semi-Markov approaches [7] may also be considered in such frameworks.

In this work, within the switching framework, we introduce the classical likelihood combined with a penalty term controlled by a properly chosen tuning parameter. In other words, the switching modelling technique is combined with the so-called penalized likelihood with the parameter estimation being dealt via the Expectation-Maximization algorithm [8]. Depending on the phenomenon under investigation, a proper switching model can be used. Two states often suffice to describe the classical dynamic behaviour of incidence data or epidemics, with one state representing the normal stage of the phenomenon and the other the outbreak stage. In such a case, the frequency (usually of daily or weekly data) changes (increases) considerably (and in some cases dramatically) when

the system enters into the second stage. Such a change is considered statistically significant, and therefore the unobserved state variable ignites the switch. In addition, possible covariates may affect the variable of interest, which is denoted by y_t and represents either the frequency or the associated rate.

2. The Modelling

The model used in this work is the 2-state switching model of conditional mean, the general form of which is given by

$$x_t = \mu_{s_t} + \sum_{i=1}^p \phi_{is_t} x_{t-i} + \epsilon_t, \quad (1)$$

where μ_{s_t} is a switching intercept, ϕ_{is_t} , $i = 1, \dots, p$, are autoregressive (AR) switching coefficients, s_t represents the state variable that takes the values 1 (normal or typical state), and 2 (the extreme or outbreak state) and ϵ_t are *i.i.d.* random variables with zero mean and variance σ_ϵ^2 .

If k covariates are allowed to enter into the model, (1) extends to

$$x_t = \mu_{s_t} + \sum_{i=1}^p \phi_{is_t} x_{t-i} + \sum_{j=1}^k \theta_{js_t} W_j + \epsilon_t, \quad (2)$$

where θ_{js_t} the coefficient associated with the W_j covariate.

As it is clear from the presentation of the above model, a different set of parameters is involved for each state considered. It is important to state that the set of covariates involved in each state may or may not be the same.

3. The Algorithmic Procedure

The approach we choose to follow for modelling phenomena that exhibit a dynamic switching behaviour consists of three steps, which are briefly discussed in this section.

3.1. Step 1-The Change Point Detection

The detection of a change point in a time series and in general in events over time constitutes an integral part of time series analysis, since their identification is directly related to a distributional change. Such changes, even light ones, should cause alarm due to the fact that they may alter the data generating process in such a way that the process under investigation may fail to fulfill the purpose for which it is intended. Applications can be found in most scientific fields from finance and business to engineering, biosciences, climatology, geosciences etc. [9–11].

The proposed methodology requires a preliminary analysis to identify a set of possible change points. It should be noted that such analysis involves only the response variable, and no covariates are involved. The method to be used may be the classical method of change-point identification [12].

For the change point detection, an offline algorithm is used to examine the entire set of observations in a single step to recognize where the change occurred. The online approach could be chosen instead, as long as a certain number of new data are available for the algorithm to function properly and satisfactorily.

To check the performance of the selected change points, we have used the mean absolute error (MSE) according to which predicted and actual values are compared. The general expression is given by

$$MSE = \frac{1}{T} \sum_{i=1}^T (\hat{x}_t - x_t)^2.$$

3.2. Step 2-The Variable Identification

For the identification of the statistically significant covariates, a kind of model selection technique can be implemented. In this work, we propose the use of computationally advanced regularization methods, such as Lasso, Ridge or Elastic-Net with the latter considered to be a generalization of the former ones that overcomes their disadvantages. For the interested reader, a number of articles investigate the interrelation of time series and regularization techniques [13–15].

The generalized regularization method used in this work is given by

$$SSE + 2T\lambda \left[\alpha \left(\sum_{i=1}^p \sum_{j=1}^k (|\phi_{is_t}| + |\theta_{js_t}|) \right) + \left(\frac{1-\alpha}{2} \right) \left(\sum_{i=1}^p \sum_{j=1}^k (\phi_{is_t}^2 + \theta_{js_t}^2) \right) \right], \tag{3}$$

where SSE is the sum of squared errors or any other loss function chosen by the researcher, T the sample size, $\alpha \in [0, 1]$, and λ the tuning parameters that result the penalty in the loss function. Note that α balances the amount of emphasis given to minimize the loss function versus minimizing the sum of squared coefficients and/or the sum of absolute coefficients.

Observe that the above generalized regularization method is reduced to

- The Lasso method for $\alpha = 1$;
- The Ridge method for $\alpha = 0$; and
- To Elastic-Net for $\alpha \in (0, 1)$.

Note that a proper weighted version of (3) can be used if it is needed, for instance, to resolve a heteroscedasticity issue. In such a case, (3) takes the general form

$$SSE + 2T\lambda \left[\alpha \left(\sum_{i=1}^p \sum_{j=1}^k (|w_i^\phi \phi_{is_t}| + w_j^\theta |\theta_{js_t}|) \right) + \left(\frac{1-\alpha}{2} \right) \left(\sum_{i=1}^p \sum_{j=1}^k (w_i^\phi \phi_{is_t}^2 + w_j^\theta \theta_{js_t}^2) \right) \right], \tag{4}$$

where w_i^ϕ and w_j^θ appropriate weights, $i = 1, \dots, p$ and $j = 1, \dots, k$.

3.3. Step 3—The Switching

The selected model for each state is obtained together with the parameter estimates and the associated standard errors.

Note that in practice, we do not know and we do not observe the state s_t , but we could infer it from the observed data. Indeed, although the state variable s_t is an unobserved variable, the process y_t is observed. To make an inference about s_t , we need to make an assumption about the process s_t , which usually is assumed to follow a first order Markov chain. For the 2-state case, the probabilities of transition are also obtained. Thus, the transition probability to state j at time t , given that the process was in state i at the time point $t - 1$, is given by $P(s_t = j | s_{t-1} = i) = p_{ij}$, $i, j = 1, 2$.

4. An Application on Epidemiology

Using a data set of 105 weekly influenza-like-illness (ILI) consultation rate data for Greece for the period 2014–2016, including a series of meteorological and climatological covariates such as temperature and wind, we were able to identify an ideal tuning parameter λ and the best value of the index α in the sense that the minimum mean squared error is achieved. Figure 1 shows that the ideal value of λ is around one (1), while the best value of

α is between 0.5 and 1. A more detailed analysis reveals that $\lambda = 1$ and $\alpha = 0.729$, with the corresponding value of MSE being equal to 0.0734.

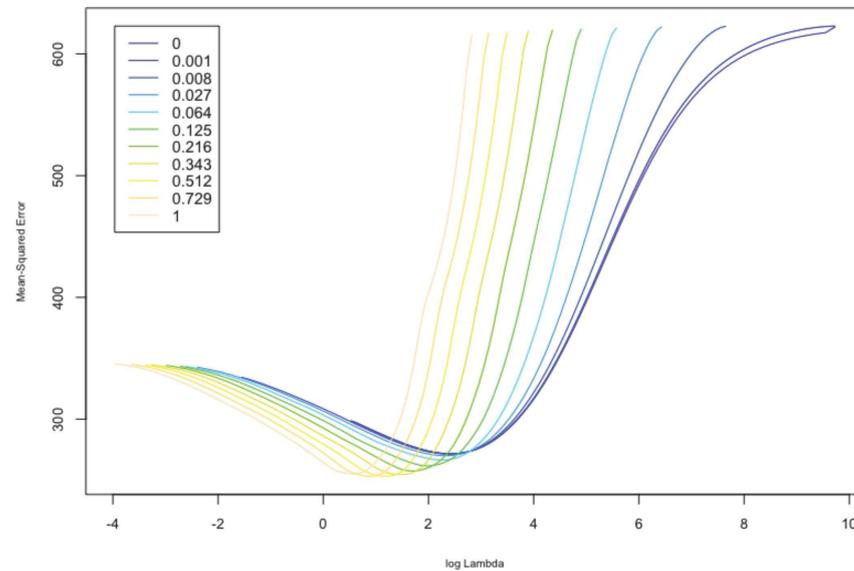


Figure 1. Behaviour for various values of a as opposed to different values of $\log \lambda$ and MSE.

Under this setting, the regimes of the data and consequently the form of the selected models are identified together with the estimates of the parameters involved.

The models obtained with the use of the MSwM R-package [16] are as follows (with three decimal points):

Regime 1-typical period/state

$$\hat{y}_t = 146.480 - 0.375t - 37.488\sin\left(\frac{2\pi t}{n}\right) - 2.682\cos\left(\frac{2\pi t}{n}\right) - 44.002\sin\left(\frac{4\pi t}{n}\right) - 23.945\cos\left(\frac{4\pi t}{n}\right) + 8.330\sin\left(\frac{8\pi t}{n}\right) - 14.244\cos\left(\frac{8\pi t}{n}\right) + 0.035T1 + 15.326T2 - 11.081T3 + 10.042WF - 0.595\hat{y}_{t-1} \quad (5)$$

Regime 2-Outbreak period/state

$$\hat{y}_t = 19.552 - 0.005t + 18.933\sin\left(\frac{2\pi t}{n}\right) - 10.652\cos\left(\frac{2\pi t}{n}\right) - 5.119\sin\left(\frac{4\pi t}{n}\right) - 2.244\cos\left(\frac{4\pi t}{n}\right) + 0.971\sin\left(\frac{8\pi t}{n}\right) - 0.313\cos\left(\frac{8\pi t}{n}\right) - 0.236T1 + 4.006T2 - 4.139T3 + 1.657WF + 0.697\hat{y}_{t-1} \quad (6)$$

It is worth mentioning that the same set of covariances are found to be significant for both regimes together with a first-order autoregressive, a first degree trend polynomial (linear trend) and a periodic (seasonal) part. The covariates chosen are the minimum, mean, and median temperature, denoted, respectively, by $T1$, $T2$, and $T3$ and the mean

of the wind force denoted by WF , implying that the influenza is closely connected to meteorological/climatological factors like the temperature and the wind.

References

1. Kalligeris E.N.; Karagrigoriou, A.; Parpoula, C. On Mixed PARMA Modeling of Epidemiological Time Series Data. *Commun. Stat. Case Stud. Data Anal.* **2019**, *6*, 36–49. [[CrossRef](#)]
2. Pelat, C.; Boëlle, P.Y.; Cowling, B.J.; Carrat, F.; Flahault, A.; Ansart, S.; Valleron, A.J. Online Detection and Quantification of Epidemics. *BMC Med. Inform. Decis. Mak.* **2007**, *5*, 29.
3. Tong, H. Nonlinear Time Series Analysis Since 1990: Some Personal Reflections. *Acta Math. Appl. Sin.* **2002**, *18*, 177. [[CrossRef](#)]
4. Kalligeris, E.N.; Karagrigoriou, A.; Parpoula, C. Periodic-Type Auto-Regressive Moving Average Modeling with Covariates for Time-Series Incidence Data via Change-point Detection. *Stat. Meth. Med. Res.* **2020**, *29*, 1639–1649. [[CrossRef](#)]
5. Lindgren, G. Markov Regime Models for Mixed Distributions and Switching Regressions. *Scand. J. Stat.* **1978**, *5*, 81–91.
6. Huber, C. Efficient Regression Estimation Under General Censoring and Truncation. In *Mathematical and Statistical Models and Methods in Reliability; Statistics for Industry and Technology*; Birkhäuser: Boston, MA, USA, 2010; Volume 12, pp. 235–241.
7. Barbu, V.S.; Karagrigoriou, A.; Makrides, A. Semi Markov Modelling for Multi State Systems. *Methodol. Comput. Appl. Probab.* **2017**, *19*, 1011–1028. [[CrossRef](#)]
8. Green, P.J. On Use of the EM Algorithm for Penalized Likelihood Estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1990**, *52*, 443–452. [[CrossRef](#)]
9. Karagrigoriou, A.; Makrides, A.; Tsapanos, T.; Vougiouka, G. Earthquake Forecasting Based on Multi State System Methodology. *Methodol. Comput. Appl. Probab.* **2016**, *18*, 547–561. [[CrossRef](#)]
10. Votsi, I.; Limnios, N.; Tsaklidis, G.; Papadimitriou, E. Hidden Markov Models Revealing the Stress Field Underlying the Earthquake Generation. *Phys. A* **2013**, *392*, 2868–2885. [[CrossRef](#)]
11. Shaby, B.A.; Reich, B.; Cooley, D.; Kaufman, C.G. A Markov Switching Model for Heat Waves. *Ann. Appl. Stat.* **2016**, *10*, 74–93. [[CrossRef](#)]
12. Page, E.S.: Continuous Inspection Schemes. *Biometrika* **1954**, *41*, 100–115. [[CrossRef](#)]
13. Nardi, Y.; Rinaldo, A. Autoregressive Process Modeling via the Lasso Procedure. *J. Multivar. Anal.* **2011**, *102*, 528–549. [[CrossRef](#)]
14. Chen, K.; Chan, K.S. Subset ARMA Selection via the Adaptive Lasso. *Stat. Interface.* **2011**, *4*, 197–205. [[CrossRef](#)]
15. Medeiros, C.M.; Eduardo, M. L1-Regularization of High-Dimensional Time-Series Models with Non-Gaussian and Heteroskedastic Errors. *J. Econom.* **2016**, *191*, 255–271. [[CrossRef](#)]
16. Sanchez-Espigares, J.A.; Lopez-Moreno, A. MSwM: Fitting Markov Switching Models. CRAN 2018. R Package Version 14. Available online: <https://CRAN.R-project.org/package=MSwM> (accessed on 25 May 2021).