



Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/cstda

Hidden semi-Markov-switching quantile regression for time series

Antonello Maruotti ^{a,b,*}, Lea Petrella ^c, Luca Sposito ^c

^a Dipartimento di Giurisprudenza, Economia, Politica e Lingue Moderne, Libera Università Maria Ss. Assunta, Rome, Italy

^b Department of Mathematics, University of Bergen, Bergen, Norway

^c MEMOTEF, Sapienza Università di Roma, Rome, Italy

ARTICLE INFO

Article history:

Received 8 May 2020

Received in revised form 9 February 2021

Accepted 21 February 2021

Available online 5 March 2021

Keywords:

Asymmetric Laplace distribution

Markov-switching models

Sojourn distribution

Quantile regression

EM algorithm

Maximum likelihood estimation

ABSTRACT

A hidden semi-Markov-switching quantile regression model is introduced as an extension of the hidden Markov-switching one. The proposed model allows for arbitrary sojourn-time distributions in the states of the Markov-switching chain. Parameters estimation is carried out via maximum likelihood estimation method using the Asymmetric Laplace distribution. As a by product of the model specification, the formulae and methods for forecasting, the state prediction, decoding and model checking that exist for ordinary hidden Markov-switching models can be applied to the proposed model. A simulation study to investigate the behaviour of the proposed model is performed covering several experimental settings. The empirical analysis studies the relationship between the stock index from the emerging market of China and those from the advanced markets, and investigates the determinants of high levels of pollution in an Italian small city.

© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Financial, economic, environmental and medical data are often analysed through linear regression models aiming at identifying determinants (covariates) which explain differences between observations. Nevertheless, the traditional regression models are based on assumptions that are not close to the real behaviour of the data at hand, often characterized by specific features like skewness, heavy tails and heterogeneity. Addressing these data features has led to a large number of proposal to properly model different distributional shapes (Rigby and Stasinopoulos, 2005; Stasinopoulos et al., 2018; Punzo et al., 2018). When the interest lies not only in the centre of the response distribution and/or when the observed data may include some outliers, quantile regression represents an interesting alternative to standard mean regression. Indeed calculating the impact of covariates on different parts of the response variable distribution may give insights eventually missed if only the mean is evaluated; see e.g. Koenker and Bassett Jr (1978), Marino and Farcomeni (2015), Bernardi et al. (2018) and Waldmann (2018) for reviews and introductions to quantile regression.

Standard quantile regression methods assume homogeneity across observations, all observations are considered generated by the same probability distribution model. In some scenarios, however, it is of substantive interest to learn if the probability distribution model generating the data, i.e. the data generating mechanism, is different across observations (Bernardi et al., 2012; Miljkovic and Grün, 2016; Punzo et al., 2018). To capture heterogeneity across

* Correspondence to: Dipartimento di Giurisprudenza, Economia, Politica e Lingue Moderne, Libera Università Maria Ss. Assunta, Via Pompeo Magno 22, 00192, Rome, Italy.

E-mail addresses: a.maruotti@lumsa.it, antonello.maruotti@uib.no (A. Maruotti).

observations it is common to assume that latent classes/states are present. In this case, each state can be characterized by a different data generating mechanism (McLachlan and Peel, 2000; Bartolucci et al., 2013; Zucchini et al., 2016). All observations within the same state are characterized by the same data generating mechanism.

During the last few years, the basic homogeneous quantile regression model (Koenker and Bassett Jr, 1978) has been extended to deal with heterogeneity by introducing mixed-effects models (Geraci and Bottai, 2007, 2014) and finite mixtures (Merlo et al., 2021; Del Sarto et al., 2019; Alfò et al., 2017; Wu and Yao, 2016). When time series data are modelled, time-dependent heterogeneity is a crucial data feature to be modelled. Farcomeni (2012) introduces hidden Markov-switching quantile regression models, also known as hidden Markov quantile regression models, where heterogeneity is modelled through a discrete hidden distribution, such that this distribution evolves over time according to a first-order finite-state hidden Markov chain. A similar approach is considered also in Ye et al. (2016) and extended in Marino and Alfò (2015), Marino et al. (2018), Liu and Luger (2018) and Adam et al. (2019).

The general hidden Markov-switching (HMS) regression model is a bivariate discrete-time stochastic process $\{S_t, Y_t\}_{t \geq 0}$ such that

- (1) $\{S_t\}$ is a first-order finite-state Markov chain, i.e. $\Pr(S_t | S_{t-1}, \dots, S_1) = \Pr(S_t | S_{t-1})$;
- (2) $\{Y_t\}$ is a sequence of conditional independent random variables given $\{S_t\}$ and a set of exogenous covariates, such that the parameters of the distribution of Y_t are linked to the exogenous covariates via regression.

However, assumption (1) may be too strong to hold in practice, as it indicates that the current latent state depends only on the most recent latent state in the past; beyond that, it is memoryless. Moreover, the number of consecutive time points that the Markov chain spends in a given state, the so-called sojourn distribution, follows a geometric distribution. Bulla and Bulla (2006) and Maruotti et al. (2019) illustrated some drawbacks of the HMS approach and overcome the lack of flexibility of the HMS model to model the sojourn distribution by using a hidden semi-Markov-switching (HSMS) model (Barbu and Limnios, 2009; Yu, 2015; Bulla et al., 2010), where the sojourn distribution is explicitly modelled rather than implicitly assumed to be geometric.

Motivated by these considerations, in this work we introduce the hidden semi-Markov-switching quantile regression, which has the work of Farcomeni (2012) as special case. Our proposal is characterized by the fact that the sojourn distribution of the process could be arbitrary compared to its Markov-switching counterpart. The greater flexibility obtained employing hidden semi-Markov-switching models carries a computational cost; indeed these models are more demanding to apply than are Markov-switching ones. To reduce the computational burden that often limits the use of HSMS models, we estimate the sojourn probabilities non-parametrically, following Langrock and Zucchini (2011). By introducing a time-series data model with heterogeneous regression coefficients, we are able to capture many attractive features of the data which models currently applied in economics and environmetrics are not often able to grasp. Moreover the time-dependent heterogeneity is captured by using heterogeneous regression coefficients, allowed to vary across quantiles. Up to our knowledge, this represents the first attempt to introduce quantile regression into a semi-Markov-switching setting for the analysis of time series data. Of course, the proposed approach is developed for data showing time-varying heterogeneity and can be viewed also as a further contribution on clustering time-dependent data, where the interest is in the evolution over time of the clustering structure underlying the data at hand (Maharaj et al., 2019; Bartolucci and Farcomeni, 2010; Berchtold and Raftery, 2002).

For inferential purposes, we propose to estimate model parameters through maximum likelihood assuming the Asymmetric Laplace (AL) distribution (Yu and Moyeed, 2001) as inferential tool. We will show how to calculate the observed and the complete data likelihood function considering the observed covariates and hidden states. This can be achieved by slightly modifying the Baum-Welch algorithm, widely used in the HMS and HSMS literature.

We illustrate the proposal by a large-scale simulation study to investigate the empirical behaviour of the proposed approach with respect to several factors, such as the number of observed times, the sojourn distribution and different quantiles as well as different distributions for the error terms. We further apply the proposed methodology to two datasets from different fields. For the first application, we consider pollution data in an Italian urban area; while the second one deals with the emerging financial market of China.

We model the hourly concentrations of $PM_{2.5}$ (i.e. the particulate matter with aerodynamic diameter ≤ 2.5) measured in Battipaglia, in Italy, as a function of some atmospheric variables. In urban areas, the study of pollutant concentrations is of high interest since negative effects of air pollution have been observed on human health (Lee and Sahu, 2016); this has, in particular, been reported for major pollutants such as particulate matter. Local and atmospheric sources of pollution may influence pollution concentrations either directly, indirectly, or both. Empirical and theoretical studies applied sophisticated strategies to link pollution with (observable and unobservable) atmospheric and other sources (Lagona et al., 2011; Park et al., 2001) and modelled the fluctuations of concentrations over long periods. Often pollutant concentrations show extreme values (Martinez-Zaroso and Maruotti, 2013; Merlo et al., 2020) and it is rather important to properly understand if any atmospheric conditions favour the increase of pollution. Moreover, pollution data often comprises periods of flat stretches (also known as background concentrations) and sudden bursts. Therefore, not only the identification of different levels of exposure, but also the investigation of their transition patterns is of interest. These aspects are addressed by the employed HSMS model. Results show that wind speed plays the main role in reducing pollution, and its effect is time-varying and non-linear. The relative humidity also plays a role, while the temperature is not significant.

For financial data, it is well known that the behaviour of the tails of the distribution are important for investment decisions. In this context the use of quantile regression is quite appropriate and indeed several standard measures of market risk are based on quantiles (see for example [Jorion, 1996](#)). In particular, the use of quantile regression avoids strict assumptions such as normality or i.i.d. of the returns. Moreover, it models non-linearity over different quantiles without setting up a complex non-linear parametric framework and, compared to least squares methods, the quantile regression method is more robust to non-Gaussian, leptokurtic, and heavy-tailed distributions, which are common for financial variables ([Maruotti et al., 2019](#)). In the financial data investigation we try to detect possible contagion effects of advanced markets on the Chinese stock market exchange at different quantile levels. Daily returns are modelled via hidden semi-Markov-switching quantile regression, to allow for structural changes ([Xiao, 2009](#)), also considering stock indices from several mature markets on the Asian Pacific rim as possible covariates. Results show the financial market integration is found for some quantiles between China and the other countries.

The remainder of the paper is as follows. In Section 2, we specify the proposed model. Section 3 provides the computational aspects of the adopted maximum likelihood algorithm. In Section 4, we show evidence of the performance of the proposal under different data generation schemes by means of a simulation study. We present two empirical applications on financial and pollution data in Section 5.2 and 5.1 respectively. In Section 6, we conclude pointing out some remarks, along with drawbacks that may arise by adopting the proposed methodology.

2. Methodology

Let $Y_t, t = 1, \dots, T$ denote a real-value observation and S_t be the state of a finite-state semi-Markov chain, defined on the state space $\{1, \dots, k, \dots, K\}$, at time t . A hidden semi-Markov model (HSMM) is a particular kind of time-dependent mixture. The underlying hidden process $\{S_t; t = 1, \dots, T\}$ is constructed as follows. A homogeneous Markov chain with K states models the transitions between different states, with initial probabilities $\pi_k = \Pr(S_1 = k)$ and transition probabilities

$$\pi_{k|j} = \Pr(S_{t+1} = k \mid S_t = j, S_{t+1} \neq k), \tag{1}$$

with $\sum_{k=1}^K \pi_{k|j} = 1$ and $\pi_{j|j} = 0$, i.e. the diagonal elements of the transition probability matrix are zeros. In (1), k refers to the current state, whereas j refers to the one previously visited; this convention will be used throughout the paper. We collect the initial probabilities in the K -dimensional vector $\boldsymbol{\pi}$, whereas the time-homogeneous transition probabilities are collected in the $K \times K$ transition matrix $\boldsymbol{\Pi}$.

We explicitly model the sojourn-time distribution, associated with each hidden state i.e. the time the hidden process S_t spends in the k th state defined as

$$d_k(u) = \Pr(S_{t+u} \neq k, S_{t+u-1} = k, \dots, S_{t+1} = k \mid S_t = k, S_{t-1} \neq k). \tag{2}$$

Hence we assume that the state occupancy distribution is concentrated on finite sets of time points.

The combination of a Markov chain, modelling state changes, and the explication of the sojourn-time distributions defining the hidden process, illustrate the main difference between the HMS model (in which the sojourn-time distribution is implicitly geometric) and the HSMS model. It is worth noting that the semi-Markovian hidden process $\{S_t\}$ of a HSMS model does not have the Markov property at each time t , but is Markovian at the times of state changes only. In most of the empirical settings, we have also a sequence of P exogenous covariates $\mathbf{X}_t = \{X_{t1}, \dots, X_{tP}, \dots, X_{tP}\}$ explaining the sequences of response random vector Y_t . In such a case, a hidden semi-Markov regression model is possible. Moreover, since we are facing financial and environmental time series data, it may be interesting to understand the behaviour of the process around the tails of the distribution, so for this reason it may be useful to embed the problem in a quantile regression framework. In particular for each hidden state k and at a given quantile level τ , we assume the following quantile regression model

$$y_t = \boldsymbol{\beta}'_k(\tau) \mathbf{x}_t^* + \epsilon_{tk}(\tau), \tag{3}$$

with $\boldsymbol{\beta}_k(\tau)$ being a vector of state-specific regression coefficients, $\mathbf{x}_t^* = (1, \mathbf{x}_t)$ the set of covariates accounting for the intercept and $\epsilon_{tk}(\tau)$ is the error term whose τ th conditional quantile equals zero.

It is well known in the literature that the univariate quantile regression approach has a direct link with the Asymmetric Laplace (AL) distribution. That is, the loss function minimization used in quantile regression to estimate regression parameters ([Koenker, 1996](#)) is equivalent (in terms of parameter estimates) to the maximization of the likelihood associated with the AL density; see, e.g., [Yu and Moyeed \(2001\)](#). Therefore, the AL distribution could offer a convenient device to implement a likelihood-based inferential approach when dealing with quantile regression analysis. In our context the AL distribution becomes

$$f(y_t \mid S_t = k; \mathbf{x}_t^*, \boldsymbol{\beta}_k, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp \left\{ -\rho_\tau \left(\frac{y_t - \boldsymbol{\beta}'_k(\tau) \mathbf{x}_t^*}{\sigma} \right) \right\}, \tag{4}$$

where σ is a scale parameter, $\boldsymbol{\beta}'_k(\tau) \mathbf{x}_t^*$ is the quantile at state k and $\rho_\tau(v)$ is the quantile loss function given by

$$\rho_\tau(v) = \begin{cases} \tau v, & v \geq 0 \\ (\tau - 1)v, & v < 0 \end{cases}. \tag{5}$$

The selection of the most appropriate sojourn distribution is instead a complicated problem. Possible solutions may be assessed but the comparison of the empirical sojourn times with a theoretical distribution can lead to unsatisfactory detection. We set all the latent state duration densities to be discrete non-parametric distributions with arbitrary point mass assigned to the feasible duration values so as to allow for the most flexibility (Langrock and Zucchini, 2011). This approach is particularly useful in cases where we do not know the sojourn distribution. We can estimate a non-parametric sojourn distribution, perhaps as an initial step before deciding on a parametric distribution, see e.g. Sansom and Thomson (2001). However, the use of a non-parametric sojourn allows to represent any given sojourn distribution and to approximate any parametric sojourn distributions accurately by choosing the maximum times spent in state k , say U_k , sufficiently large. This comes at the price of using a large number of parameters, but not of the computational times (see for example Langrock and Zucchini (2011)). The model selection procedure becomes a bit more complex as not only the number of hidden states should be selected, but also the U_k values, which may differ across states. However, in practice, $U_k = U = 30$ is large enough in most empirical situations. The total number of parameters is thus given by

- the number of regression parameters $K \times \dim(\boldsymbol{\beta}_k)$;
- $K \times (K - 2)$, which corresponds to the number of independent transition probabilities collected in Π ;
- the sojourn distribution probabilities, $K \times (U - 1)$.

Of course it is not reasonable to try to estimate the initial distribution from just one observation at time 1, especially as the state of the semi-Markov chain itself is not observed. As discussed in Levinson et al. (1983), the maximum likelihood estimator of initial distribution is a unit vector: one entry is 1 and the others 0. Following Leroux and Puterman (1992), no matter which starting values we use for the initial distribution probabilities, it approaches the unit vector.

3. Likelihood inference

Once the number of latent states K has been assigned/fixed, the algorithm basically works on the complete-data likelihood

$$\mathcal{L}_c(\boldsymbol{\vartheta}) = \pi_{s_1^*}^* d_{s_1^*}^*(u_1) \left\{ \prod_{r=2}^{R-1} \pi_{s_r^* | s_{r-1}^*}^* d_{s_r^*}^*(u_r) \right\} \pi_{s_R^* | s_{R-1}^*}^* D_{s_R^*}^*(u_R) \prod_{t=1}^T f(y_t | S_t; \mathbf{x}_t^*, \boldsymbol{\beta}_{s_t}, \sigma, \tau),$$

where s_r^* is the r th visited state, u_r is the time spent in that state (i.e. the duration of the r th visit), $R - 1$ is the number of state changes up to time T , and $\boldsymbol{\vartheta}$ designates the vector of all parameters. In his paper Guédon (2003) proposed to use of the survivor function

$$D_k(u) = \sum_{v \geq u} d_k(v) \tag{6}$$

for the sojourn time in state k . To estimate $\boldsymbol{\vartheta}$, we use the Baum et al. (1970) version of the Expectation–Maximization (EM) algorithm, though alternatives are available in the literature (see e.g. Rydén, 2008; Bulla and Berzel, 2008; MacDonald, 2014). For example, direct numerical maximization of the likelihood using Newton-type algorithms generally converges faster than the EM algorithm, especially in the neighbourhood of a maximum; however, it requires more accurate starting values than the EM to converge. Similarly, Gibbs sampling and other MCMC algorithms, working in a Bayesian context, are likely to require less computation time, but this approach is not the one adopted throughout the paper. The EM algorithm instead provides an iterative solution to maximum likelihood extensively used and preferred when latent variables are presented and when the likelihood evaluation is prohibitively expensive. The EM algorithm alternates the following steps until convergence:

E-step : compute the conditional expected value of the complete-data log-likelihood given the observed data and $\hat{\boldsymbol{\vartheta}}$, the current estimate of $\boldsymbol{\vartheta}$;

M-step : maximize the preceding expected value with respect to $\boldsymbol{\vartheta}$.

In our case with the E-step we calculate

$$Q(\boldsymbol{\vartheta} | \boldsymbol{\vartheta}^{(h)}) = E_{\boldsymbol{\vartheta}^{(h)}} \{ \ln [\mathcal{L}_c(\boldsymbol{\vartheta})] | \mathbf{y}_1, \dots, \mathbf{y}_T \}, \tag{7}$$

the expected complete data log-likelihood given the observed data. In (7), the expectation operator E has the subscript $\boldsymbol{\vartheta}^{(h)}$ to explicitly convey that this expectation is being effected using $\boldsymbol{\vartheta}^{(h)}$ for $\boldsymbol{\vartheta}$. To calculate (7) it is required the calculation of the following three quantities:

1. the probability

$$\gamma_t^{(h)}(k) = P_{\boldsymbol{\vartheta}^{(h)}}(S_t = k | \mathbf{y}_1, \dots, \mathbf{y}_T) \tag{8}$$

of being in state k at time t given the observed sequence;

2. the probability

$$\xi_t^{(h)}(j, k) = P_{\vartheta^{(h)}}(S_{t-1} = j, S_t = k | \mathbf{y}_1, \dots, \mathbf{y}_T) \tag{9}$$

that the process left state j at time $t - 1$ and entered state k at t given the observed sequence;

3. the expected number of times a process spends u time steps in state k ,

$$\eta_t^{(h)}(k, u) = P_{\vartheta^{(h)}}(S_{t-u+1} = k, \dots, S_t = k | \mathbf{y}_1, \dots, \mathbf{y}_T). \tag{10}$$

The first two quantities, $\gamma_t^{(h)}(j)$ and $\xi_t^{(h)}(j, k)$, can be calculated via a dynamic programming method known as the forward-backward algorithm, see [Appendix](#).

3.1. Operational aspects

The likelihood equation will likely have multiple roots corresponding to local maxima. The EM algorithm should be applied from a wide choice of starting values in any search for the maximum, aiming at excluding the so-called spurious maxima. The problem of choosing reliable initial values for the parameters has been often overlooked in the HMSM and HSMSM literature, though few papers recognize potential issues in likelihood inference due to poor initialization of the parameters. [Dunmur and Titterton \(1998\)](#) investigate the failure of reaching the maximum for the EM algorithm, and to this end a simple binary Markov chain structure is studied, which allows detailed exploration of the parameter space. As a result, the choice of initial conditions is found to be highly influential. [Bartolucci and Farcomeni \(2009\)](#) suggest a deterministic rule to initialize the parameters and, then, perform a sensitivity analysis by randomly perturbing those obtained by the deterministic rule to investigate the EM behaviour. [Leroux and Puterman \(1992\)](#) propose a simple procedure for automatically generating good starting value based on data partitions. [Lagona et al. \(2015\)](#) notice that the convergence to spurious maxima is fast and can be detected by monitoring mixing proportions. As remarked by [Maruotti and Rocci \(2012\)](#), a key factor in the fitting of H(S)MSMs is the accuracy of the estimate of the Markov chain parameters; the remaining parameters can be estimated from the inferred partition in accordance with the estimates of the Markov chain parameters. A comparison of several initialization strategies is given in [Maruotti and Punzo \(2021\)](#).

Here, to fit the HSMM with K states, we initialize the EM algorithm by providing the initial parameters, say $\vartheta^{(0)}$, to the first E-step of the algorithm. The initial parameters are computed in the following way. An initial state partition, say $\{S_t^{(0)}\}$, is obtained by using a convenient method. The initial $\Pi^{(0)}$ is computed based on this partition. For the HSMSM, the off-diagonal elements of $\Pi^{(0)}$ are computed as proportions of transition. The sequence of sojourn times in state k is computed from $\{S_t^{(0)}\}$. Finally, as concerns the initial parameters $\beta_k^{(0)}$ and $\sigma_k^{(0)}$, the quantile regression is fitted on the observations in state k . Several strategies can be followed to obtain the initial state partition $\{S_t^{(0)}\}$. Here, we consider a random partition mixed with a short-EM strategy. We classify observations into K clusters according to a Multinomial distribution with parameters $1/K$. From the generated partition we obtain $\vartheta^{(0)}$ as described above. Each run of the EM algorithm is "short" because it is executed for a small number of iterations, without waiting for convergence. Then, the EM algorithm is run from the parameter vector $\vartheta^{(0)}$ providing the largest likelihood from these short runs of EM. Starting the algorithm with good initial values certainly helps reduce the computational cost. The E- and M-steps are alternated repeatedly until the following difference

$$\ell^{(r+1)} - \ell^{(r)} < \epsilon, \quad \epsilon > 0,$$

changes by an arbitrary small amount, i.e. a stopping criterion based on the sequence of likelihood values $\ell^{(r)}$, $r = 1, \dots$ is considered. Since $\ell^{(r+1)} > \ell^{(r)}$, convergence is obtained with a sequence of likelihood value, bounded above.

The algorithm is defined for K fixed. Various authors have discussed algorithms for joint estimation of K and model parameters in a finite mixture context ([Wang, 2010](#)); a possible solution is to update estimates for a fixed K and improving step by step. Once the algorithm has reached its end for a given K , we can proceed to increase the number of components to $K + 1$ and estimate corresponding model parameters. After that, a formal comparison could be done using penalized-likelihood criteria (such as AIC or BIC) or by employing a formal likelihood-ratio test. In the latter case, the number of states is selected by bootstrapping the likelihood ratio test (LRT) statistic for different number of components, since usual regularity conditions for the asymptotic distribution of the LRT statistic do not hold.

The estimation procedure outlined above does not produce standard errors of the estimates, because approximations based on the observed information matrix often require a very large sample size. The parametric bootstrap scheme can also be used to approximate the covariance matrix of parameter estimates to obtain reliable estimation of confidence bands ([Visser et al., 2000](#)). We further investigate the behaviour of the parametric bootstrap under the considered setting in the simulation study.

4. Simulation study

To evaluate the behaviour of the proposed model, a simulation study has been conducted under several different scenarios. In all scenarios, we simulated 500 samples from a two-state HSMS model with varying parameters. In the different scenarios we have analysed three different experimental factors: the sample size ($T = 500; 1000$), the sojourn distribution (non-parametric; Geometric; Shifted Poisson, Shifted Negative Binomial) and the distribution of the error terms (standard Gaussian, skew Student's t with five degrees of freedom). We further provide results with respect to the HMS approach.

First, we simulate from the following model

$$y_t = \begin{cases} \beta_{01} + \beta_{11}x + \epsilon_1 = 1 - 1x + \epsilon_1, & S_t = 1 \\ \beta_{02} + \beta_{12}x + \epsilon_2 = -1 + 1x + \epsilon_2, & S_t = 2 \end{cases} \quad (11)$$

with the error terms $\epsilon_1 = \epsilon_2$ distributed as $N(0; 0.5)$, skew- t with 5 degrees of freedom, alternatively, and $x \sim N(0, 1)$. Different sojourn distributions are considered. Firstly, we generate sojourns from

$$d_k(u) = \begin{cases} 0.2; 0.3; 0.5, & S_t = 1 \\ 0.1; 0.2; 0.7, & S_t = 2 \end{cases}, u = 1, 2, 3;$$

then we simulate from a shifted-Poisson distribution, i.e.

$$d_k(u) = \exp(-\lambda_k) \frac{\lambda_k^{u-1}}{(u-1)!}, u = 1, 2, \dots$$

with $\lambda_1 = 10$ and $\lambda_2 = 5$; and shifted-Negative Binomial, i.e.

$$d_k(u) = \frac{\Gamma(u + \lambda_k - 1)}{(u-1)! \Gamma(\lambda_k)} p_k^{\lambda_k} (1 - p_k)^{u-1}, u = 1, 2, \dots$$

with $\lambda_1 = 10$, $\lambda_2 = 5$ and $p_1 = 0.5$, $p_2 = 0.6$. A geometric sojourn

$$d_k(u) = p_k(1 - p_k)^{u-1}, u = 1, 2, \dots$$

is also considered, with $p_1 = p_2 = 0.2$. The latter corresponds to the Markov-switching quantile regression introduced by Farcomeni (2012). For each chosen combination of sample size, sojourn distribution and error terms distribution we generate 500 datasets and fit our HSMS model and its HMS counterpart. To evaluate model performance in recovering the true model structure, we compared the models on the basis of their ability to accurately estimate the slopes of state-dependent regression coefficient and the latent process parameters, summarized looking at the goodness of classification measures by the Adjusted Rand Index (ARI; Hubert and Arabie, 1985). We measure the accuracy by reporting point estimates and standard errors of the estimators. We started each run from a randomly chosen set of parameters as described in Section 3.1 and stopped the algorithms when the increase in the log-likelihood at two consecutive iterations was less than 10^{-4} .

We report the goodness of classification in Figs. 1 and 2. The inferred classification is obtained via a simple maximum a posteriori rule, also known as local decoding. Firstly, we notice that a slightly better classification is estimated for the HMS under the Geometric sojourn (see the bottom part of Fig. 1). This is not surprising, as an implicit assumption of the HMS is indeed the Geometric sojourn distribution. The HSMS approach so far proposed provides an approximation to any sojourn distributions and its superiority in terms of goodness of classification is rather clear if we compare the two approaches under the non-parametric sojourn. We further consider Shifted Poisson and Shifted Negative Binomial sojourns (see Fig. 2). The HSMS approach provides slightly better ARI values, but the HSM approach is still comparable in terms of classification performance. We further notice that the errors distribution also plays a crucial role in recovering the underlying classification. As expected, assuming Skew- t errors, i.e. heavy tails, makes the classification a bit fuzzier. At last, the goodness of classification is also partially quantile-specific, i.e. the ARI is better at the median and a bit worse if quantiles in the tails are modelled.

In Tables 1 and 2 we report the estimated of the slopes (as in Farcomeni, 2012) at different quantiles ($\tau = 0.1; 0.25; 0.5; 0.75; 0.9$) for all combinations of sojourn and error distributions. Looking at these tables, we may notice that the precision of the estimates is higher at the centre of the distribution when compared to other quantiles. Such differences are due to the reduced amount of information at the tails of the distribution. This effect tends to decrease slightly with increasing sample size (i.e. for $T = 1000$, not reported here for sake of brevity). Overall, the recovering of the data generation parameters by the HSMSM is satisfactory, no matter what sojourn distribution is considered. As a further remark, better goodness of classification leads to more accurate parameters estimates, in line with the findings of Maruotti and Rocci (2012).

We check how well the parametric sojourn distributions parameters are estimated by the non-parametric approach so far proposed (see Table 3). We adopt a two-step procedure. Firstly, we determine the state membership via local decoding; then we fit the inferred sojourn distributions and obtain the provided estimates by maximum likelihood. The results are mixed for $T = 500$. The non-parametric approach is able to reasonably approximate shifted Poisson and Geometric sojourns, for all quantiles and error distributions, whilst its performance is a bit worse if the sojourns follow a

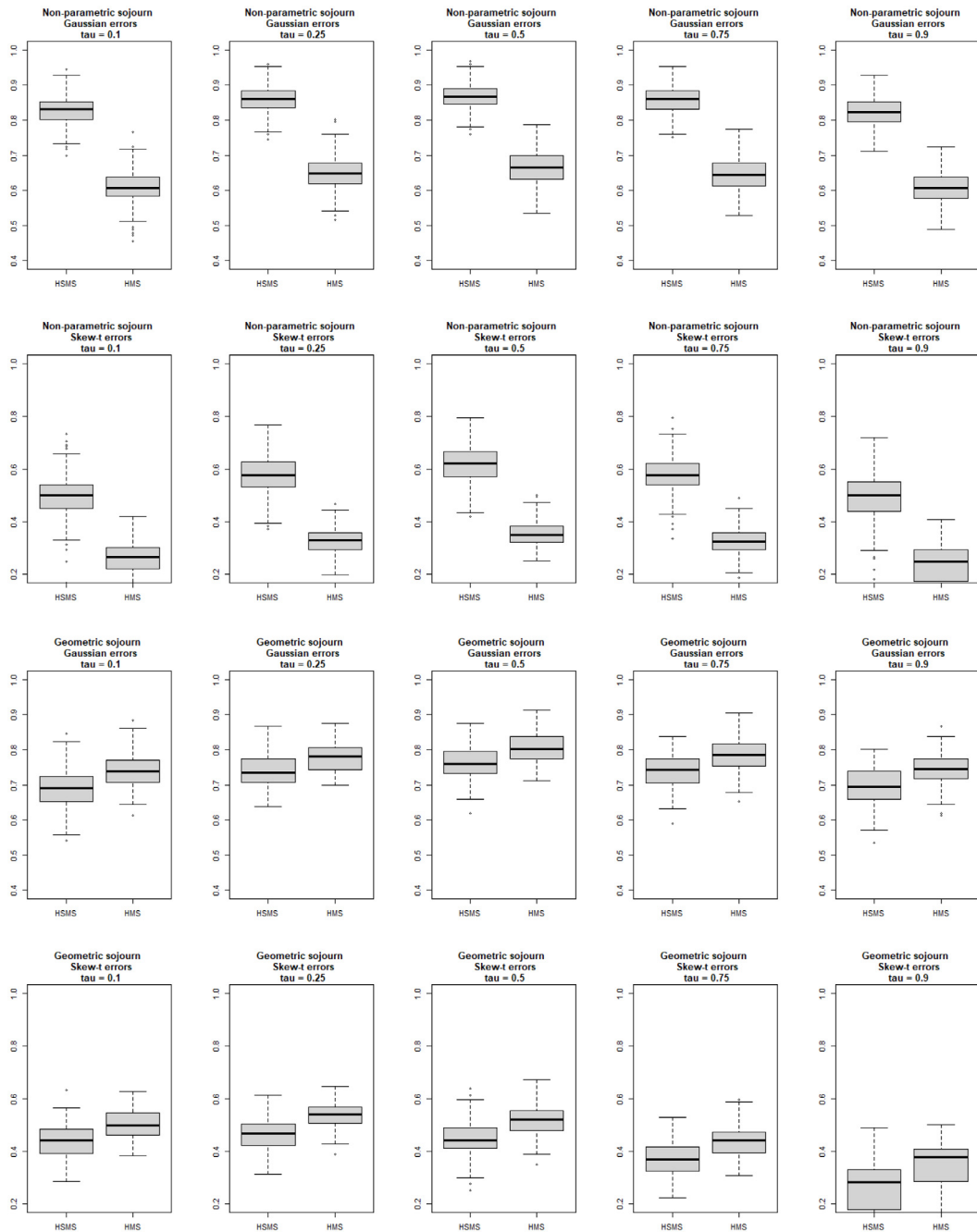


Fig. 1. Adjusted Rand Index: Non-parametric and geometric sojourns.

shifted Negative Binomial distribution. These estimates improve for $T = 1000$, i.e. increasing the sample size (results are not reported here and available upon request). This part of the simulation study aims at showing that there might also be some drawbacks and/or weaknesses in using a non-parametric approach with *small* sample sizes; while it is well-known that its performance increases with the sample size.

At last, we investigate a specific computational aspect related to the estimate of the standard errors. As mentioned before, the EM algorithm does not provide standard errors of the estimates. In the empirical applications, however, we need them to make inference on the regression parameters. An estimation of the regression parameters can be accurate and precise, but if the associated estimation of variance is poor, then coverage by the 95% confidence interval may falsely

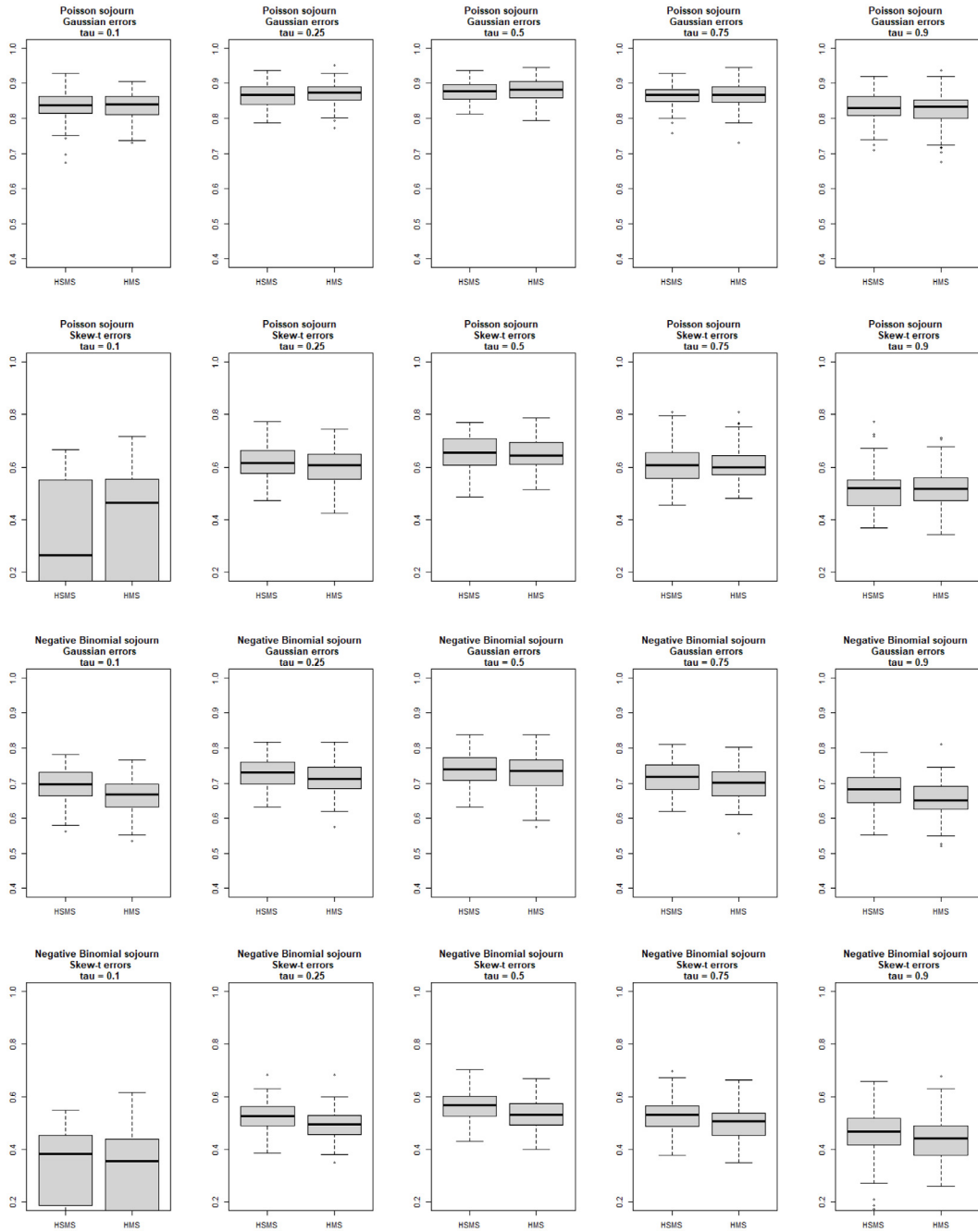


Fig. 2. Adjusted Rand Index: Poisson and negative binomial sojourns.

indicate poor estimation by the point estimator, that is, the point estimator may result in a poor coverage rate. For each of the simulated datasets, using the estimated parameters, 100 more datasets of the same size were simulated. Then, the corresponding standard errors are computed and then used to compute the average standard deviation. We report the results about the HSMS model, with Gaussian errors, in Table 4, along with the 95% coverage probabilities. We compare the standard errors of the estimates and the average standard errors obtained from the parametric bootstrap approach. The parametric bootstrap performs very well and represents a more than reasonable choice for the estimation of the uncertainty surrounding the regression parameters estimates. Its implementation is straightforward, though time consuming, and its behaviour does not qualitatively vary for the skew-t error case.

Table 1
Estimated coefficients (Est.) and standard errors (Std.Err.) over 500 simulations with $T = 500$.

Non-parametric sojourn – Gaussian distributed errors																				
State 1 - $\beta_{11} = 1$										State 2 - $\beta_{12} = -1$										
$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		
Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	
HSMS	1.00	0.06	1.00	0.05	1.00	0.04	1.00	0.05	1.01	0.07	-1.00	0.06	-1.00	0.04	-1.00	0.04	-1.01	0.05	-1.01	0.06
HMM	1.00	0.06	1.00	0.06	0.97	0.20	0.91	0.41	1.02	0.07	-1.01	0.06	-0.99	0.17	-0.99	0.21	-1.01	0.05	-1.00	0.06
Non-parametric sojourn – Skew-t distributed errors																				
State 1 - $\beta_{11} = 1$										State 2 - $\beta_{12} = -1$										
$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		
Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	
HSMS	0.94	0.19	0.94	0.11	0.97	0.10	1.03	0.11	1.09	0.27	-1.09	0.20	-1.04	0.12	-1.01	0.08	-1.00	0.10	-0.99	0.17
HMM	0.92	0.19	0.86	0.31	1.02	0.73	1.01	0.41	0.88	0.71	-1.02	0.55	-1.00	0.40	-1.14	0.86	-1.02	0.37	-0.93	0.22
Geometric sojourn – Normal distributed errors																				
State 1 - $\beta_{11} = 1$										State 2 - $\beta_{12} = -1$										
$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		
Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	
HSMS	0.99	0.06	1.00	0.06	0.98	0.10	1.01	0.05	1.01	0.06	-1.01	0.05	-0.95	0.09	-0.98	0.10	-1.00	0.05	-1.01	0.06
HMS	0.99	0.06	1.00	0.05	1.00	0.04	0.98	0.19	1.01	0.06	-1.01	0.05	-1.00	0.05	-1.00	0.04	-1.00	0.05	-1.01	0.06
Geometric sojourn – Skew-t distributed errors																				
State 1 - $\beta_{11} = 1$										State 2 - $\beta_{12} = -1$										
$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		
Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	
HMSM	0.99	0.11	0.98	0.09	1.05	0.12	1.08	0.32	0.94	0.80	-1.00	0.28	-1.01	0.18	-0.95	0.12	-0.90	0.22	-0.92	0.28
HMS	0.98	0.11	0.97	0.21	1.02	0.19	1.05	0.45	1.02	0.67	-1.01	0.33	-0.96	0.29	-1.04	0.70	-0.85	0.32	-0.92	0.29

Table 2
Estimated coefficients (Est.) and standard errors (Std.Err.) over 500 simulations with $T = 500$.

Poisson sojourn – Gaussian distributed errors																				
State 1 - $\beta_{11} = 1$										State 2 - $\beta_{12} = -1$										
$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		
Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	
HSMS	1.00	0.05	1.00	0.04	1.00	0.04	1.00	0.05	1.01	0.06	-1.02	0.07	-1.00	0.06	-1.00	0.04	-0.99	0.05	-1.00	0.07
HMS	1.01	0.05	1.00	0.04	1.00	0.04	1.00	0.05	1.01	0.06	-0.98	0.24	-1.00	0.06	-1.00	0.04	-0.99	0.05	-1.00	0.07
Poisson sojourn – Skew-t distributed errors																				
State 1 - $\beta_{11} = 1$										State 2 - $\beta_{12} = -1$										
$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		
Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	
HSMS	0.97	0.28	0.96	0.21	0.99	0.08	1.04	0.10	0.94	0.54	-0.37	0.78	-0.96	0.40	-0.98	0.10	-0.95	0.12	-0.75	0.56
HMS	0.97	0.19	0.98	0.10	0.99	0.08	1.04	0.09	0.97	0.46	-0.70	0.91	-1.00	0.36	-0.98	0.11	-0.93	0.13	-0.74	0.48
Negative Binomial sojourn – Normal distributed errors																				
State 1 - $\beta_{11} = 1$										State 2 - $\beta_{12} = -1$										
$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		
Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	
HSMS	1.00	0.07	1.00	0.05	0.98	0.02	0.88	0.05	1.01	0.01	-1.01	0.05	-1.00	0.05	-1.00	0.05	-1.00	0.06	-1.00	0.07
HMS	0.99	0.07	1.00	0.05	0.91	0.04	0.90	0.04	1.00	0.02	-1.01	0.06	-1.01	0.05	-0.97	0.22	-1.01	0.07	-1.00	0.06
Negative Binomial sojourn – Skew-t distributed errors																				
State 1 - $\beta_{11} = 1$										State 2 - $\beta_{12} = -1$										
$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$		
Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	
HSMS	1.00	0.22	0.94	0.29	0.99	0.08	1.02	0.09	0.99	0.37	-0.60	0.70	-1.00	0.33	-0.99	0.13	-0.92	0.15	-0.81	0.35
HMS	0.97	0.21	0.98	0.11	1.00	0.08	1.02	0.09	0.99	0.42	-0.81	0.70	-1.05	0.16	-0.99	0.11	-0.90	0.14	-0.77	0.36

5. Real data

5.1. The analysis of pollution data

The data considered in this example originate from a regional monitoring network system developed in Italy. The data are collected in their raw form, but this leads to difficulties in communicating the overall exposure conditions. Therefore, quantitative methods (models/summary statistics) with the ability to condense the large volume data into a small number of categorical summaries are required. Exposure conditions are, in our case, identified by the hidden states and the corresponding conditional distributions, which provide a condensed overview of their characteristics via the respective concentration profiles. Moreover, we link pollution levels to observed atmospheric variables, which improves the characterization of the different levels of exposure. Atmospheric variables may play a major role in determining the level of exposure to particular pollutants, but also for capturing time dependence as proxies for daily variations. Data are displayed in Fig. 3. From the visual perspective, it is apparent that patterns of non-linear relationship can be detected and that the correlation structure is rather heterogeneous. Existing analyses relate the mean of (log-transformed)

Table 3

Sojourn distributions parameters. Hidden semi-Markov switching model. Estimates with standard errors in brackets.

	$\epsilon_1 = \epsilon_2 \sim N(0; 1)$					$\epsilon_1 = \epsilon_2 \sim skew - t_5$				
	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
Geometric sojourn										
$p_1 = 0.2$	0.149 (0.019)	0.152 (0.021)	0.153 (0.021)	0.156 (0.021)	0.164 (0.023)	0.133 (0.029)	0.130 (0.029)	0.147 (0.034)	0.155 (0.038)	0.209 (0.116)
$p_2 = 0.2$	0.162 (0.025)	0.152 (0.028)	0.151 (0.021)	0.148 (0.023)	0.149 (0.022)	0.151 (0.047)	0.142 (0.046)	0.132 (0.033)	0.133 (0.050)	0.125 (0.043)
Shifted Poisson sojourn										
$\lambda_1 = 10$	10.272 (0.804)	10.172 (0.734)	10.083 (0.708)	9.914 (0.704)	9.661 (0.740)	9.137 (3.704)	10.686 (1.294)	10.316 (1.004)	9.315 (1.052)	7.777 (1.689)
$\lambda_2 = 5$	4.870 (0.435)	4.957 (0.409)	4.999 (0.384)	5.028 (0.401)	5.015 (0.477)	4.644 (3.048)	4.890 (0.976)	5.107 (0.594)	5.329 (0.777)	5.471 (1.594)
Shifted Negative Binomial sojourn										
$\lambda_1 = 10$	11.283 (3.109)	10.615 (1.112)	10.402 (0.873)	10.111 (0.932)	9.413 (1.132)	8.332 (2.420)	9.265 (1.497)	9.492 (1.049)	8.277 (1.059)	7.969 (1.496)
$\lambda_2 = 5$	3.317 (0.753)	3.663 (1.504)	4.509 (0.418)	3.507 (0.439)	3.390 (0.545)	3.339 (2.174)	3.571 (0.824)	3.714 (0.645)	3.769 (0.676)	3.985 (1.977)
$p_1 = 0.5$	0.419 (0.178)	0.492 (0.172)	0.513 (0.163)	0.529 (0.164)	0.503 (0.175)	0.411 (0.239)	0.473 (0.216)	0.518 (0.197)	0.600 (0.222)	0.593 (0.213)
$p_2 = 0.6$	0.400 (0.297)	0.424 (0.265)	0.487 (0.203)	0.429 (0.254)	0.406 (0.288)	0.387 (0.314)	0.414 (0.277)	0.499 (0.213)	0.501 (0.299)	0.490 (0.388)

Table 4

Standard errors, average standard errors obtained from the parametric bootstrap procedure and Coverage (%) probabilities with a nominal 95% level.

τ	Non-parametric sojourn – Gaussian distributed errors					
	$\beta_{11} = 1$			$\beta_{21} = -1$		
	Std. Err.	Av. Std. Err.	Coverage (%)	Std. Err.	Av. Std. Err.	Coverage (%)
0.1	0.06	0.05	96	0.06	0.05	94
0.25	0.05	0.04	94	0.04	0.04	96
0.5	0.04	0.04	95	0.04	0.04	95
0.75	0.05	0.04	97	0.05	0.04	94
0.9	0.07	0.05	95	0.06	0.05	94
τ	Geometric sojourn – Gaussian distributed errors					
	$\beta_{11} = 1$			$\beta_{21} = -1$		
	Std. Err.	Av. Std. Err.	Coverage (%)	Std. Err.	Av. Std. Err.	Coverage (%)
0.1	0.06	0.05	95	0.05	0.04	97
0.25	0.06	0.04	97	0.09	0.05	97
0.5	0.10	0.10	93	0.10	0.10	95
0.75	0.05	0.05	93	0.05	0.04	96
0.9	0.06	0.04	96	0.06	0.05	96
τ	Poisson sojourn – Gaussian distributed errors					
	$\beta_{11} = 1$			$\beta_{21} = -1$		
	Std. Err.	Av. Std. Err.	Coverage (%)	Std. Err.	Av. Std. Err.	Coverage (%)
0.1	0.05	0.04	96	0.07	0.09	95
0.25	0.04	0.04	95	0.06	0.08	97
0.5	0.04	0.04	95	0.04	0.06	95
0.75	0.05	0.04	95	0.05	0.05	94
0.9	0.06	0.06	93	0.07	0.05	96
τ	Negative Binomial sojourn – Gaussian distributed errors					
	$\beta_{11} = 1$			$\beta_{21} = -1$		
	Std. Err.	Av. Std. Err.	Coverage (%)	Std. Err.	Av. Std. Err.	Coverage (%)
0.1	0.07	0.04	96	0.05	0.08	91
0.25	0.07	0.03	96	0.05	0.06	98
0.5	0.02	0.04	96	0.05	0.05	94
0.75	0.05	0.04	96	0.06	0.05	97
0.9	0.01	0.04	97	0.07	0.05	97

concentrations to atmospheric variables, but linear regression is known to perform badly with heavy-tailed and skewed data, also implicitly imposing restrictions on the entire distribution. By using quantile regression, instead, we are able

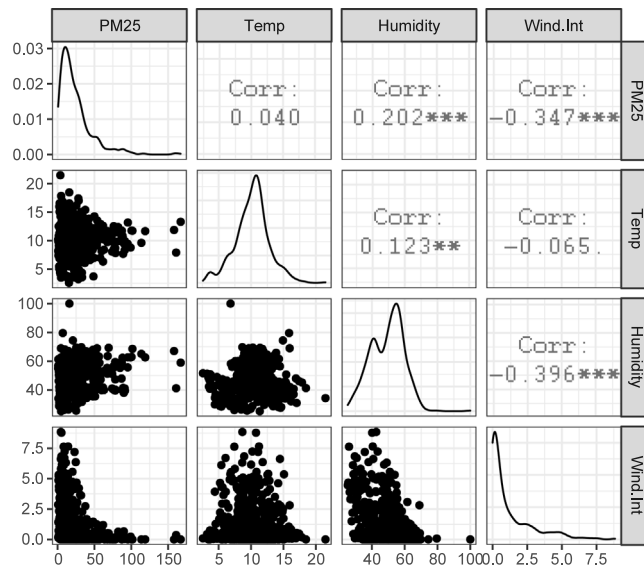


Fig. 3. Pollution data: marginal distributions and bivariate correlations.

Table 5

AIC and BIC values for varying number of hidden states and quantiles. Pollution data.

K	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$
AIC					
2	5412.36	5484.25	5589.70	5686.66	5862.00
3	5259.49	5312.54	5387.52	5506.38	5722.56
4	5205.40	5253.25	5307.26	5387.47	5524.81
5	5149.17	5194.47	5334.70	5369.62	5574.51
BIC					
2	5618.79	5690.69	5796.13	5893.10	6068.44
3	5582.61	5635.65	5710.63	5829.49	6045.67
4	5654.17	5702.02	5756.03	5836.24	6049.58
5	5732.57	5777.87	5918.10	5953.02	6157.91

to model the entire distribution of $PM_{2.5}$ concentrations, focusing on the highest quantiles, where exposure to high concentrations may represent a risk for human health.

The dependent variable is the hourly average of $PM_{2.5}$ concentrations, which is known to be highly skewed, i.e. we observed several $PM_{2.5}$ low concentrations with a very long right tail. The recorded concentrations may strongly be influenced by temperature, relative humidity and wind intensity. The association between $PM_{2.5}$ and atmospheric variables constitutes a crucial aspect of the analysis as well, and should not be neglected or treated as a nuisance. Here, air temperature (Temp), relative humidity (Humidity) and wind intensity (Wind. Int) are considered as potential explanatory variables. Thus, we consider the following empirical model

$$PM_{2.5,t} = \beta_{0k}(\tau) + \beta_{1k}(\tau)Temp_t + \beta_{2k}(\tau)Humidity_t + \beta_{3k}(\tau)Wind.Int_t + \epsilon_{tk}(\tau). \tag{12}$$

The AIC and BIC do not select the same number of hidden states. The BIC is in favour of the 3 States model, while the AIC is increasing up to 5 states (and maybe more), see Table 5.

We comment on the BIC-chosen model. The three states correspond to different levels of (increasing) exposures. State 1 indicates a safe condition, with low pollution. State 2 is the alarming state, because though pollution is not at a warning level, there is a probability of around 0.3 to move to State 3, according to the estimated transition probability matrices. State 3 identifies at-risk exposures, with high and very high pollution. Over the considered period, i.e. 1/1/2019–31/1/2019, State 3 is visited only the 13% of hours, while State 1 and State 2 are visited 49% and 38% of hours, respectively.

We summarize the estimated quantile-specific regression coefficients in Fig. 4. At first hand, it is a general belief that temperature, influencing air stability, would play an important role in particle accumulation, causing an increase in the concentration of $PM_{2.5}$ and probably further aggravates haze pollution. This is not the case in our example. The temperature plays no role in estimating the quantiles of the $PM_{2.5}$ concentrations. On the other hand, the relative humidity and the wind intensity both have significant effects for most of the quantiles in all the 3 hidden states. There are large

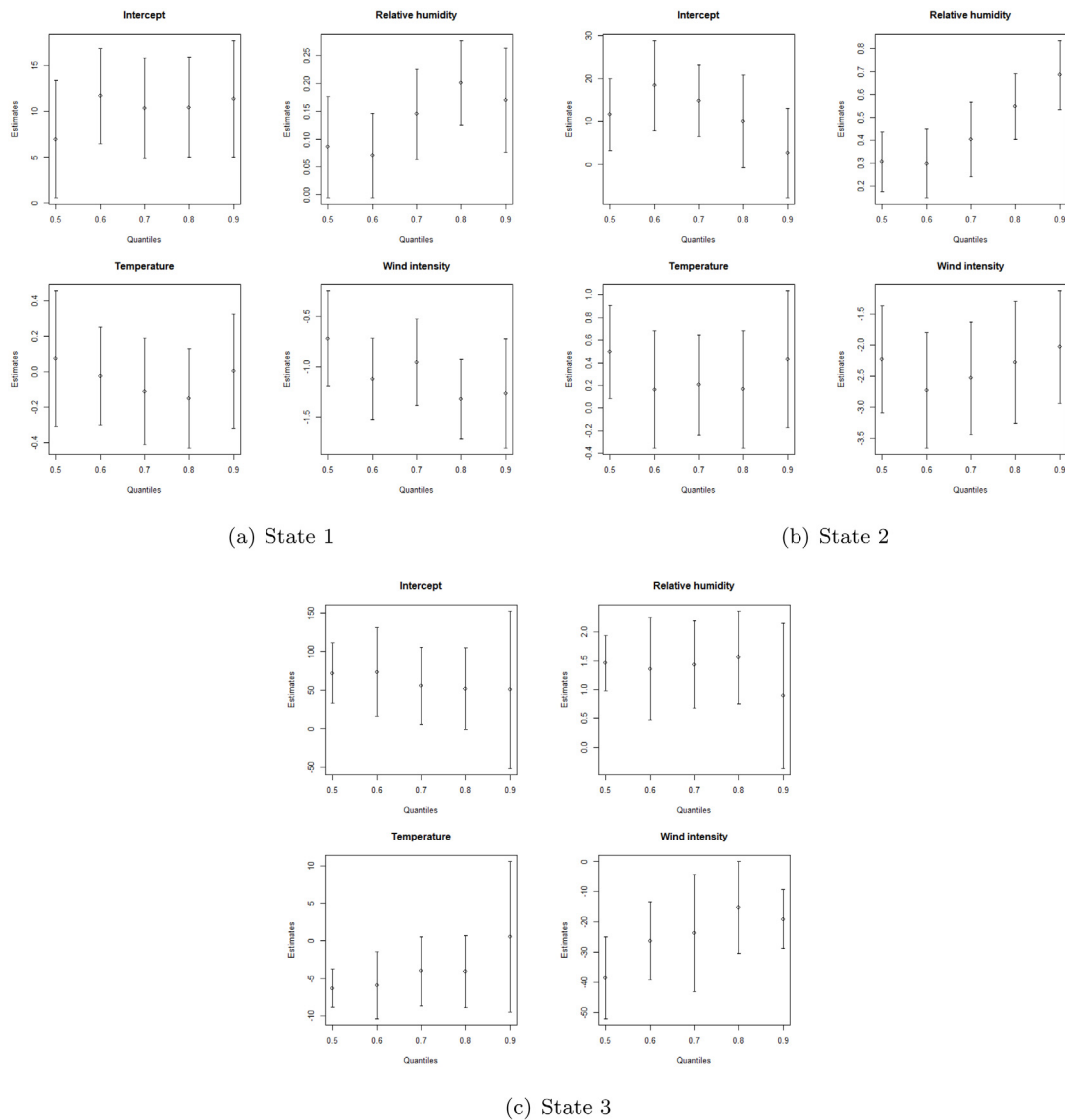


Fig. 4. Pollution data: Regression parameters estimates for varying τ_s . The \circ indicate the parameter estimates and the lines indicate 95% confidence intervals.

gaps in understanding the relationship between relative humidity and $PM_{2.5}$ concentrations, mainly due to the uncertainty about $PM_{2.5}$ sensitivity to relative humidity (Wang et al., 2015; Du et al., 2016; Zhu et al., 2016). The estimated effect of relative humidity is quantile-specific, being non-linear, and state-specific, as the magnitude of the effects varies across states. While relative humidity strongly contributes to the accumulation, development and sustained growth of $PM_{2.5}$ concentrations, its effect is partially alleviated for extreme pollution. This is because an increase in the water in the air could enhance the condensation of water and the formation of rainfall, all of which in turn would mitigate $PM_{2.5}$ concentrations, though increasing the concentrations to a certain extent. As expected, wind intensity plays the major role in reducing pollution, in particular in at-risk situations as those identified by State 3. Wind-blowing is often the only way to reduce pollution from high-risk situations to more affordable ones.

At last, we checked for the usefulness of the non-parametric estimation of the sojourn probabilities using the fast and accurate method proposed by Dimitrova et al. (2020) to compute the cumulative density function of the Kolmogorov–Smirnov statistic when discrete functions are considered, obtaining the exact p values of the Kolmogorov–Smirnov test. Table 6 shows the p -values of the Kolmogorov–Smirnov test statistic. This is rather useful to get more insights on the sojourn distributions. Each state has different sojourn characteristics. It is rather clear that the Shifted Poisson is not a good candidate for the data at hand, i.e. the sojourn times show overdispersion. The geometric and Shifted Negative Binomial may be reasonable at some quantiles to model the sojourns in State 1 and, partially, in State 2, but not in State

Table 6
Kolmogorov–Smirnov tests: p-values. Pollution data.

Null hypothesis	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$
State 1					
Geometric	0.414	0.391	0.326	0.021	0.002
Shifted Poisson	0.001	0.001	<0.001	<0.001	<0.001
Shifted Negative Binomial	0.998	0.995	0.247	<0.001	<0.001
State 2					
Geometric	0.037	0.264	0.153	0.376	0.139
Shifted Poisson	0.001	0.002	<0.001	0.003	<0.001
Shifted Negative Binomial	<0.001	0.500	<0.001	0.600	0.006
State 3					
Geometric	0.025	0.004	0.018	0.121	0.014
Shifted Poisson	<0.001	<0.001	<0.001	<0.001	<0.001
Shifted Negative Binomial	<0.001	<0.001	<0.001	0.042	<0.001

3. So, considering a unique and common (to all quantiles and states) sojourn distribution may be too restrictive. It is not surprising that different sojourns apply to different pollution conditions (Lagona et al., 2011). The non-parametric approach allows for a lot more flexibility. Used jointly with the Kolmogorov–Smirnov test can also be used to select a proper parametric sojourn distribution, if the case.

5.2. The analysis of the Shanghai stock exchange composite index

In this section, we illustrate the practical relevance of our methodology by modelling the Shanghai Stock Exchange (SSE) index and those from the advanced markets such as the United States (SP500) and Japan (NIKKEI). This study also considers stock indices from several mature markets on the Asian Pacific rim, including Hong Kong (HANGSENG), Korea (KOSPI), Singapore (STRAITS), Australia (ASX), and New Zealand (NZAO). We use $T = 3849$ daily returns calculated, spanning the period from January 3, 2005, to October 2, 2019. For each stock market index, daily returns are computed as $[\log(P_t) - \log(P_{t-1})] * 100$, where P_t is the closing price on day t . In line with the ideas discussed in Ye et al. (2016), we consider the following model:

$$\begin{aligned}
 Returns_{SSE,t} = & \beta_{0k}(\tau) + \beta_{1k}(\tau)Returns_{SP500,t} + \beta_{2k}(\tau)Returns_{NIKKEI,t} + \beta_{3k}(\tau)Returns_{HANGSENG,t} \\
 & + \beta_{4k}(\tau)Returns_{KOSPI,t} + \beta_{5k}(\tau)Returns_{STRAITS,t} + \beta_{6k}(\tau)Returns_{ASX,t} \\
 & + \beta_{7k}(\tau)Returns_{NZAO,t} + \epsilon_{tk}(\tau)
 \end{aligned}
 \tag{13}$$

with $Returns_{SSE,t}$ corresponding to returns at time t for the SSE index, and similarly for the other indexes. As it is well known, financial data are characterized by stylized facts like heavy tails, skewness, presence of outliers (Bulla, 2011); this makes the selection of an appropriate parametric distribution challenging. For these reasons a quantile regression framework may be an appropriate modelling approach to be considered. Moreover the presence of changes in volatility has a fundamental impact on the behaviour of returns (Rydén et al., 1998; Ang and Timmermann, 2012; Maruotti et al., 2019).

With the rapid growth China, many investors would certainly consider investing in this market rather than in the advanced or developed markets. However, the question of whether the Chinese market is integrated with other stock markets so that investing in China will provide the benefit of diversification is a major concern for investors. This lead to investigate whether SSE is interdependent with other stock markets.

Comparison of models with differing number of states, and thus parameters, is accomplished via the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), see Table 7. The two chosen model selection criteria lead to different preferred models, and at different quantile we have different levels of heterogeneity, as measured by the number of hidden states. This is not surprising, as e.g. Maruotti et al. (2019) show that different criteria may lead to different chosen models when analysing financial time-series. A $K = 2$ -state model is discussed in the following for ease of discussion for all τ -levels. Of course, model selection is a crucial aspect of any real data analyses and some contributions are present in the Markov-switching literature (see e.g. Pohle et al., 2017). It deserves a specific further investigation under the quantile regression framework.

Results are summarized in Fig. 5. As could be expected, the state-specific intercepts are increasing somewhat with τ , with State 1 having lower values than State 2 for all τ s. Hidden states are often used as a mathematical tool to model time-dependent heterogeneity; however, here, they have a physical meaning being State 1 the *at-risk* state and State 2 the *non-at-risk* state. According to the initial probabilities, at $t = 1$ we start from the non-at-risk state, which is rather reasonable as the series starts from January 2005.

The integration between the Hong Kong and Chinese market is rather evident. The main role in estimating the returns quantiles of the SSE index is played by the Hang Seng index. This would imply that there is a driving force that brings these markets together. Hence, the benefit of diversification between the two markets would be limited. This is not a surprising

Table 7
AIC and BIC values for varying number of hidden states and quantiles. Financial data.

K	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$
AIC					
2	7587.51	6896.47	6429.18	6175.50	6096.87
3	6823.47	6601.94	6224.96	6060.50	6010.62
4	6566.49	6182.56	6132.55	6086.50	5961.07
BIC					
2	8050.40	7359.37	6892.07	6638.39	6559.77
3	7536.57	7315.05	6938.07	6773.61	6723.72
4	7360.92	6976.99	6926.98	6880.93	6755.49

Table 8
Kolmogorov–Smirnov tests: p-values. Financial data.

Null hypothesis	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$
State 1					
Geometric	<0.001	<0.001	<0.001	0.109	<0.001
Shifted Poisson	<0.001	<0.001	<0.001	<0.001	<0.001
Shifted Negative Binomial	<0.001	<0.001	<0.001	<0.001	<0.001
State 2					
Geometric	0.971	0.999	0.999	0.772	0.508
Shifted Poisson	0.002	0.070	0.723	<0.001	<0.001
Shifted Negative Binomial	0.998	0.995	0.997	<0.001	0.799

result. The proximity between the two markets allows the investors to switch investments between the two markets. Moreover, the Hang Seng index in Hong Kong is often seen as *back door* to access to the Chinese market; allowing greater access to investors hoping to get in on growth in mainland China. This also partially shared with the KOSPI in Korea, though the relationship is weaker. Evidence is found that a significant effect exists between the daily returns generated by United States and the SSE. This is not surprising as U.S is the world's foremost stock market and has large influence on other stock markets. Beside that U.S. is also one of the main trading partners for both China. The quantiles of the SSE returns are less affected by the NIKKEI of Japan. Its effect is rather small and significant in the not-at risk state only. There is also an effect of other emerging markets, in particular in State 1, like they share a common systematic risk.

The main achieved goal of our empirical analysis is the evidence of the quantile effects of a large number of financial time series on the SSE. Indeed we found that the number of markets influencing the Chinese one varies across quantiles. The way the other markets influence the SSE is not symmetric and the number of significant markets in the lower tail is larger than that in the upper tail.

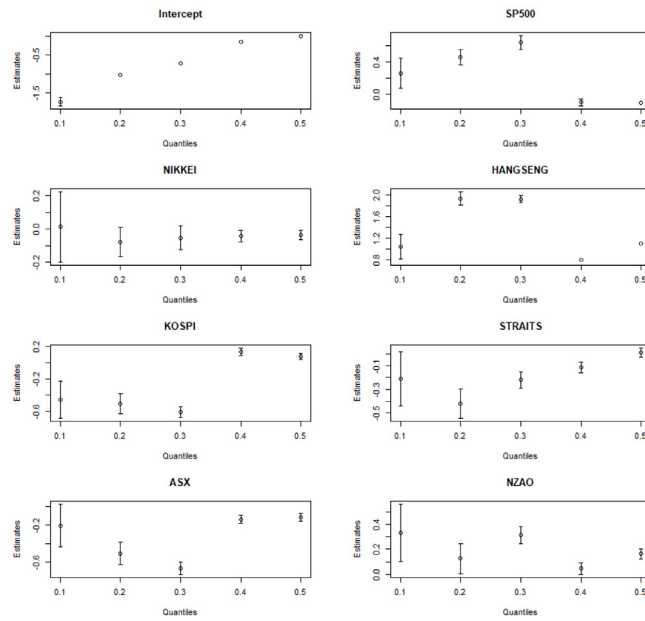
The number of significant markets also varies over time. This is one of the reasons that understanding current financial market structures is important because the market structure changes over time. As a by-product of using quantile regression is the investigation of the relationships between the entire SSE returns distributions and the other markets. The relationship among markets is rather complex, and time-varying. It is worth mentioning that some markets, e.g. the NIKKEI in State 1, the KOSPI and the NZAO in State 2, affect the SSE index only at certain quantiles. In particular the NZAO shows a significant effects in the at-risk State only at extreme quantiles, i.e. the SSE is more connected to Asia Pacific regional markets under at-risk periods, while the NIKKEI affects the SSE only at the centre of the distribution.

The usefulness of relaxing the geometric assumption for sojourn distribution is widely acknowledged in the literature (see e.g. Bulla and Bulla, 2006; Maruotti et al., 2019). This is because the HMS model is not able to capture some stylized facts of the return series (Rydén et al., 1998).

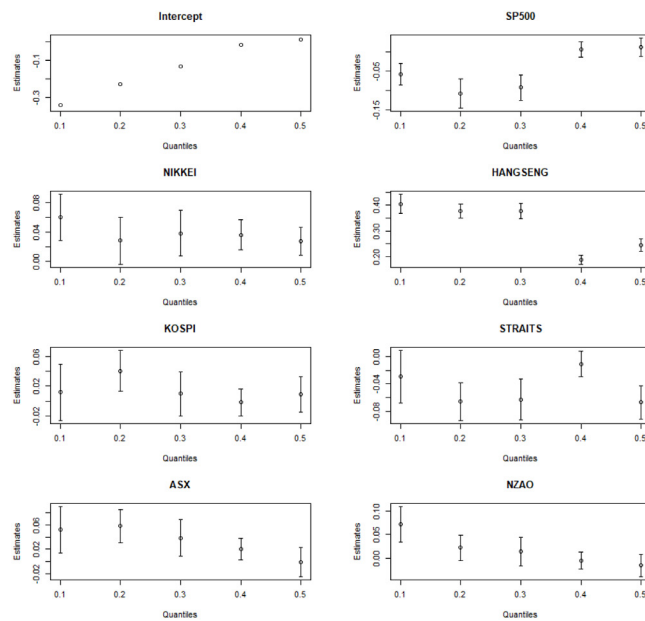
As for the environmental data, we test (Dimitrova et al., 2020) if the estimated sojourn probabilities follow a known parametric distribution, see Table 8. It is rather clear that none of the considered parametric distributions can reproduce the sojourns in State 1; while the sojourn in State 2 can be approximated by the geometric or the Negative Binomial, even better by assuming a non-parametric head and a geometric/negative binomial tail (as in Bulla et al., 2010).

6. Conclusions

We introduced a hidden semi-Markov quantile regression as a model-based clustering approach which overcame the limitations inherent in the Markov-switching one. The model has attractive features, including the feasibility to infer cluster-specific covariate effects on various quantiles, and the flexible way in which the entire response conditional distribution may be modelled and detected. Parameter estimation can be obtained in a maximum likelihood framework, slightly modifying the well-know Baum–Welch algorithm. We further provide the usefulness of the proposal in different empirical fields like finance and environmental frameworks.



(a) State 1



(b) State 2

Fig. 5. Financial data: Regression parameters estimates for varying τ_s . The \circ indicate the parameter estimates and the lines indicate 95% confidence intervals.

The model could be extended towards several directions. Firstly, in a multivariate setting, multiple quantiles can be jointly modelled by using the Multivariate Asymmetric Laplace distribution under proper constraints of the parameters (Petrella and Raponi, 2019). Time dependence can be further included in the quantile regression model allowing the evolution of the quantile over time to vary according to an autoregressive process. The effect of external, independent, covariates may also be included in the latent part of the model, allowing the semi-Markov chain parameters to depend on external information.

The proposal is introduced under the time-series setting. Of course, and as already mentioned in the Introduction, similar approaches are available under the longitudinal setting for HMMs (Marino and Alfó, 2015; Marino et al., 2018).

Our hidden semi-Markov-switching quantile regression model could also be extended to the longitudinal setting as further research work, extending the mixed HSMM approach proposed in Langrock et al. (2012) to quantile regression.

In the analysis of non-normal data in a regression setting, it is relevant to note that Markov-switching generalized additive models for location, shape and scale considered in Hambuckers et al. (2018) and Langrock et al. (2018) could be used as a promising alternative to fit the data, ensuring a lot of flexibility in a parametric setting and a parsimonious approach, as it has fewer model parameters than the proposed Markov-switching quantile model. Of course, a key advantage of Markov-switching quantile regression is that no distributional assumption are required.

A possible limitation of the suggested approach may be the well-known quantile crossing problem that may arise when quantile regression models are fitted separately for each considered quantile probability level. This common practice of treating the quantile levels independently of one another can yield fitted quantile curves that cross one another, thereby leading to a nonsensical response distribution. Indeed, quantile monotonicity requires the quantiles to be increasing as a function of τ , meaning that any well-defined distribution must necessarily have non-crossing quantiles. This quantile crossing problem is potentially worse for non-linear quantile models, like the (semi-)Markov-switching specification developed here. Fortunately, the approach developed by Frumento and Bottai (2016) makes it simple to control for quantile crossing. Further research investigations could be made in this direction.

Acknowledgements

We would like to warmly thank the associate editor and two anonymous reviewers for their valuable comments. We also express deep gratitude to Luca Merlo for the discussions on the draft version of the paper. The research of Antonello Maruotti has been partially supported by the Financial Market Fund (Norwegian Research Council), project number 309218 ‘Statistical modelling and inference for (high-dimensional) financial data’.

Appendix. The forward-backward algorithm

Let us define the forward variable

$$\alpha_t(k, u) = \Pr(S_{t-u+1} = k, \dots, S_t = k, y_1, \dots, y_t)$$

where $k = 1, \dots, K, t = 1, \dots, T$ and $u = \{1, \dots, \min(U_k, t)\}$. Initialize

$$\alpha_1(k, u) = \pi_k$$

and define the recursion

$$\alpha_t(k, u) = \sum_{k=1}^K \sum_{u'=1}^{\min(U_k, t)} \alpha_{t-u}(k, u') \pi_{k|ij} d_k(u) f(y_{t-u+1}, \dots, y_t | S_{t-u+1} = k, \dots, S_t = k; \cdot).$$

Similarly, we define the backward variable

$$\zeta_t(k, u) = \Pr(y_{t+1}, \dots, y_T | S_{t-u+1} = k, \dots, S_t = k)$$

where $k = 1, \dots, K, t = 1, \dots, T$ and $u = \{1, \dots, \min(U_k, t)\}$. Initialize

$$\zeta_T(k, u) = 1$$

and define the recursion

$$\zeta_t(k, u) = \sum_{k=1}^K \sum_{u'=1}^{\min(U_k, T-t)} \pi_{k|ij} d_k(u) f(y_{t+1}, \dots, y_{t+u'} | S_{t+u'} = k, \dots, S_t = k; \cdot) \zeta_{t+u'}(k, u').$$

The quantities used in the E-step can be thus rewritten as

$$\xi_t(j, k) = \sum_{u'=1}^{\min(U_k, t)} \sum_{u=1}^{\min(U_k, T-t)} \alpha_t(k, u') \pi_{k|ij} f(y_{t+1}, \dots, y_{t+u'} | S_{t+u'} = k, \dots, S_t = k; \cdot) \zeta_{t+u}(k, u)$$

$$\eta_t(k, u) = \alpha_t(k, u) \zeta_t(k, u)$$

$$\gamma_t(k) = \sum_{u=1}^{\min(U_k, t, T-t)} \eta_t(k, u)$$

Assuming U_k being the maximum times in state k , Eq. (7) becomes

$$Q(\vartheta | \vartheta^{(h)}) = \sum_{k=1}^K \gamma_1^{(h)}(k) \log \pi_k \tag{A.1}$$

$$+ \sum_{t=1}^T \sum_{k=1}^K \sum_{k \neq j} \frac{\xi_t^{(h)}(j, k)}{\gamma_t^{(h)}(j)} \log \pi_{kj} \quad (\text{A.2})$$

$$+ \sum_{t=1}^T \sum_{k=1}^K \sum_{u=1}^{U_k} \frac{\eta_t^{(h)}(k, u)}{\gamma_t^{(h)}(k)} \log d_k(u) \quad (\text{A.3})$$

$$+ \sum_{t=1}^T \sum_{k=1}^K \gamma_t^{(h)}(k) \log f(y_t | S_t = k; \mathbf{x}_t^*, \boldsymbol{\beta}, \sigma, \tau). \quad (\text{A.4})$$

In the M-step, we can harness the separability of parameters in $Q(\boldsymbol{\vartheta} | \boldsymbol{\vartheta}^{(h)})$ to maximize each component individually as follows

$$\pi_k = \frac{\gamma_1(k)}{\sum_{k=1}^K \gamma_1(k)} \quad (\text{A.5})$$

$$\pi_{kj} = \frac{\sum_{t=1}^T \xi_t(j, k)}{\sum_{k \neq j} \sum_{t=1}^T \xi_t(j, k)} \quad (\text{A.6})$$

$$d_k(u) = \frac{\sum_{t=1}^T \eta_t(k, u)}{\sum_{u=1}^{U_k} \sum_{t=1}^T \eta_t(k, u)} \quad (\text{A.7})$$

$$\boldsymbol{\beta}_k = \arg \min_{\boldsymbol{\beta}_k} \sum_{t=1}^T \gamma_t(k) \rho_\tau(y_t - \boldsymbol{\beta}_k'(\tau) \mathbf{x}_t^*) \quad (\text{A.8})$$

$$\sigma = \frac{1}{T} \sum_{t=1}^T \sum_{k=1}^K \gamma_t(k) \rho_\tau(y_t - \boldsymbol{\beta}_k'(\tau) \mathbf{x}_t^*) \quad (\text{A.9})$$

The existence of (A.8) is readily available in many statistical packages such as the R package `quantreg` (Koenker et al., 2020).

It is known (Levinson et al., 1983) that the maximum likelihood estimator of $\boldsymbol{\pi}$ is a unit vector: one entry is 1 and the others 0. Thus, in practice, the likelihood is maximized conditional on the Markov chain starting in state k , for each of $k = 1, 2, \dots, K$, and then chooses the largest result.

References

- Adam, T., Langrock, R., Kneib, T., 2019. Model-based clustering of time series data: a flexible approach using nonparametric state-switching quantile regression models. In: Proceedings of the 12th Scientific Meeting on Classification and Data Analysis, pp. 19–22.
- Alfö, M., Salvati, N., Ranalli, M.G., 2017. Finite mixtures of quantile and M-quantile regression models. *Stat. Comput.* 27 (2), 547–570. <http://dx.doi.org/10.1007/s11222-016-9638-1>.
- Ang, A., Timmermann, A., 2012. Regime changes and financial markets. *Annu. Rev. Financ. Econ.* 4 (1), 313–337.
- Barbu, V.S., Limnios, N., 2009. *Semi-Markov Chains and Hidden Semi-Markov Models Toward Applications: their Use in Reliability and DNA Analysis*, Vol. 191. Springer Science & Business Media.
- Bartolucci, F., Farcomeni, A., 2009. A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *J. Amer. Statist. Assoc.* 104 (486), 816–831.
- Bartolucci, F., Farcomeni, A., 2010. A note on the mixture transition distribution and hidden Markov models. *J. Time Series Anal.* 31 (2), 132–138.
- Bartolucci, F., Farcomeni, A., Pennoni, F., 2013. *Latent Markov Models for Longitudinal Data*. CRC Press.
- Baum, L.E., Petrie, T., Soules, G., Weiss, N., 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.* 41 (1), 164–171.
- Berchtold, A., Raftery, A.E., 2002. The mixture transition distribution model for high-order Markov chains and non-Gaussian time series. *Statist. Sci.* 328–356.
- Bernardi, M., Bottone, M., Petrella, L., 2018. Bayesian Quantile regression using the skew exponential power distribution. *Comput. Statist. Data Anal.* 126, 92–111.
- Bernardi, M., Maruotti, A., Petrella, L., 2012. Skew mixture models for loss distributions: A Bayesian approach. *Insurance Math. Econom.* 51 (3), 617–623. <http://dx.doi.org/10.1016/j.insmathco.2012.08.002>, URL: <http://www.sciencedirect.com/science/article/pii/S0167668712001035>.
- Bulla, J., 2011. Hidden Markov models with t components. Increased persistence and other aspects. *Quant. Finance* 11 (3), 459–475.
- Bulla, J., Berzel, A., 2008. Computational issues in parameter estimation for stationary hidden Markov models. *Comput. Statist.* 23 (1), 1–18. <http://dx.doi.org/10.1007/s00180-007-0063-y>.
- Bulla, J., Bulla, I., 2006. Stylized facts of financial time series and hidden semi-Markov models. *Comput. Statist. Data Anal.* 51 (4), 2192–2209. <http://dx.doi.org/10.1016/j.csda.2006.07.021>, URL: <http://www.sciencedirect.com/science/article/pii/S0167947306002374> Nonlinear Modelling and Financial Econometrics.
- Bulla, J., Bulla, I., Nenadić, O., 2010. **Hsmm** — An R package for analyzing hidden semi-Markov models. *Comput. Statist. Data Anal.* 54 (3), 611–619.
- Del Sarto, S., Marino, M.F., Ranalli, M.G., Salvati, N., 2019. Using finite mixtures of M-quantile regression models to handle unobserved heterogeneity in assessing the effect of meteorology and traffic on air quality. *Stoch. Environ. Res. Risk Assess.* 33 (7), 1345–1359.
- Dimitrova, D.S., Kaishev, V.K., Tan, S., 2020. Computing the Kolmogorov-Smirnov distribution when the underlying CDF is purely discrete, mixed, or continuous. *J. Stat. Softw.* 95 (1), 1–42.
- Du, H., Song, X., Jiang, H., Kan, Z., Wang, Z., Cai, Y., 2016. Research on the cooling island effects of water body: A case study of shanghai, China. *Ecol. Indic.* 67, 31–38.

- Dunmur, A., Titterton, D., 1998. The influence of initial conditions on maximum likelihood estimation of the parameters of a binary hidden Markov model. *Statist. Probab. Lett.* 40 (1), 67–73.
- Farcomeni, A., 2012. Quantile regression for longitudinal data based on latent Markov subject-specific parameters. *Stat. Comput.* 22 (1), 141–152. <http://dx.doi.org/10.1007/s11222-010-9213-0>.
- Frumento, P., Bottai, M., 2016. Parametric modeling of quantile regression coefficient functions. *Biometrics* 72 (1), 74–84.
- Geraci, M., Bottai, M., 2007. Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics* 8 (1), 140–154.
- Geraci, M., Bottai, M., 2014. Linear quantile mixed models. *Statistics and computing* 24 (3), 461–479.
- Guédou, Y., 2003. Estimating hidden semi-Markov chains from discrete sequences. *J. Comput. Graph. Statist.* 12 (3), 604–639.
- Hambuckers, J., Kneib, T., Langrock, R., Silbersdorff, A., 2018. A Markov-switching generalized additive model for compound Poisson processes, with applications to operational loss models. *Quant. Finance* 18 (10), 1679–1698.
- Hubert, L., Arabie, P., 1985. Comparing partitions. *J. Classification* 2 (1), 193–218.
- Jorion, P., 1996. Risk2: Measuring the risk in value at risk. *Financ. Anal. J.* 52 (6), 47–56.
- Koenker, R., 1996. *Quantile Regression*. Cambridge University Press.
- Koenker, R., Bassett Jr, G., 1978. Regression quantiles. *Econometrica* 33–50.
- Koenker, R., Portnoy, S., Ng, P.T., Zeileis, A., Grosjean, P., Ripley, B.D., 2020. Package 'Quantreg', Version 5.61. Technical Report, Last accessed April 21th.
- Lagona, F., Maruotti, A., Padovano, F., 2015. Multilevel multivariate modelling of legislative count data, with a hidden Markov chain. *J. R. Statist. Soc. - Ser. A* 178 (3), 705–723.
- Lagona, F., Maruotti, A., Picone, M., 2011. Hidden Markov Models: Theory and Applications. InTech, pp. 207–222, Chapter A Non-Homogeneous Hidden Markov Model for the Analysis of Multi-Pollutant Exceedances Data.
- Langrock, R., Adam, T., Leos-Barajas, V., Mews, S., Miller, D.L., Papastamatiou, Y.P., 2018. Spline-based nonparametric inference in general state-switching models. *Stat. Neerl.* 72 (3), 179–200.
- Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., Morales, J.M., 2012. Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology* 93 (11), 2336–2342.
- Langrock, R., Zucchini, W., 2011. Hidden Markov models with arbitrary state dwell-time distributions. *Comput. Statist. Data Anal.* 55 (1), 715–724.
- Lee, D., Sahu, S., 2016. In: Lawson, A., Banerjee, S., Haining, R., Ugarte, L. (Eds.), *Handbook of Spatial Epidemiology*.
- Leroux, B.G., Puterman, M.L., 1992. Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics* 48 (2), 545–558.
- Levinson, S.E., Rabiner, L.R., Sondhi, M.M., 1983. An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *Bell Syst. Tech. J.* 62 (4), 1035–1074.
- Liu, X., Luger, R., 2018. Markov-Switching quantile autoregression: a gibbs sampling approach. *Stud. Nonlinear Dyn. Econometr.* 22 (2).
- MacDonald, I.L., 2014. Numerical maximisation of likelihood: A neglected alternative to EM? *Internat. Statist. Rev.* 82 (2), 296–308.
- Maharaj, E.A., D'Urso, P., Caiado, J., 2019. *Time Series Clustering and Classification*. CRC Press.
- Marino, M.F., Alfó, M., 2015. Latent drop-out based transitions in linear quantile hidden Markov models for longitudinal responses with attrition. *Adv. Data Anal. Classif.* 9 (4), 483–502. <http://dx.doi.org/10.1007/s11634-015-0222-x>.
- Marino, M.F., Farcomeni, A., 2015. Linear quantile regression models for longitudinal experiments: an overview. *METRON* 73 (2), 229–247. <http://dx.doi.org/10.1007/s40300-015-0072-5>.
- Marino, M.F., Tzavidis, N., Alfó, M., 2018. Mixed hidden Markov quantile regression models for longitudinal data with possibly incomplete sequences. *Stat. Methods Med. Res.* 27 (7), 2231–2246. <http://dx.doi.org/10.1177/0962280216678433>, arXiv:<https://doi.org/10.1177/0962280216678433> PMID: 27899706.
- Martinez-Zarzoso, I., Maruotti, A., 2013. The environmental kuznets curve: functional form, time-varying heterogeneity and outliers in a panel setting. *Environmetrics* 24 (7), 461–475.
- Maruotti, A., Punzo, A., 2021. Initialization of hidden Markov and semi-Markov models: a critical evaluation of several strategies. *Internat. Statist. Rev.* <http://dx.doi.org/10.1111/insr.12436>, (in press).
- Maruotti, A., Punzo, A., Bagnato, L., 2019. Hidden Markov and semi-Markov models with multivariate leptokurtic-normal components for robust modelling of daily returns series. *J. Financ. Econ.* 17, 91–117.
- Maruotti, A., Rocci, R., 2012. A mixed non-homogeneous hidden Markov model for categorical data, with application to alcohol consumption. *Stat. Med.* 31 (9), 871–886.
- McLachlan, G.J., Peel, D., 2000. *Finite Mixture Models*. John Wiley & Sons, New York, p. 419.
- Merlo, L., Maruotti, A., Petrella, L., 2021. A two-part finite mixture quantile regression model for semi-continuous longitudinal data. *Statistical Modelling to appear*.
- Merlo, L., Petrella, L., Raponi, V., et al., 2020. Sectoral decomposition of CO2 world emissions: A joint quantile regression approach. *Int. Rev. Environ. Resour. Econ.* 14 (2–3), 197–239.
- Miljkovic, T., Grün, B., 2016. Modeling loss data using mixtures of distributions. *Insurance Math. Econom.* 70, 387–396. <http://dx.doi.org/10.1016/j.insmatheco.2016.06.019>, URL: <http://www.sciencedirect.com/science/article/pii/S0167668715303140>.
- Park, E.S., Guttorp, P., Henry, R.C., 2001. Multivariate receptor modeling for temporally correlated data by using MCMC. *J. Amer. Statist. Assoc.* 96 (456), 1171–1183.
- Petrella, L., Raponi, V., 2019. Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress. *J. Multivariate Anal.* 173, 70–84.
- Pohle, J., Langrock, R., van Beest, F.M., Schmidt, N.M., 2017. Selecting the number of states in hidden Markov models: pragmatic solutions illustrated using animal movement. *J. Agric. Biol. Environ. Stat.* 22 (3), 270–293.
- Punzo, A., Bagnato, L., Maruotti, A., 2018. Compound unimodal distributions for insurance losses. *Insurance Math. Econom.* 81, 95–107. <http://dx.doi.org/10.1016/j.insmatheco.2017.10.007>, URL: <http://www.sciencedirect.com/science/article/pii/S016766871730207X>.
- Rigby, R.A., Stasinopoulos, D.M., 2005. Generalized additive models for location, scale and shape. *J. R. Stat. Soc. Ser. C. Appl. Stat.* 54 (3), 507–554. <http://dx.doi.org/10.1111/j.1467-9876.2005.00510.x>.
- Rydén, T., 2008. EM versus Markov chain Monte Carlo for estimation of hidden Markov models: a computational perspective. *Bayesian Anal.* 3 (4), 659–688. <http://dx.doi.org/10.1214/08-BA326>.
- Rydén, T., Terasvirta, T., Asbrink, S., 1998. Stylized facts of daily return series and the hidden Markov model. *J. Appl. Econom.* 13 (3), 217–244.
- Sansoni, J., Thomson, P., 2001. Fitting hidden semi-Markov models to breakpoint rainfall data. *J. Appl. Probab.* 142–157.
- Stasinopoulos, M.D., Rigby, R.A., Bastiani, F.D., 2018. GAMLSS: a distributional regression approach. *Statist. Model.* 18 (3–4), 248–273.
- Visser, I., Raijmakers, M.E.J., Molenaar, P.C.M., 2000. Confidence intervals for hidden Markov model parameters. *Br. J. Math. Stat. Psychol.* 53 (2), 317–327. <http://dx.doi.org/10.1348/000711000159240>.
- Waldmann, E., 2018. Quantile regression: A short story on how and why. *Statist. Model.* 18 (3–4), 203–218. <http://dx.doi.org/10.1177/1471082X18759142>, arXiv:<https://doi.org/10.1177/1471082X18759142>.
- Wang, Y., 2010. Maximum likelihood computation for fitting semiparametric mixture models. *Stat. Comput.* 20 (1), 75–86.

- Wang, P., Cao, J., Tie, X., Wang, G., Li, G., Hu, T., Wu, Y., Xu, Y., Xu, G., Zhao, Y., et al., 2015. Impact of meteorological parameters and gaseous pollutants on PM_{2.5} and PM₁₀ mass concentrations during 2010 in xian, China. *Aerosol Air Qual. Res.* 15 (5), 1844–1854.
- Wu, Q., Yao, W., 2016. Mixtures of quantile regressions. *Comput. Statist. Data Anal.* 93, 162–176.
- Xiao, Z., 2009. Quantile cointegrating regression. *J. Econometrics* 150 (2), 248–260.
- Ye, W., Zhu, Y., Wu, Y., Miao, B., 2016. Markov Regime-switching quantile regression models and financial contagion detection. *Insurance Math. Econom.* 67, 21–26. <http://dx.doi.org/10.1016/j.insmatheco.2015.11.002>, URL: <http://www.sciencedirect.com/science/article/pii/S0167668715301013>.
- Yu, S.-Z., 2015. *Hidden Semi-Markov Models: Theory, Algorithms and Applications*. Morgan Kaufmann.
- Yu, K., Moyeed, R.A., 2001. Bayesian quantile regression. *Statist. Probab. Lett.* 54 (4), 437–447.
- Zhu, L., Liu, J., Cong, L., Ma, W., Ma, W., Zhang, Z., 2016. Spatiotemporal characteristics of particulate matter and dry deposition flux in the cuihu wetland of Beijing. *PLoS One* 11 (7).
- Zucchini, W., MacDonald, I.L., Langrock, R., 2016. *Hidden Markov Models for Time Series: An Introduction Using R*, second ed. Chapman & Hall, Boca Raton, FL.