Statistics and Computing (2022) 32:61
https://doi.org/10.1007/s11222-022-10130-1

Quantile hidden semi-Markov models for multivariate time series

Luca Merlo[1] · Antonello Maruotti[2,3] · Lea Petrella[4] · Antonio Punzo[5]

Received: 21 September 2021 / Accepted: 24 July 2022 / Published online: 9 August 2022

Abstract
This paper develops a quantile hidden semi-Markov regression to jointly estimate multiple quantiles for the analysis of multivariate time series. The approach is based upon the Multivariate Asymmetric Laplace (MAL) distribution, which allows to model the quantiles of all univariate conditional distributions of a multivariate response simultaneously, incorporating the correlation structure among the outcomes. Unobserved serial heterogeneity across observations is modeled by introducing regime-dependent parameters that evolve according to a latent finite-state semi-Markov chain. Exploiting the hierarchical representation of the MAL, inference is carried out using an efficient Expectation-Maximization algorithm based on closed form updates for all model parameters, without parametric assumptions about the states' sojourn distributions. The validity of the proposed methodology is analyzed both by a simulation study and through the empirical analysis of air pollutant concentrations in a small Italian city.

Keywords EM algorithm · Latent process · Maximum likelihood · Multivariate asymmetric Laplace distribution · Quantile regression · Sojourn distribution

1 Introduction

Since their introduction in the 1960s, Hidden Markov Models (HMMs, see MacDonald and Zucchini 1997; Cappé et al. 2006; Zucchini et al. 2016) have been successfully implemented in a wide range of applications for the analysis of time series data. This class of models is described by an observable stochastic process whose dynamic is governed, completely or partially, by a latent unobservable Markov chain. Owing to their mathematical tractability and the availability of efficient computational procedures, the use of HMMs is well justified when the researcher is interested in inference and/or predictions about the latent process based on the observed one. For a detailed survey of the literature and fields of application, please see MacDonald and Zucchini 1997; Ephraim and Merhav 2002; Cappé et al. 2006; Maruotti 2011; Bartolucci et al. 2012; Zucchini et al. 2016 and Maruotti and Punzo (2021).

One immediate consequence of the Markov property is that in any HMM, the sojourn time (also defined as state duration or dwell-time), that is, the number of consecutive time points that the Markov chain spends in a given state, is implicitly geometrically distributed (Langrock and Zucchini 2011; Zucchini et al. 2016). Despite the popularity of HMMs, this assumption may not be realistic in many applications which can lead to biased parameter estimates and deteriorate the states classification performance due to a misspecification of the dynamic of the hidden process. Bulla and Bulla (2006) and Maruotti et al. (2019), for instance, show the inability of HMMs to model temporal dependence and reproduce empirical characteristics in real-world data, especially when the probability mass function of sojourn times is far from being geometric.

Motivated by these considerations, Hidden Semi-Markov Models (HSMMs, see Yu 2015) are designed to relax this condition by allowing the Sojourn Distributions (SDs) to be modeled directly by the researcher using more flexible parametric or nonparametric distributions.

✉ Luca Merlo
luca.merlo@uniroma1.it

1 Department of Statistical Sciences, Sapienza University of Rome, Piazzale Aldo Moro, 5, 00185 Rome, Italy

2 Department of Mathematics, University of Bergen, Bergen, Norway

3 Department of Law, Economics, Political Sciences and Modern Languages, LUMSA University, Rome, Italy

4 MEMOTEF Department, Sapienza University of Rome, Rome, Italy

5 Department of Economics and Business, University of Catania, Catania, Italy

Typical choices include families of discrete (semi-)parametric distributions or, alternatively, one can avoid distributional assumptions by estimating the SD probability mass functions based on the observations (Sansom and Thomson 2001; Guédon 2003; Pohle et al. 2022). Combining the increased flexibility to capture a wide range of distributional shapes of the SDs with the well-known advantages of HMMs, HSMMs constitute a versatile framework in several spheres of application (see Guédon 2003; Barbu and Limnios 2009; Bulla et al. 2010; O'Connell and Højsgaard 2011; Yu 2015; Maruotti and Punzo 2021 and the references therein).

In the regression context where covariates are available, both HMMs and HSMMs have also been extended to include a set of predictors by introducing state-dependent regression parameters that evolve over time according to the unobserved process (Hamilton 1989; Yu 2015; Zucchini et al. 2016).

This model specification permits to investigate the dynamics of the hidden state sequence and, at the same time, allows to examine state-specific covariate effects in the observable process, providing a useful modeling framework to capture unobserved time-dependent heterogeneity. Typically, a linear model targeting the conditional mean of the dependent variable given covariates is specified. The assumptions underlying traditional linear regression models, however, are seldom satisfied in real data which often exhibit skewness, heavy tails and outliers. Moreover, the effect of the covariates can differ greatly between different parts of the response distribution. Therefore, when the aim of the research focuses not only at the center of the response distribution but also, and especially in the tails, quantile regression (Koenker and Bassett 1978) represents an interesting alternative to standard mean regression. This method provides a way to model the conditional quantiles of a response variable with respect to the covariates in order to have a more complete picture of the entire conditional distribution compared to ordinary least squares. In the univariate quantile regression framework, both the classical and Bayesian inferential approaches have been proposed in the literature to estimate the model parameters. In the frequentist setting, the inferential approach relies on the minimization of the asymmetric loss function (see Koenker and Bassett 1978) while, in the Bayesian setting and in a likelihood inferential approach, the Asymmetric Laplace (AL) distribution has been introduced as a likelihood inferential tool. The two approaches are well-justified by the relationship between the quantile loss function and the AL density. Indeed, Yu and Moyeed (2001) showed that the minimization of the quantile loss function is equivalent, in terms of parameter estimates, to the maximization of the likelihood associated with the AL density. For a detailed review and list of references, Koenker (2005); Luo et al. (2012); Bernardi et al. (2015) and Koenker et al. (2017) provide an overview of the most used quantile regression techniques in both the classical and Bayesian settings. In addition, quantile regression methods have also been generalized to account for serial heterogeneity. In the analysis of longitudinal data, Farcomeni (2012) and Marino et al. (2018) consider univariate linear quantile models where unobserved sources of time-varying heterogeneity are captured by means of state-dependent coefficients evolving according to a finite-state homogeneous hidden Markov chain. Further, Ye et al. (2016) and Maruotti et al. (2021) propose a (semi-)Markov quantile regression to model the regime-switching effect of the regression coefficients in financial and environmental time series.

When multivariate response variables are concerned, the existing literature on quantile regression is less extensive since there is no "natural" ordering in a p-dimensional space, for $p > 1$. As a consequence, the univariate quantile regression method does not straightforwardly extend to higher dimensions. Nevertheless, in most situations of practical interest, the purpose of the matter being investigated lies in describing the distribution of a multivariate response variable.

For this reason the search for a satisfactory notion of multivariate quantile has led to a flourishing literature on this topic despite its definition is still a debatable issue (see Serfling 2002; Kong and Mizera 2012; Koenker et al. 2017; Stolfi et al. 2018; Chavas 2018; Charlier et al. 2020; Merlo et al. 2021, 2022 and the references therein for relevant studies).

Recently, Petrella and Raponi (2019) generalized the AL distribution inferential approach of the univariate quantile regression to a multivariate framework by using the Multivariate Asymmetric Laplace (MAL) distribution defined in Kotz et al. (2012). Employing the MAL distribution as a likelihood based inferential tool, the authors sidestep the problem of defining the quantiles of a multivariate distribution, and instead implement joint estimation for the univariate quantiles of the conditional distribution of a multivariate response variable given covariates, accounting for possible correlation among the responses.

The purpose of this article is to extend the work of Petrella and Raponi (2019) by introducing a HSMM for the analysis of multivariate time series. More formally, we develop a Quantile Hidden Semi-Markov Model (QHSMM) to jointly estimate the quantiles of the univariate conditional distributions of a multivariate response, accounting for the dependence structure between the outcomes. In particular, to capture the temporal evolution of unobserved heterogeneity, we introduce state-dependent coefficients in the regression model that evolve over time according to a latent semi-Markov process. In order to prevent inconsistent parameter estimates due to misspecification of the SDs, we adopt the nonparametric approach of Guédon (2003) where they are left unspecified and approximated by discrete distributions concentrated on a finite set of time points estimated from the data. Within this scheme, our modeling framework can be thought of as a model-based clustering approach for data

showing time-varying heterogeneity, where the interest lies in the effect of cluster-specific covariates on various quantile levels.

Throughout the paper we propose to estimate the model parameters with a Maximum Likelihood (ML) approach by using the MAL distribution as working likelihood in a regression framework. Specifically, as in Petrella and Raponi (2019) and Merlo et al. (2022), we consider the mixture representation of the MAL distribution which allows us to build an efficient Expectation-Maximization (EM) algorithm with the E- and M-step updates in closed form for all model parameters.

Using simulation experiments, we illustrate the validity of our approach under different data generating processes and evaluate its ability in recovering the true values of the regression coefficients, the true classification and number of latent states.

In the empirical analysis, we apply the proposed methodology to investigate the effect of a collection of atmospheric variables on the daily concentrations of three major pollutants, i.e., particulate matter, ozone and nitrogen dioxide measured in Rieti (Italy) from 2019 to 2021. Our method allows us to: (i) assess how the effects of atmospheric variables can vary across different (more extreme) quantiles of the conditional distribution of air pollutants, accounting for their dependence structure; (ii) summarize the data by means of a reduced number of latent regimes associated with different concentration levels of chemicals.

The paper is organized as follows. In Sect. 2, we introduce the proposed model. Sect. 3 illustrates the EM-based ML approach to estimating the model parameters and the computational details of the algorithm. In Sect. 4 we present the simulation results, while Sect. 5 discusses the empirical application. Finally, Sect. 6 concludes.

2 Methodology

Let $\{S_t\}_{t=1}^T$ denote a finite-state hidden semi-Markov chain defined over a discrete state space $\mathcal{S} = \{1, \ldots, K\}$. The latent process $\{S_t\}_{t=1}^T$ is constructed as follows. A homogeneous hidden Markov chain with K states models the transitions between different states, with initial probabilities, $\pi_k = \Pr(S_1 = k)$, and transition probabilities

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

with $\sum_{k=1}^K \pi_{jk} = 1$, $\pi_{jk} \geq 0$, for every $k = 1, \ldots, K$ and $\pi_{jj} = 0$, i.e., the diagonal elements of the transition probability matrix are zeros. More concisely, we collect the initial and transition probabilities in the K-dimensional vector $\boldsymbol{\pi}$ and in the $K \times K$ matrix $\mathbf{Q}$, respectively. Because the unobserved process is semi-Markovian, only transitions from one state to another are governed by the transition probabilities in (1), but the duration of a stay in a state is modeled by a separate SD. Specifically, let us denote by $d_k(u)$ the SD, i.e., the probability the hidden process $\{S_t\}_{t=1}^T$ spends u consecutive time steps in the k-th state, as follows:

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

where U_k corresponds to the maximum sojourn time of the hidden chain in state k. Let us also denote $\mathbf{U} = (U_1, \ldots, U_K)$ the K-dimensional vector collecting all state-specific maximum sojourn times.

HSMMs allow for great flexibility as the SD in (2) is directly specified by the researcher and estimated from the observed data. The SD can be chosen from a large variety of parametric distributions, such as the shifted-Poisson, the shifted-negative binomial distributions or, in the particular case where $d_k(u)$ is assumed to be geometrically distributed, a HSMM reduces to a HMM with the most likely sojourn time for every state being 1 (Zucchini et al. 2016). Parametric distributions, however, might lack the flexibility to capture key features of empirical SDs in the data, increasing state misclassification rates and inducing substantial bias. Alternatively, semi- and nonparametric data-driven approaches can be adopted (see Sansom and Thomson 2001; Guédon 2003; Langrock and Zucchini 2011; Maruotti et al. 2021) to provide sufficient additional flexibility in comparison to HMMs and accommodate complex distributional shapes.

To build the proposed model, let $\mathbf{Y}_t = (Y_t^{(1)}, \ldots, Y_t^{(p)})'$ be a continuous observable p-variate response variable and $\mathbf{X}_t = (1, X_t^{(2)}, \ldots, X_t^{(m)})'$ be a m-dimensional vector of covariates, with the first element being the intercept, at time $t = 1, \ldots, T$. The process $\{\mathbf{Y}_t\}_{t=1}^T$ represents the state-dependent process of the HSMM and, conditional on the hidden states, fulfills the independence property:

$$f_{\mathbf{Y}}(\mathbf{y}_t \mid \mathbf{x}_t, \mathbf{y}_1, \ldots, \mathbf{y}_{t-1}, S_1 = s_1, \ldots, S_t = s_t) = f_{\mathbf{Y}}(\mathbf{y}_t \mid \mathbf{x}_t, S_t = s_t), \tag{3}$$

where $f_{\mathbf{Y}}(\mathbf{y}_t \mid \mathbf{x}_t, S_t = s_t)$ is the conditional distribution of $\mathbf{Y}_t$ given the covariates $\mathbf{X}_t$ and the hidden state occupied at time t.

As mentioned in Section 1, our objective is to provide joint estimation of the p quantiles of the univariate conditional distributions of $\mathbf{Y}_t$, taking into account time-dependent heterogeneity and potential correlation among the components of $\mathbf{Y}_t$. Given p quantile indexes $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)$, with $\tau_j \in (0, 1)$, $j = 1, \ldots, p$, the Quantile Hidden Semi-Markov Model (QHSMM) is defined as follows:

$$\mathbf{Y}_t = \boldsymbol{\beta}_k(\boldsymbol{\tau})\mathbf{X}_t + \epsilon_{tk}(\boldsymbol{\tau}),$$

$$t = 1, \ldots, T \quad \text{and} \quad k = 1, \ldots, K, \tag{4}$$

with $\boldsymbol{\beta}_k(\boldsymbol{\tau}) = (\boldsymbol{\beta}_k^{(1)}(\boldsymbol{\tau}), \ldots, \boldsymbol{\beta}_k^{(p)}(\boldsymbol{\tau}))$ being a state-specific $p \times m$ matrix of unknown regression coefficients that evolves over time according to the hidden process S_t and takes one of the values in the set $\{\boldsymbol{\beta}_1(\boldsymbol{\tau}), \ldots, \boldsymbol{\beta}_K(\boldsymbol{\tau})\}$, and where $\epsilon_{tk}(\boldsymbol{\tau})$ denotes a p-dimensional vector of error terms with univariate component-wise quantiles (at fixed levels $\tau_1, \ldots, \tau_p$, respectively) equal to zero.

Generalizing the approach of Petrella and Raponi (2019), as conditional distribution of $\mathbf{Y}_t$ we consider a Multivariate Asymmetric Laplace (MAL) distribution (see Kotz et al. 2012). In detail, based on (4) we assume $\mathbf{Y}_t \mid \mathbf{X}_t = \mathbf{x}_t, S_t = k \sim \mathcal{MAL}_p(\boldsymbol{\mu}_{tk}, \mathbf{D}_k\tilde{\boldsymbol{\xi}}, \mathbf{D}_k\boldsymbol{\Sigma}_k\mathbf{D}_k)$ whose probability density function is given by:

$$f_{\mathbf{Y}}(\mathbf{y}_t \mid \mathbf{x}_t, S_t = k) = \frac{2\exp\left\{(\mathbf{y}_t - \boldsymbol{\mu}_{tk})'\mathbf{D}_k^{-1}\boldsymbol{\Sigma}_k^{-1}\tilde{\boldsymbol{\xi}}\right\}}{(2\pi)^{p/2}|\mathbf{D}_k\boldsymbol{\Sigma}_k\mathbf{D}_k|^{1/2}} \left(\frac{\tilde{m}_{tk}}{2+\tilde{d}_k}\right)^{\nu/2} K_\nu\left(\sqrt{(2+\tilde{d}_k)\tilde{m}_{tk}}\right), \tag{5}$$

where, for each time occasion $t = 1, \ldots, T$ and state $k = 1, \ldots, K$, the location parameter $\boldsymbol{\mu}_{tk}$ is defined by the linear model:

$$\boldsymbol{\mu}_{tk} = \boldsymbol{\mu}(S_t = k, \boldsymbol{\tau}) = \boldsymbol{\beta}_k(\tau)\mathbf{X}_t, \tag{6}$$

with $\mathbf{D_k}\tilde{\boldsymbol{\xi}}$ being the skewness parameter, $\mathbf{D}_k = \text{diag}[\delta_{k1}, \ldots, \delta_{kp}]$, $\delta_{kj} > 0$, and $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_1, \ldots, \tilde{\xi}_p)'$ having generic element $\tilde{\xi}_j = \frac{1-2\tau_j}{\tau_j(1-\tau_j)}$, $j = 1, \ldots, p$. $\boldsymbol{\Sigma}_k$ is a $p \times p$ positive definite matrix such that $\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda}\boldsymbol{\Psi}_k\boldsymbol{\Lambda}$, with $\boldsymbol{\Psi}_k$ being a $p \times p$ state-specific correlation matrix and $\boldsymbol{\Lambda} = \text{diag}[\sigma_1, \ldots, \sigma_p]$, with $\sigma_j^2 = \frac{2}{\tau_j(1-\tau_j)}$, $j = 1, \ldots, p$. Moreover, $\tilde{m}_{tk} = (\mathbf{y}_t - \boldsymbol{\mu}_{tk})'(\mathbf{D}_k\boldsymbol{\Sigma}_k\mathbf{D}_k)^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_{tk})$, $\tilde{d}_k = \mathbf{Q}\boldsymbol{\Sigma}_k^{-1}\tilde{\boldsymbol{\xi}}$, and $K_\nu(\cdot)$ denotes the modified Bessel function of the third kind with index parameter $\nu = (2-p)/2$.

One of the key benefits of the MAL distribution is that, using (4) and (5), and following Kotz et al. (2012), the $\mathcal{MAL}_p(\boldsymbol{\mu}, \mathbf{D}\tilde{\boldsymbol{\xi}}, \mathbf{D}\boldsymbol{\Sigma}\mathbf{D})$ can be written as a location-scale mixture with the following representation:

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{D}\tilde{\boldsymbol{\xi}}\tilde{C} + \sqrt{\tilde{C}}\mathbf{D}\mathbf{6}^{1/2}\mathbf{Z} \tag{7}$$

where $\mathbf{Z} \sim \mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$ denotes a p-variate standard Normal distribution and $\tilde{C} \sim \text{Exp}(1)$ has a standard exponential distribution, with $\mathbf{Z}$ being independent of $\tilde{C}$. In particular, the constraints imposed on $\tilde{\boldsymbol{\xi}}$ and $\boldsymbol{\Lambda}$ guarantee that the j-th element of $\boldsymbol{\mu}_{tk}$, μ_{tkj}, is the τ_j-th conditional quantile function of $Y_t^{(j)}$ given $S_t = k$, for $k = 1, \ldots, K$ and $j = 1, \ldots, p$, and represent necessary conditions for model identifiability for any fixed quantile level $\tau_1, \ldots, \tau_p$, as stated in the next proposition.

Proposition 1 *Let $\mathbf{Y}_t \mid \mathbf{X}_t = \mathbf{x}_t, S_t = k \sim \mathcal{MAL}_p(\boldsymbol{\mu}_{tk}, \mathbf{D}_k\tilde{\boldsymbol{\xi}}, \mathbf{D}_k\boldsymbol{\Sigma}_k\mathbf{D}_k)$, for $k = 1, \ldots, K$, where K is a positive integer, $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_1, \ldots, \tilde{\xi}_p)'$ with $\tilde{\xi}_j = \frac{1-2\tau_j}{\tau_j(1-\tau_j)}$ being known for any fixed value of τ_j. Furthermore, $\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda}\boldsymbol{\Psi}_k\boldsymbol{\Lambda}$ is a $p \times p$ positive definite matrix with $\boldsymbol{\Psi}_k$ being an unknown $p \times p$ correlation matrix and $\boldsymbol{\Lambda} = diag[\sigma_1, \ldots, \sigma_p]$, with fixed element $\sigma_j^2 = \frac{2}{\tau_j(1-\tau_j)}$, $j = 1, \ldots, p$. Then, the model in (2)-(6) is identified.*

Proof See Proof of Proposition 1 in Appendix. □

In comparison with other methods in the literature, the proposed modeling framework includes the homogeneous joint quantile regression approach of Petrella and Raponi (2019) when $K = 1$ and it reduces to the univariate hidden semi-Markov-switching quantile regression of Maruotti et al. (2021) when $p = 1$. Naturally, when a geometric SD is assumed for all latent states, we call our methodology the Quantile Hidden Markov Model (QHMM).

3 Maximum likelihood estimation and inference

In this section we introduce a ML approach to making inference on model parameters. As is usually done in the literature in the presence of latent variables, we propose a suitable likelihood-based EM algorithm (Dempster et al. 1977). To hedge against possibly biased inference from incorrect parametric assumptions on the SDs, we estimate the sojourn probabilities nonparametrically following Guédon (2003). In addition, we show that both the E- and M-step updates of the algorithm can be obtained in closed form by exploiting the hierarchical representation of the MAL distribution in (7) under the constraints on $\mathbf{Q}$ and $\boldsymbol{\Lambda}$, hence reducing the computational burden compared to direct maximization of the likelihood. We illustrate the EM algorithm to fit the more general QHSMM but it can also be employed for the QHMM by assuming a geometric SD. To ease the notation, unless specified otherwise, we omit the quantile levels vector $\boldsymbol{\tau}$, yet all model parameters are allowed to depend on it. All the proofs are collected in the Appendix.

Let us denote by $\boldsymbol{\Phi}_{\boldsymbol{\tau}} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \mathbf{D}_1, \ldots, \mathbf{D}_K, \boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_K, \boldsymbol{\pi}, \mathbf{Q}, d_1(u), \ldots, d_K(U_K))$ the set of model parameters. For any fixed $\boldsymbol{\tau}$, number of hidden states K and maximum sojourn times $\mathbf{U}$, we use the MAL representation

in (7) to express the complete-data likelihood as follows:

$$L_c(\boldsymbol{\Phi}_{\boldsymbol{\tau}}) = \pi_{s_1^\star} d_{s_1^\star}(u_1) \left\{ \prod_{r=2}^{R-1} \pi_{s_r^\star | s_{r-1}^\star} d_{s_r^\star}(u_r) \right\} \pi_{s_R^\star | s_{R-1}^\star} D_{s_R^\star}(u_R) \prod_{t=1}^{T} f_{\mathbf{Y}}(\mathbf{y}_t \mid \mathbf{x}_t, s_t, \tilde{c}_t, \tau) f_{\tilde{C}}(\tilde{c}_t), \tag{8}$$

where $\tilde{C}$ is a latent variable that follows an exponential distribution with parameter 1, $s_r^\star$ is the r-th visited state, u_r is the time spent in that state (i.e., the duration of the r-th visit) and $R-1$ is the number of state changes up to time T. Following Guédon (2003), the survivor function $D_k(u)$ for the sojourn time in state k is defined as:

$$D_k(u) = \sum_{v \geq u} d_k(v). \tag{9}$$

The survivor function sums up the individual probability masses of all possible sojourns of length $v \geq u$ and it has several advantages. Firstly, we do not have to assume that the process is leaving a state immediately after the upper endpoint T. Secondly, it provides a more accurate prediction of the last state visited, which is important when the data analysis wishes to estimate the most recently visited state, and improves parameter estimation (O'Connell and Højsgaard 2011).

3.1 The EM algorithm

The EM algorithm alternates between performing an expectation (E) step, which defines the expectation of the complete log-likelihood function evaluated using the current estimates of the parameters, and a maximization (M) step, which computes parameter estimates by maximizing the expected complete log-likelihood obtained in the E-step. The expected complete log-likelihood function and the optimal parameter updates are given in the following propositions.

Given the representation in (8), for the implementation of the algorithm we introduce the following quantities. We define the probability of being in state k at time t given the observed sequence as:

$$\gamma_{tk} = \Pr(S_t = k \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_T). \tag{10}$$

The probability the process left state j at time $t-1$ and entered state k at time t given the observed sequence is:

$$v_{tjk} = \Pr(S_{t-1} = j, S_t = k \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_T). \tag{11}$$

Finally, let us denote by $\eta_k(u)$ the expected number of times the process spends u consecutive time steps in state k as:

$$\begin{aligned} \eta_k(u) = \Pr(&S_{1+u} \neq k, S_{1+u-v} = k, \\ &v = 1, \ldots, u \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_T) \\ &+ \sum_{t=2}^{T} \Pr(S_{t+u} \neq k, S_{1+u-v} = k, v = 1, \ldots, u, \\ &S_{t-1} \neq k \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_T). \end{aligned} \tag{12}$$

Then, the expected log-likelihood for the complete data is presented in the following proposition.

Proposition 2 *For any fixed* $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)$*, number of hidden states* K *and maximum sojourn times* $\mathbf{U} = (U_1, \ldots, U_K)$*, the expected complete log-likelihood function (up to additive constants) is:*

$$\begin{aligned} \mathcal{O}(\boldsymbol{\Phi}_{\boldsymbol{\tau}}) = &\sum_{k=1}^{K} \gamma_{1k} \log \pi_k + \sum_{t=1}^{T} \sum_{j=1}^{K} \sum_{k \neq j}^{K} v_{tjk} \log \pi_{jk} + \sum_{k=1}^{K} \sum_{u=1}^{U_k} \eta_k(u) \log d_k(u) - \frac{1}{2} T \sum_{k=1}^{K} \log \mid \mathbf{D}_k \boldsymbol{\Sigma}_k \mathbf{D}_k \mid \\ &+ \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_{tk} \tilde{z}_{tk} (\mathbf{Y}_t - \boldsymbol{\mu}_{tk})' \mathbf{D}_k^{-1} \boldsymbol{\Sigma}_k^{-1} \tilde{\boldsymbol{\xi}} - \frac{1}{2} \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_{tk} \tilde{z}_{tk} (\mathbf{Y}_t - \boldsymbol{\mu}_{tk})' (\mathbf{D}_k \boldsymbol{\Sigma}_k \mathbf{D}_k)^{-1} (\mathbf{Y}_t - \boldsymbol{\mu}_{tk}) - \frac{1}{2} \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_{tk} \tilde{c}_{tk} \tilde{\boldsymbol{\xi}}' \boldsymbol{\Sigma}_k^{-1} \tilde{\boldsymbol{\xi}}, \end{aligned} \tag{13}$$

where

$$\tilde{c}_{tk} = \left(\frac{\tilde{m}_{tk}}{2 + \tilde{d}_k} \right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2 + \tilde{d}_k)\tilde{m}_{tk}} \right)}{K_{\nu}\left(\sqrt{(2 + \tilde{d}_k)\tilde{m}_{tk}} \right)},$$

$$\tilde{z}_{tk} = \left(\frac{2 + \tilde{d}_k}{\tilde{m}_{tk}} \right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2 + \tilde{d}_k)\tilde{m}_{tk}} \right)}{K_{\nu}\left(\sqrt{(2 + \tilde{d}_k)\tilde{m}_{tk}} \right)} - \frac{2\nu}{\tilde{m}_{tk}}, \tag{14}$$

with

$$\tilde{m}_{tk} = (\mathbf{y}_t - \boldsymbol{\mu}_{tk})' (\mathbf{D}_k \boldsymbol{\Sigma}_k \mathbf{D}_k)^{-1} (\mathbf{y}_t - \boldsymbol{\mu}_{tk}), \; \tilde{d}_k = \tilde{\boldsymbol{\xi}}' \boldsymbol{\Sigma}_k^{-1} \tilde{\boldsymbol{\xi}}. \tag{15}$$

Therefore, the EM algorithm can be implemented as follows:

E-step: At the generic r-th iteration of the algorithm, let $\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(r-1)}$ denote the current parameter estimates. Then, conditionally on the observed data and $\hat{\boldsymbol{\Phi}}_{\tau}^{(r-1)}$, the quantities, γ_{tk} in (10) and v_{tjk} in (11), can be calculated via a dynamic programming method known as the forward-backward algorithm (see, e.g., Levinson et al. 1983), while $\eta_k(u)$ in (12)

can be computed using the efficient adaptation of the forward-backward algorithm provided by Guédon (2003). Similarly, the conditional expectations $\tilde{c}_{tk}$ and $\tilde{z}_{tk}$ in (14) are considered; see the Appendix. We denote such quantities as $\hat{\gamma}_{tk}^{(r)}$, $\hat{v}_{tjk}^{(r)}$, $\hat{\eta}_k^{(r)}(u)$, $\hat{\tilde{c}}_{tk}^{(r)}$ and $\hat{\tilde{z}}_{tk}^{(r)}$.

M-step: Substitute them in (13) to maximize $\mathcal{O}(\boldsymbol{\Phi}_{\boldsymbol{\tau}} \mid \hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(r-1)})$ with respect to $\boldsymbol{\Phi}_{\boldsymbol{\tau}}$, and obtain the updated parameter estimates. Because the expected complete-data log-likelihood in (13) decomposes into orthogonal subproblems, the maximization with respect to the regression coefficients, the parameters of the MAL distribution and the hidden process, can be performed separately. Thus, the initial probabilities π_j and transition probabilities π_{jk} are estimated by:

$$\hat{\pi}_j^{(r)} = \hat{\gamma}_{1j}^{(r)} \quad \text{and} \quad \hat{\pi}_{jk}^{(r)} = \frac{\sum_{t=1}^{T} \hat{v}_{tjk}^{(r)}}{\sum_{t=1}^{T} \sum_{j \neq k}^{K} \hat{v}_{tjk}^{(r)}}. \tag{16}$$

To update the state-specific SD, we follow the nonparametric approach of Guédon (2003). In particular, we set the latent state duration densities to be discrete nonparametric distributions with arbitrary point mass assigned to the feasible duration values, that is, the SD is estimated as follows:

$$\hat{d}_k^{(r)}(u) = \frac{\hat{\eta}_k^{(r)}(u)}{\sum_{v=1}^{U_k} \hat{\eta}_k^{(r)}(v)}. \tag{17}$$

Finally, the M-step updates of the parameters in the regression equation $\{\boldsymbol{\beta}_j, \mathbf{D}_j, \boldsymbol{\Sigma}_j\}_{j=1}^{K}$, are given in the following proposition.

Proposition 3 *At the generic r-th iteration, the values of* $\boldsymbol{\beta}_j, \mathbf{D}_j$ *and* $\boldsymbol{\Sigma}_j$ *maximizing* (13) *are:*

$$\hat{\boldsymbol{\beta}}_j^{\prime(r)} = \Big(\sum_{t=1}^{T} \hat{\gamma}_{tj}^{(r)} \hat{\tilde{z}}_{tj}^{(r)} \mathbf{X}_t \mathbf{X}_t^{\prime}\Big)^{-1} \left(\sum_{t=1}^{T} \hat{\gamma}_{tj}^{(r)} \hat{\tilde{z}}_{tj}^{(r)} \mathbf{X}_t \mathbf{Y}_t^{\prime} - \sum_{t=1}^{T} \hat{\gamma}_{tj}^{(r)} \mathbf{X}_{it} \tilde{\boldsymbol{\xi}}^{\prime} \hat{\mathbf{D}}_j^{(r-1)}\right). \tag{18}$$

where $\rho_\tau(\cdot)$ *is the quantile check function of Koenker and Bassett (1978):*

$$\rho_\tau(u) = u(\tau - \mathbf{1}(u < 0)), \tag{21}$$

with $\mathbf{1}(\cdot)$ *being the indicator function and* $\hat{\mu}_{tjk}^{(r)}$ *being the k-th element of the vector* $\hat{\boldsymbol{\mu}}_{tj}^{(r)}$.

The E- and M-steps are alternated until convergence, that is when the observed likelihood between two consecutive iterations is smaller than a predetermined threshold. In this paper, we set this threshold criterion equal to 10^{-5}.

Following Maruotti et al. (2021), for fixed $\boldsymbol{\tau}$, K and $\mathbf{U}$, we initialize the EM algorithm by providing the initial states partition, $\{S_t^{(0)}\}_{t=1}^{T}$, according to a Multinomial distribution with probabilities $1/K$. From the generated partition, the off-diagonal elements of $\hat{\mathbf{Q}}^{(0)}$ are computed as proportions of transition. We obtain $\hat{\boldsymbol{\beta}}_k^{(0)}$ and $\hat{\mathbf{D}}_k^{(0)}$ by fitting univariate quantile regressions on observations within state k, while $\hat{\boldsymbol{\Psi}}_k^{(0)}$ is set equal to the empirical correlation computed on observations in the k-th state. The initial SDs are estimated from $\{S_t^{(0)}\}_{t=1}^{T}$ assuming a geometric distribution as in HMMs. To avoid convergence to local maxima and better explore the parameter space, we fit the proposed QHSMM using a multiple random starts strategy with different starting partitions and retain the solution corresponding to the maximum likelihood value.

Once we computed the ML estimates of the model parameters $\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}$, we calculate standard errors using a parametric bootstrap approach (Visser et al. 2000). That is, we refitted the model to H bootstrap samples and approximate the standard error of each model parameter with its corresponding standard deviation computed on bootstrap samples. Hence, standard error estimates for $\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}$ are given by the diagonal elements of:

$$\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}) = \sqrt{\frac{1}{H-1} \sum_{h=1}^{H} (\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(h)} - \bar{\boldsymbol{\Phi}}_{\boldsymbol{\tau}})(\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(h)} - \bar{\boldsymbol{\Phi}}_{\boldsymbol{\tau}})^{\prime}}, \tag{22}$$

$$\hat{\boldsymbol{\Sigma}}_j^{(r)} = \frac{1}{T} \sum_{t=1}^{T} \hat{\gamma}_{tj}^{(r)} \hat{\tilde{z}}_{tj}^{(r)} \hat{\mathbf{D}}_j^{-1(r-1)} (\mathbf{Y}_t - \hat{\boldsymbol{\mu}}_{tj}^{(r)})(\mathbf{Y}_t - \hat{\boldsymbol{\mu}}_{tj}^{(r)})^{\prime} \hat{\mathbf{D}}_j^{-1(r-1)} + \frac{1}{T} \sum_{t=1}^{T} \hat{\gamma}_{tj}^{(r)} \hat{\tilde{c}}_{tj}^{(r)} \tilde{\boldsymbol{\xi}} \tilde{\boldsymbol{\xi}}^{\prime} - \frac{2}{T} \hat{\mathbf{D}}_j^{-1(r-1)} \sum_{t=1}^{T} \hat{\gamma}_{tj}^{(r)} (\mathbf{Y}_t - \hat{\boldsymbol{\mu}}_{tj}^{(r)}) \tilde{\boldsymbol{\xi}}^{\prime}, \tag{19}$$

where $\hat{\boldsymbol{\mu}}_{tj}^{(r)} = \hat{\boldsymbol{\beta}}_j^{(r)} \mathbf{X}_t$.

For the j-th state, the elements $\delta_{jk}, k = 1, \ldots, p$, *of the diagonal scale matrix* $\mathbf{D}_j$ *are estimated by:*

$$\hat{\delta}_{jk}^{(r)} = \frac{1}{T} \sum_{t=1}^{T} \hat{\gamma}_{tj}^{(r)} \rho_\tau(Y_t^{(k)} - \hat{\mu}_{tjk}^{(r)}), \tag{20}$$

where $\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(h)}$ is the set of parameter estimates for the h-th bootstrap sample and $\bar{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}$ denote the sample mean of all $\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(h)}, h = 1, \ldots, H$.

3.2 Model selection

In the EM algorithm discussed above, the number of hidden states K and maximum length of state durations $\mathbf{U} = (U_1, \ldots, U_K)$ are unknown. From an applied perspective, choosing an adequate number of states is a crucial aspect of the data analysis which shall take into account the data structure and research question at hand. In particular, K is typically selected using penalized-likelihood criteria or cross-validation methods, which can become demanding to fit computationally (Pohle et al. 2017). In HSMMs, not only the number of hidden states shall be selected, but also the maximum length of state durations. In practice, this amounts to fixing $U_k = U$, $k = 1, \ldots, K$, with U being large enough to capture the main support of the SD in each state (see Maruotti et al. 2021). The major disadvantages of this approach are the large number of parameters to be estimated and the fact that different states may require substantially different maximum sojourn times. For these reasons, herein we simultaneously select the optimal values of K and vector $\mathbf{U}$ using penalized likelihood criteria, such as AIC (Akaike 1998), BIC (Schwarz 1978) and ICL (Biernacki et al. 2000) which penalizes the BIC for the estimated mean entropy and it is given by:

$$ICL_{(K,\mathbf{U})} = BIC_{(K,\mathbf{U})} - 2\sum_{t=1}^{T}\sum_{k=1}^{K}\widehat{\gamma}_{tk}\log\widehat{\gamma}_{tk}, \tag{23}$$

where $BIC_{(K,\mathbf{U})}$ in (23) is defined as $BIC_{(K,\mathbf{U})} = -2\ell(\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}) + \log(T)\nu_f$, with $\ell(\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}) = \log\sum_{s_1,\ldots,s_T}\sum_u L_c(\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}})$ being the observed data log-likelihood in correspondence of the ML estimate of $\boldsymbol{\Phi}_{\boldsymbol{\tau}}$, T corresponds to the number of observations and ν_f denotes the number of free model parameters in $\boldsymbol{\Phi}_{\boldsymbol{\tau}}$. Computing $\ell(\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}})$ may prove to be difficult to evaluate directly because it involves the sum on every possible state sequence of length T, $\sum_{s_1,\ldots,s_T}$, and the sum on every supplementary duration from time $T+1$ spent in the state occupied at time T, $\sum_u$. Therefore, to compute $\ell(\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}})$ we use the variables in (10)-(12) required for the EM algorithm (please see Guédon 2003).

All criteria involve penalization terms depending on the number of parameters ν_f, which is given by the sum of:

- the number of regression parameters in $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K\}$: $p \times m \times K$,
- the number of scale parameters in $\mathbf{D}_k$: $p \times K$,
- the number of correlation parameters in $\boldsymbol{\Psi}_k$: $p \times K(K-1)/2$,
- the number of independent transition probabilities in $\mathbf{Q}$: $K \times (K-2)$,
- the unconstrained sojourn distribution probabilities: $\sum_{k=1}^{K}(U_k - 1)$.

To select the order of the hidden process, we first define a sequence of values of K and construct a K-dimensional grid of maximum sojourn distributions, $\mathcal{U} \subset \mathbb{R}^K_{\geq 0}$, and then fit the model using the EM algorithm described above for fixed $\boldsymbol{\tau}$, K and a vector $\mathbf{U}$ in $\mathcal{U}$. Because a full search over $\mathcal{U}$ might be computationally infeasible, we employ the greedy search algorithm considered in Langrock et al. (2015) and Adam et al. (2019) and select the best combination of $(K, \mathbf{U})$ corresponding to the lowest value of the penalized likelihood criteria.

4 Simulation study

We conduct a simulation study to evaluate the finite sample properties of the proposed QHSMM. This simulation exercise addresses the following issues: (i) study the performance of the model under different distributional choices for the error term and SDs, when either a linear or nonlinear quantile regression function of $\mathbf{Y}$ given $\mathbf{X}$ is considered; (ii) assess the classification performance of the proposed model; (iii) evaluate the performance of penalized likelihood criteria in selecting the optimal number of hidden states K and maximum sojourn times $\mathbf{U}$. Additional simulation studies are illustrated in the Supplementary Materials.

We consider $T = 1000$, a continuous response variable of dimension $p = 2$ and one explanatory variable $X_t \sim \mathcal{N}(0, 1)$. The observations are generated from a two state HSMM, i.e., $K = 2$, using the following data generating process:

$$\mathbf{Y}_t = \boldsymbol{\beta}_k \mathbf{X}_t + \boldsymbol{\epsilon}_{tk}, \tag{24}$$

where $\mathbf{X}_t = (1, X_t)'$ and the true values of the state-dependent parameters, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, are given by:

$$\boldsymbol{\beta}_1 = \begin{pmatrix} 4 & 2 \\ -3 & -1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\beta}_2 = \begin{pmatrix} 5 & -2 \\ -4 & 1 \end{pmatrix}. \tag{25}$$

We consider the following two distributions for the error terms $\boldsymbol{\epsilon}_{tk}$ in (24):

- $(\mathcal{N})$: $\boldsymbol{\epsilon}_{tk}$ are generated from a bivariate Normal random variable with zero mean vector and variance-covariance matrix equal to $\tilde{\boldsymbol{\Omega}}_k$, for $k = 1, 2$;
- $(\mathcal{T})$: $\boldsymbol{\epsilon}_{tk}$ are generated from a bivariate Student t distribution with 5 degrees of freedom, zero mean and scale matrix equal to $\tilde{\boldsymbol{\Omega}}_k$, for $k = 1, 2$.

The state-specific covariance matrices $\tilde{\boldsymbol{\Omega}}_k$, $k = 1, 2$, are set equal to low ($\tilde{\boldsymbol{\Omega}}_1 = \left(\begin{smallmatrix} 1 & 0.3 \\ 0.3 & 1 \end{smallmatrix}\right)$) and high ($\tilde{\boldsymbol{\Omega}}_2 = \left(\begin{smallmatrix} 1 & 0.8 \\ 0.8 & 1 \end{smallmatrix}\right)$) correlation between the responses.

Table 1 ARB (in percentage) and RMSE (in brackets) for state-parameter estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ with normal errors for the QHSMM

True Coef.	$\boldsymbol{\tau}$				
	(0.10, 0.10)	(0.25, 0.25)	(0.50, 0.50)	(0.75, 0.75)	(0.90, 0.90)
Panel A: SPO					
−4	2.584 (0.081)	0.653 (0.062)	0.095 (0.058)	−0.452 (0.066)	−2.227 (0.090)
−3	3.576 (0.051)	0.447 (0.039)	−0.053 (0.035)	−0.549 (0.038)	−3.771 (0.053)
−2	0.243 (0.081)	0.149 (0.068)	0.150 (0.065)	0.040 (0.070)	0.214 (0.086)
−1	0.055 (0.064)	0.146 (0.052)	0.201 (0.050)	0.161 (0.051)	0.293 (0.065)
1	−0.588 (0.085)	−0.344 (0.068)	−0.271 (0.067)	−0.074 (0.071)	−0.344 (0.088)
2	0.047 (0.063)	−0.085 (0.051)	−0.152 (0.050)	−0.205 (0.051)	−0.265 (0.063)
4	−2.658 (0.052)	−0.342 (0.038)	0.023 (0.035)	0.374 (0.039)	2.674 (0.054)
5	−1.994 (0.081)	−0.487 (0.061)	−0.077 (0.058)	0.415 (0.062)	1.797 (0.085)
Panel B: SNB					
−4	2.851 (0.092)	0.738 (0.072)	0.100 (0.067)	−0.579 (0.077)	−2.414 (0.110)
−3	3.390 (0.050)	0.424 (0.036)	−0.021 (0.032)	−0.493 (0.035)	−3.689 (0.050)
−2	0.298 (0.097)	0.457 (0.079)	0.044 (0.073)	−0.179 (0.080)	−0.061 (0.098)
−1	−0.335 (0.061)	0.111 (0.048)	0.087 (0.047)	−0.190 (0.050)	−0.062 (0.061)
1	−0.457 (0.098)	−0.867 (0.078)	0.007 (0.071)	0.352 (0.078)	0.257 (0.096)
2	0.255 (0.061)	−0.043 (0.050)	−0.131 (0.047)	−0.010 (0.049)	0.073 (0.063)
4	−2.496 (0.050)	−0.274 (0.036)	0.047 (0.033)	0.323 (0.036)	2.545 (0.052)
5	−2.252 (0.095)	−0.578 (0.074)	−0.076 (0.070)	0.488 (0.074)	1.881 (0.103)
Panel C: GEO					
−4	2.771 (0.076)	0.774 (0.060)	0.027 (0.056)	−0.589 (0.065)	−2.306 (0.090)
−3	3.174 (0.062)	0.235 (0.044)	−0.037 (0.040)	−0.556 (0.044)	−3.769 (0.064)
−2	0.030 (0.083)	−0.033 (0.068)	−0.082 (0.063)	−0.068 (0.068)	0.090 (0.082)
−1	−1.221 (0.069)	−0.284 (0.053)	−0.134 (0.049)	−0.055 (0.053)	0.189 (0.067)
1	−0.292 (0.082)	−0.231 (0.067)	−0.154 (0.063)	−0.189 (0.069)	−0.249 (0.081)
2	0.841 (0.070)	0.264 (0.055)	−0.019 (0.051)	−0.015 (0.052)	−0.005 (0.066)
4	−2.420 (0.059)	−0.219 (0.044)	0.038 (0.039)	0.300 (0.044)	2.417 (0.061)
5	−2.137 (0.076)	−0.619 (0.060)	−0.041 (0.058)	0.474 (0.062)	1.779 (0.087)

Similarly to Maruotti et al. (2021), for each scenario we further consider three SDs:

(SPO) : a shifted-Poisson, i.e.:

$$d_k(u) = \exp(-\lambda_k)\frac{\lambda_k^{u-1}}{(u-1)!}, \qquad u = 1, 2, \ldots, \tag{26}$$

with $\lambda_1 = 10$ and $\lambda_2 = 5$;

(SNB) : a shifted-negative binomial, i.e.:

$$d_k(u) = \frac{\Gamma(u + \lambda_k - 1)}{(u-1)!\Gamma(\lambda_k)} p_k^{\lambda_k}(1-p_k)^{u-1}, \; u = 1, 2, \ldots, \tag{27}$$

with $\lambda_1 = 8$, $\lambda_2 = 4$, $p_1 = 0.5$ and $p_2 = 0.6$;

(GEO) : a geometric sojourn, i.e.:

$$d_k(u) = p_k(1-p_k)^{u-1}, \qquad u = 1, 2, \ldots, \tag{28}$$

with $p_1 = 0.2$ and $p_2 = 0.3$.

We fit the proposed QHSMM for five quantile levels, i.e., $\boldsymbol{\tau} = (0.10, 0.10)$, $\boldsymbol{\tau} = (0.25, 0.25)$, $\boldsymbol{\tau} = (0.50, 0.50)$, $\boldsymbol{\tau} = (0.75, 0.75)$ and $\boldsymbol{\tau} = (0.90, 0.90)$. For each model, we carry out $B = 1000$ Monte Carlo replications and report the following indicators. The Average Relative Bias (ARB), expressed as a percentage:

$$ARB(\hat{\theta}_{\boldsymbol{\tau}}) = \frac{1}{B}\sum_{b=1}^{B}\frac{(\hat{\theta}_{\boldsymbol{\tau}}^{(b)} - \theta_{\boldsymbol{\tau}})}{\theta_{\boldsymbol{\tau}}} \times 100, \tag{29}$$

where $\hat{\theta}_{\boldsymbol{\tau}}^{(b)}$ is the estimated parameter at level $\boldsymbol{\tau}$ for the b-th replication and $\theta_{\boldsymbol{\tau}}$ is the corresponding "true" value. Secondly, the Root Mean Square Error (RMSE) of model

Table 2 ARB (in percentage) and RMSE (in brackets) for state-parameter estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ with Student t errors for the QHSMM

	$\boldsymbol{\tau}$				
True Coef.	(0.10, 0.10)	(0.25, 0.25)	(0.50, 0.50)	(0.75, 0.75)	(0.90, 0.90)
Panel A: SPO					
−4	4.150 (0.132)	0.393 (0.068)	0.082 (0.059)	−0.112 (0.069)	−3.457 (0.140)
−3	3.213 (0.080)	−0.420 (0.044)	−0.099 (0.037)	0.200 (0.044)	−4.030 (0.081)
−2	0.178 (0.119)	−0.098 (0.077)	−0.071 (0.069)	0.098 (0.078)	0.323 (0.124)
−1	−0.470 (0.088)	0.065 (0.058)	−0.164 (0.050)	−0.221 (0.057)	−0.082 (0.089)
1	−0.610 (0.119)	−0.135 (0.077)	−0.273 (0.070)	−0.469 (0.079)	−0.740 (0.126)
2	0.613 (0.086)	0.258 (0.056)	0.257 (0.048)	0.250 (0.056)	0.231 (0.087)
4	−2.558 (0.075)	0.294 (0.042)	0.083 (0.036)	−0.190 (0.042)	2.500 (0.075)
5	−2.986 (0.132)	−0.236 (0.071)	−0.075 (0.060)	0.115 (0.069)	2.574 (0.137)
Panel B: SNB					
−4	5.410 (0.158)	0.579 (0.085)	−0.054 (0.072)	−0.577 (0.084)	−5.037 (0.199)
−3	2.516 (0.080)	−0.423 (0.042)	0.075 (0.034)	0.357 (0.041)	−3.445 (0.082)
−2	0.906 (0.135)	0.177 (0.086)	−0.101 (0.076)	−0.038 (0.087)	−0.171 (0.143)
−1	−0.945 (0.084)	−0.119 (0.056)	−0.168 (0.050)	−0.059 (0.055)	0.448 (0.085)
1	−1.643 (0.137)	−0.275 (0.086)	0.206 (0.075)	0.359 (0.085)	0.744 (0.141)
2	0.840 (0.085)	0.066 (0.056)	−0.127 (0.048)	−0.022 (0.054)	−0.047 (0.084)
4	−2.102 (0.075)	0.291 (0.042)	−0.037 (0.034)	−0.395 (0.040)	1.812 (0.076)
5	−3.836 (0.171)	−0.391 (0.087)	0.015 (0.072)	0.444 (0.083)	3.787 (0.186)
Panel C: GEO					
−4	4.668 (0.129)	0.717 (0.071)	−0.012 (0.059)	−0.421 (0.071)	−3.923 (0.167)
−3	2.407 (0.102)	−0.537 (0.052)	0.015 (0.043)	0.308 (0.050)	−3.759 (0.105)
−2	−0.068 (0.115)	−0.039 (0.075)	−0.130 (0.065)	−0.173 (0.074)	0.375 (0.118)
−1	−1.090 (0.092)	−0.022 (0.060)	−0.091 (0.052)	0.119 (0.059)	0.323 (0.098)
1	0.275 (0.118)	0.076 (0.076)	0.127 (0.064)	0.545 (0.075)	−0.229 (0.118)
2	1.057 (0.093)	0.229 (0.062)	−0.094 (0.053)	−0.068 (0.062)	0.034 (0.098)
4	−2.102 (0.092)	0.429 (0.050)	0.045 (0.042)	−0.368 (0.052)	2.027 (0.098)
5	−3.366 (0.135)	−0.429 (0.072)	−0.016 (0.061)	0.287 (0.073)	2.963 (0.158)

parameters averaged across the B simulations:

$$RMSE(\hat{\theta}_{\boldsymbol{\tau}}) = \sqrt{\frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}_{\boldsymbol{\tau}}^{(b)} - \theta_{\boldsymbol{\tau}})^2}. \tag{30}$$

To assess the first and second queries of this simulation exercise, Tables 1 and 2 report the ARB and RMSE for the state-specific coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. As can be noted, the proposed model under the Normal and Student t error distributions is able to recover the true state-dependent parameters for both low and high degree of dependence and all three considered SDs. Not surprisingly, the bias effect is quite small when we analyze the median levels (see column 3). As the quantile levels become more extreme (see columns 1, 2, 4 and 5), the ARB slightly increases but it still remains reasonably small. Also, under the ($\mathcal{T}$) scenario the heavier tails of the Student t contribute to higher ARB and RMSE especially at the 10-th and 90-th percentiles.

To evaluate the classification performance of the proposed model, we report the average Adjusted Rand Index (ARI) of Hubert and Arabie 1985 and the misclassification rate (MCR). Specifically, we compare the classification obtained by the QHSMM with the one obtained from a QHMM under the assumption of a geometric SD. The state partition provided by the fitted models is obtained by taking the maximum, $\max_{k\in\mathcal{S}} \gamma_{tk}$, posteriori probability for every $t = 1, \ldots, T$. The results in Table 3 show that when the true SD of the data generating process is geometric, the QHMM provides a slightly better classification both in terms of ARI and MCR (please compare the GEO row of Panel A with that of Panel C and the GEO row of Panel B with that of Panel D). This is not surprising as the QHMM implicitly assumes geometrically distributed sojourn distributions. The QHSMM, on the contrary, outperforms the QHMM in all other cases, as it can approximate arbitrarily well any SD and does not rely on a distributional assumption for $d_k(u)$ (compare the SPO

Table 3 Average and standard deviation (in brackets) values of the ARI and MCR for the QHSMM and QHMM under the three considered SDs and two distributions for the error term

SD	$\boldsymbol{\tau}$									
	(0.10, 0.10)		(0.25, 0.25)		(0.50, 0.50)		(0.75, 0.75)		(0.90, 0.90)	
	ARI	MCR	ARI	MCR	ARI	MCR	ARI	MCR	ARI	MCR
QHSMM										
Panel A: $\mathcal{N}$										
SPO	0.905	0.024	0.908	0.023	0.908	0.023	0.909	0.023	0.907	0.024
	(0.022)	(0.006)	(0.021)	(0.006)	(0.021)	(0.006)	(0.021)	(0.005)	(0.022)	(0.006)
SNB	0.852	0.037	0.857	0.036	0.855	0.036	0.856	0.036	0.853	0.037
	(0.026)	(0.007)	(0.026)	(0.007)	(0.026)	(0.007)	(0.026)	(0.007)	(0.026)	(0.007)
GEO	0.783	0.057	0.785	0.057	0.783	0.057	0.783	0.057	0.782	0.058
	(0.028)	(0.008)	(0.029)	(0.008)	(0.028)	(0.008)	(0.029)	(0.008)	(0.028)	(0.008)
Panel B: $\mathcal{T}$										
SPO	0.874	0.032	0.885	0.030	0.885	0.029	0.885	0.029	0.876	0.032
	(0.029)	(0.008)	(0.023)	(0.006)	(0.023)	(0.006)	(0.023)	(0.006)	(0.025)	(0.007)
SNB	0.810	0.048	0.824	0.045	0.824	0.045	0.824	0.045	0.811	0.048
	(0.030)	(0.008)	(0.028)	(0.008)	(0.028)	(0.008)	(0.029)	(0.008)	(0.031)	(0.008)
GEO	0.737	0.071	0.745	0.068	0.744	0.069	0.743	0.069	0.735	0.071
	(0.031)	(0.009)	(0.031)	(0.009)	(0.031)	(0.009)	(0.031)	(0.009)	(0.031)	(0.009)
QHMM										
Panel C: $\mathcal{N}$										
SPO	0.895	0.027	0.904	0.024	0.909	0.023	0.907	0.024	0.900	0.025
	(0.022)	(0.006)	(0.021)	(0.005)	(0.020)	(0.005)	(0.020)	(0.005)	(0.021)	(0.005)
SNB	0.850	0.038	0.856	0.036	0.861	0.035	0.857	0.036	0.851	0.037
	(0.026)	(0.007)	(0.025)	(0.007)	(0.025)	(0.006)	(0.025)	(0.007)	(0.026)	(0.007)
GEO	0.799	0.053	0.802	0.052	0.804	0.052	0.802	0.052	0.798	0.053
	(0.026)	(0.007)	(0.026)	(0.007)	(0.026)	(0.007)	(0.027)	(0.007)	(0.027)	(0.008)
Panel D: $\mathcal{T}$										
SPO	0.856	0.037	0.874	0.032	0.883	0.030	0.876	0.032	0.858	0.037
	(0.026)	(0.007)	(0.024)	(0.006)	(0.023)	(0.006)	(0.023)	(0.006)	(0.026)	(0.007)
SNB	0.803	0.050	0.820	0.046	0.830	0.043	0.822	0.045	0.806	0.049
	(0.029)	(0.008)	(0.027)	(0.007)	(0.026)	(0.007)	(0.027)	(0.007)	(0.029)	(0.008)
GEO	0.752	0.066	0.762	0.063	0.768	0.062	0.763	0.063	0.752	0.066
	(0.030)	(0.009)	(0.030)	(0.008)	(0.029)	(0.008)	(0.030)	(0.008)	(0.029)	(0.008)

and SNB rows of Panel A with those of Panel C, and the SPO and SNB rows of Panel B with those of Panel D), with very few exceptions at quantile levels $\boldsymbol{\tau} = (0.50, 0.50)$ and $\boldsymbol{\tau} = (0.75, 0.75)$ where the two models give comparable results.

We further evaluate the QHSMM introduced when a nonlinear quantile regression function of $\mathbf{Y}$ given $\mathbf{X}$ is considered. Similarly to Geraci (2019), the observations are generated from a two state HSMM using the following nonlinear quantile regression models. In the first scenario, we simulated the data from the following logistic model:

$$Y_t^{(j)} = \frac{\beta_{1,jk}}{1 + \exp((\beta_{2,jk} - X_t)/\beta_{3,jk})} + \epsilon_{tk}^{(j)}, \text{ for } j = 1, 2, \tag{31}$$

where the explanatory variable is drawn from a continuous uniform distribution, $X_t \sim \mathcal{U}(0, 20)$, and where $Y_t^{(j)}$ and $\epsilon_{tk}^{(j)}$ denote the j-th component of $\mathbf{Y}_t$ and $\boldsymbol{\epsilon}_{tk}$, respectively. For each component j and hidden state k, the true values of the parameters, $\boldsymbol{\beta}_{jk}$, are given by $\boldsymbol{\beta}_{11} = (50, 12, 3)$, $\boldsymbol{\beta}_{21} = (10, 2, -1)$, $\boldsymbol{\beta}_{12} = (30, 5, 2)$ and $\boldsymbol{\beta}_{22} = (20, 11, -3)$.

In the second scenario, following El Ghouch and Genton (2009) the data are generated according to the equation:

$$Y_t^{(j)} = \beta_{1,jk} + \beta_{2,jk} X_t^2 + \beta_{3,jk} X_t^3 + \gamma_j \exp(-4(X_t - 1)^2) + \epsilon_{tk}^{(j)}, \text{ for } j = 1, 2, \tag{32}$$

where $X_t \sim \mathcal{U}(-1.1, 2.1)$ and where the parameter γ_j in (32) can be seen as a misspecification parameter that con-

trols the deviation from the polynomial function. As γ_j increases, the data structure becomes more complicated, and approximating the true curve by a polynomial becomes increasingly difficult. In this study we chose $\gamma_1 = 4$ and $\gamma_2 = 8$. The true values of the parameters are given by $\boldsymbol{\beta}_{11} = (10, -6, 2.8)$, $\boldsymbol{\beta}_{21} = (3, 5, -0.5)$, $\boldsymbol{\beta}_{12} = (-3, 5, -2)$ and $\boldsymbol{\beta}_{22} = (8, 11, -3)$.

For the error terms $\boldsymbol{\epsilon}_{tk}$ in (31) and (32), and the SDs we considered the same distributions adopted for the linear case. Examples of the simulated data are shown in Figures 1 and 2 from the two scenarios, respectively.

We fit the proposed QHSMM for five quantile levels, i.e., $\boldsymbol{\tau} = (0.10, 0.10)$, $\boldsymbol{\tau} = (0.25, 0.25)$, $\boldsymbol{\tau} = (0.50, 0.50)$, $\boldsymbol{\tau} = (0.75, 0.75)$ and $\boldsymbol{\tau} = (0.90, 0.90)$. For each model, we report the Proportion of Negative Residuals (PNR):

$$PNR(\tau_j) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\left(Y_t^{(j)} < \widehat{Q}_{Y_t^{(j)}|X_t}\right), \text{ for } j = 1, 2, \tag{33}$$

with $\widehat{Q}_{Y_t^{(j)}|X_t}$ being the fitted conditional quantile of $Y_t^{(j)}$ at level τ_j, $j = 1, 2$. If the model is correctly specified, the PNR should be approximately equal to τ_j for each outcome. The results contained in Tables 4 and 5 are averaged over $B = 1000$ Monte Carlo replications.

By looking at the results, PNR rates are in general coherent with the selected quantile level. Further, one can observe that PNRs are typically closer to the nominal values $\boldsymbol{\tau}$ in the first scenario as opposed to the second one. This may be explained by the fact that the fitted model provides a better linear approximation of the true logistic quantile regression function than the more complex nonlinear model in the second scenario. Moreover, in this latter case, the PNRs obtained for the second component Y_2 do worse than the corresponding ones for Y_1, which is mainly due to the relatively larger variance associated to the misspecification parameter γ_j in (32). Overall, in both cases the PNRs are slightly, especially at the tails, below (at $\boldsymbol{\tau} = (0.10, 0.10)$) or above (at $\boldsymbol{\tau} = (0.90, 0.90)$) the expected proportions.

Finally, to assess the performance of penalized likelihood criteria (AIC, BIC and ICL) for selecting the number of hidden states and maximum sojourn times, we considered the same simulation experiment where the sojourns are generated from a beta-binomial distribution:

$$d_k(u) = \binom{U_k}{u} \frac{\mathcal{B}(u + \alpha_k, U - u + \lambda_k)}{\mathcal{B}(\alpha_k, \lambda_k)}, \; u = 1, 2, \ldots, U_k, \tag{34}$$

where $\mathcal{B}(\cdot, \cdot)$ is the beta function, $\alpha_1 = 6$, $\alpha_2 = 0.7$, $\lambda_1 = 3$, $\lambda_2 = 2$, $U_1 = 20$ and $U_2 = 10$. For each of the simulated $B = 100$ datasets, we fit the QHSMM with $K = 2, 3, 4$ over the grid of state-dependent maximum sojourn times $(10, 15, 20) \times \cdots \times (10, 15, 20)$, and select the best combination of $(K, \mathbf{U})$ associated to the lowest penalized likelihood criteria. Table 6 reports the number of times each criterion correctly identifies the number of latent states and maximum support points of the SDs (Panel A) and the absolute frequency distributions of the selected K for each of the three criteria (Panel B).

As one can see in Panel A, all three criteria work relatively well at $\boldsymbol{\tau} = (0.50, 0.50)$. By contrast, as we move towards the tails of the distribution of the responses, the ICL outperforms both the AIC and BIC and correctly identifies the pair $(K, \mathbf{U})$ that was used to generate the data more than 96% across all simulation scenarios. Moving onto Panel B, the AIC consistently performs worse than the BIC but both mostly overestimate the true number of states. These results suggest that regardless of the distribution on the error terms and SDs, the ICL yields superior performance and captures serial heterogeneity in the data in a more parsimonious manner compared to the other criteria, easing the interpretation of the latent states.

5 Application

In this section we apply the proposed methodology to air pollution data collected by the Lazio Regional Agency for Environmental Prevention and Protection (ARPA Lazio, https://www.arpalazio.it) in Italy. The ARPA Lazio provides information regarding the regional state of the environment and environmental trends, performing scientific, technical and research functions as well as assessment, monitoring, control and supporting local and health authorities. The time series used in this research are freely available from the ARPA Lazio website (https://www.arpalazio.net/main/aria/sci/basedati/chimici/chimici.php).

5.1 Data description

The data considered originate from a regional monitoring network system developed by the Lazio Region and have already been discussed in the work of Maruotti et al. (2017). This system has been organized in order to respond to the increasing demand for environmental information, but also for providing reliable data suitable for policy-related aspects. The network recorded concentrations of nine air pollutants on hourly basis at monitoring stations in the central area of the city of Rieti, Italy. The Rieti site has been chosen as it is classified as a traffic location, although it is located at a short distance from green areas and forests that facilitate the movement of air masses and removal of pollutants. In this work we consider three major air pollutants, i.e., Particulate Matter with aerodynamic diameter less than 2.5 μm (PM_{25}),

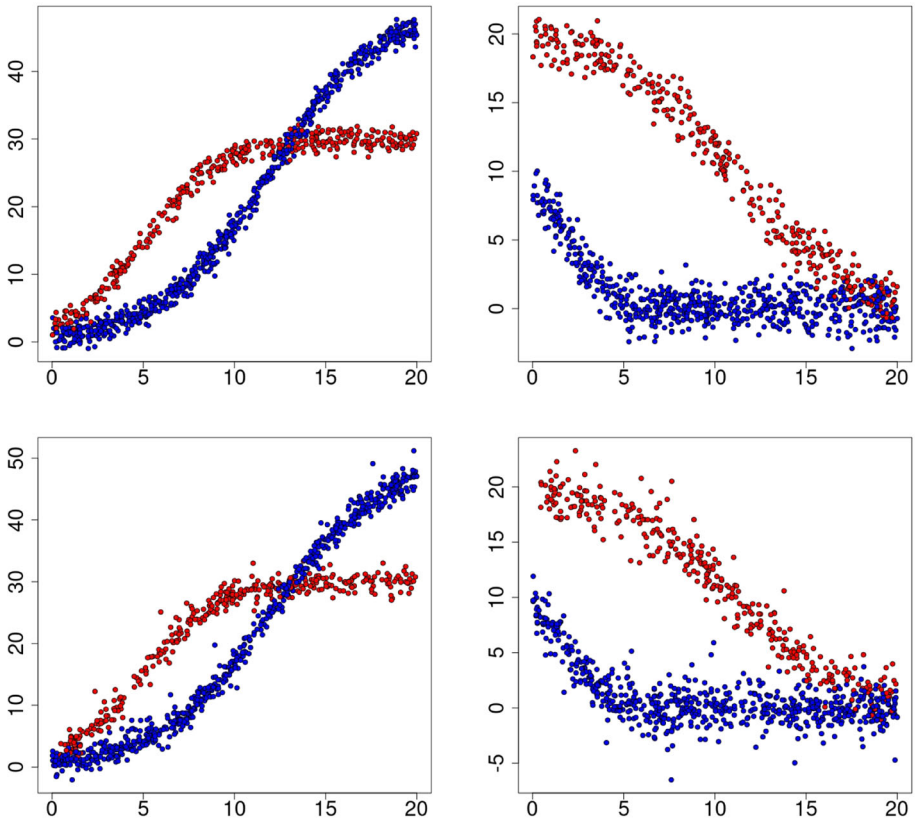

Fig. 1 Examples of data generated from the first scenario. Scatterplot of Y_1 (first column) and Y_2 (second column) under normal (first row) and Student t (second row) errors as a function of the included covariate, when using shifted-Poisson SDs. Red and blue data points distinguish the two latent states

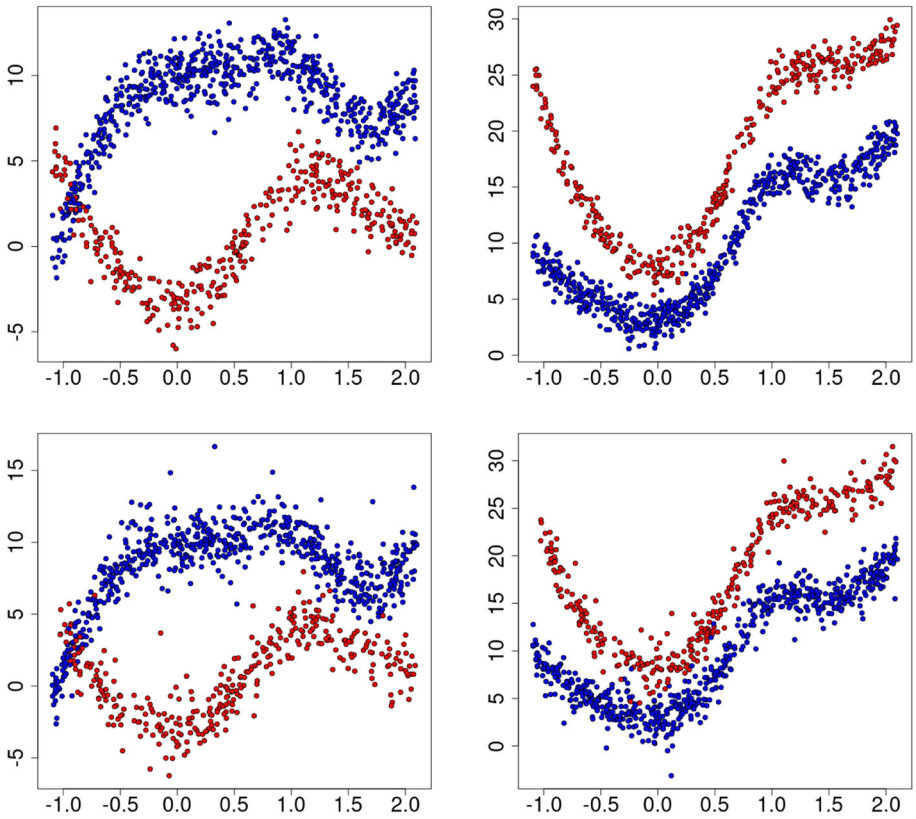

Fig. 2 Examples of data generated from the second scenario. Scatterplot of Y_1 (first column) and Y_2 (second column) under normal (first row) and Student t (second row) errors as a function of the included covariate, when using shifted-Poisson SDs. Red and blue data points distinguish the two latent states

Table 4 Average and standard deviation (in brackets) values of the PNR for Y_1 and Y_2 under the considered SDs and error term distributions for the first scenario over 1000 samples

PNR	τ									
	(0.10, 0.10)		(0.25, 0.25)		(0.50, 0.50)		(0.75, 0.75)		(0.90, 0.90)	
Panel A: $\mathcal{N}$										
SPO	0.096	0.083	0.265	0.233	0.497	0.504	0.751	0.769	0.920	0.912
	(0.005)	(0.004)	(0.008)	(0.007)	(0.007)	(0.008)	(0.006)	(0.006)	(0.005)	(0.004)
SNB	0.091	0.079	0.251	0.224	0.492	0.501	0.750	0.770	0.918	0.911
	(0.007)	(0.006)	(0.010)	(0.009)	(0.012)	(0.014)	(0.007)	(0.006)	(0.005)	(0.004)
GEO	0.088	0.076	0.246	0.231	0.499	0.502	0.753	0.769	0.920	0.913
	(0.008)	(0.006)	(0.012)	(0.017)	(0.007)	(0.008)	(0.007)	(0.006)	(0.004)	(0.005)
Panel B: $\mathcal{T}$										
SPO	0.097	0.080	0.266	0.232	0.498	0.502	0.750	0.768	0.918	0.914
	(0.006)	(0.004)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.006)	(0.005)	(0.004)
SNB	0.089	0.080	0.256	0.224	0.495	0.502	0.750	0.768	0.917	0.912
	(0.006)	(0.005)	(0.011)	(0.010)	(0.010)	(0.012)	(0.007)	(0.006)	(0.005)	(0.004)
GEO	0.095	0.080	0.249	0.231	0.500	0.500	0.752	0.769	0.918	0.914
	(0.008)	(0.005)	(0.014)	(0.017)	(0.007)	(0.009)	(0.007)	(0.006)	(0.004)	(0.005)

Table 5 Average and standard deviation (in brackets) values of the PNR for Y_1 and Y_2 under the considered SDs and error term distributions for the second scenario over 1000 samples

PNR	τ									
	(0.10, 0.10)		(0.25, 0.25)		(0.50, 0.50)		(0.75, 0.75)		(0.90, 0.90)	
Panel A: $\mathcal{N}$										
SPO	0.095	0.090	0.249	0.239	0.479	0.482	0.743	0.760	0.907	0.923
	(0.010)	(0.017)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.033)	(0.011)
SNB	0.090	0.083	0.250	0.239	0.479	0.481	0.742	0.758	0.922	0.926
	(0.008)	(0.013)	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.024)	(0.009)
GEO	0.092	0.084	0.251	0.238	0.486	0.483	0.742	0.758	0.909	0.926
	(0.008)	(0.014)	(0.006)	(0.006)	(0.008)	(0.009)	(0.008)	(0.008)	(0.042)	(0.011)
Panel B: $\mathcal{T}$										
SPO	0.090	0.082	0.254	0.245	0.483	0.485	0.746	0.756	0.933	0.926
	(0.009)	(0.015)	(0.006)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.012)	(0.009)
SNB	0.088	0.079	0.254	0.244	0.482	0.484	0.745	0.754	0.935	0.925
	(0.008)	(0.012)	(0.006)	(0.007)	(0.008)	(0.008)	(0.008)	(0.007)	(0.011)	(0.009)
GEO	0.090	0.080	0.254	0.243	0.487	0.485	0.745	0.755	0.923	0.924
	(0.008)	(0.012)	(0.006)	(0.007)	(0.008)	(0.008)	(0.008)	(0.007)	(0.031)	(0.011)

Ozone (O_3) and Nitrogen Dioxide (NO_2) from January 01, 2019 to June 14, 2021. We averaged the pollution data to daily frequency and all concentrations are expressed in $\mu g/m^3$.

Atmospheric variables also play a major role in determining the level of exposure to particular pollutants and capturing time dependence as proxies for seasonal variations and characteristics. Since pollution episodes are triggered by specific atmospheric factors, we include the following variables, namely the daily average wind speed, temperature, pressure and humidity. Table 7 presents the main descriptive statistics for the response variables and the set of included predictors. The asymmetry in the distributions of the pollutants is noted by examining the mean-median relationship and the five summary statistics indicate severe departures from the Gaussian assumptions, presenting high kurtosis and outlying values. In the same table we also report the empirical correlation coefficients between each response variable which clearly highlight a positive correlation between PM_{25} and NO_2, and a negative association with O_3. Therefore, the dependence structure among different pollutants, which can-

Table 6 Number of correctly identified hidden states K and maximum sojourn times $\mathbf{U}$ (Panel A) and absolute frequency distribution of the selected number of states (Panel B) under Normal and Student t errors over 100 replications

	τ														
	(0.10, 0.10)			(0.25, 0.25)			(0.50, 0.50)			(0.75, 0.75)			(0.90, 0.90)		
	AIC	BIC	ICL	AIC	BIC	ICL	AIC	BIC	ICL	AIC	BIC	ICL	AIC	BIC	ICL
Panel A															
$\mathcal{N}$	0	1	100	0	54	99	30	98	99	0	61	100	0	2	99
$\mathcal{T}$	0	0	82	0	27	97	72	86	97	0	22	97	0	0	92
							Panel B								
Panel B.1: $\mathcal{N}$															
2	0	1	100	0	56	100	30	99	99	0	62	100	0	2	100
3	0	70	0	0	44	0	4	1	1	0	38	0	0	80	0
4	100	29	0	100	0	0	66	0	0	100	0	0	100	18	0
Panel B.2: $\mathcal{T}$															
2	0	0	88	0	32	100	72	99	100	0	28	100	0	0	98
3	0	15	11	1	68	0	13	1	0	0	72	0	0	56	2
4	100	85	1	99	0	0	15	0	0	100	0	0	100	44	0

Table 7 Summary statistics of the sample data

Variable	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
PM_{25}	0	6	9	11.846	15	63
O_3	7	31	49	46.340	61	94
NO_2	0	8	12	14.139	20	44
Wind Speed	0	6	8	8.839	11	29
Temperature	0	11	16	16.659	22	34
Pressure	989	1011	1015	1015.127	1020	1036
Humidity	17	42	56	58.415	74	100
Correlation matrix						
	PM_{25}	O_3	NO_2			
PM_{25}	1					
O_3	−0.566	1				
NO_2	0.743	−0.656	1			

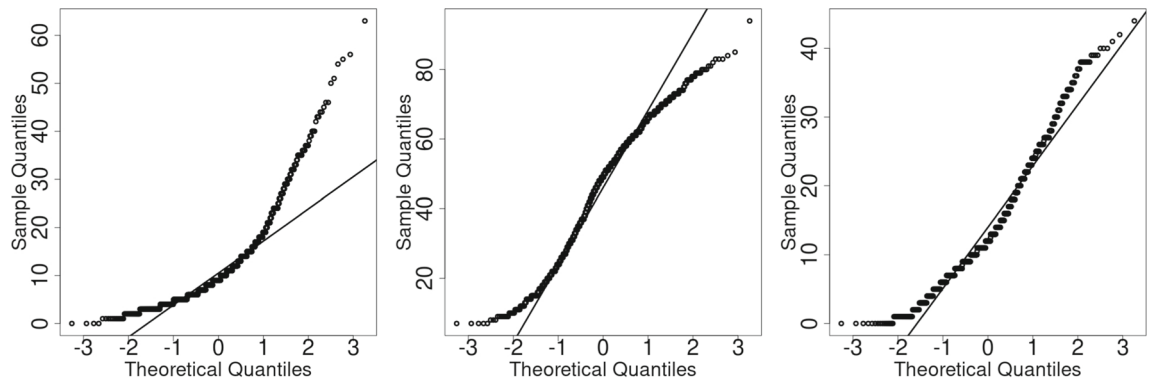

Fig. 3 From left to right, univariate normal QQ plots for PM_{25}, O_3 and NO_2

Table 8 Log-likelihood, AIC, BIC and ICL values for a varying number of hidden states. Bold font highlights the best values for the considered criteria (lower-is-better)

K	τ											
	(0.50, 0.50, 0.50)				(0.75, 0.75, 0.75)				(0.90, 0.90, 0.90)			
	Loglik	AIC	BIC	ICL	Loglik	AIC	BIC	ICL	Loglik	AIC	BIC	ICL
1	−8284.99	16611.98	16712.73	16712.73	−8553.92	17149.85	17250.61	17250.61	−9002.05	18046.10	18146.86	18146.86
2	−7893.43	15986.87	16466.66	16521.09	−8093.62	16387.25	16867.04	16937.16	−8415.43	17040.86	17544.65	17624.07
3	−7656.82	15549.64	**16115.80**	**16132.46**	−7808.18	15952.36	16758.42	16801.01	−8149.87	16635.73	17441.79	17490.93
4	−7434.64	15315.28	16385.22	16424.51	−7655.14	15756.27	16826.21	16908.00	−7926.16	16258.32	**17232.30**	**17396.04**
5	−7311.89	**15083.79**	16187.32	16240.48	−7523.67	**15507.34**	**16610.86**	**16755.08**	−7820.87	**16181.74**	17477.18	17638.80
6	−7195.51	15099.02	16797.49	16854.63	−7400.95	15509.90	17208.37	17284.46	−7761.82	16191.63	17794.14	17916.22

not be detected by univariate methods, constitutes a crucial aspect of the analysis and should not be neglected.

From a graphical standpoint, Fig. 3 shows the normal QQ plots for the PM_{25}, O_3 and NO_2 time series. These reveal the presence of potentially influential observations in the data, heavy tails and skewness for all three outcomes. This exploratory analysis and preliminary considerations motivate us to consider a joint quantile regression approach as investigative tool.

5.2 Results

We jointly model the concentrations of PM_{25}, O_3 and NO_2 as a function of Wind Speed, Temperature, Pressure and Humidity at quantile levels $\boldsymbol{\tau} = (0.50, 0.50, 0.50)$, $\boldsymbol{\tau} = (0.75, 0.75, 0.75)$ and $\boldsymbol{\tau} = (0.90, 0.90, 0.90)$. Considering the 75-th and 90-th percentiles puts emphasis on alert thresholds for ambient air pollution associated with high levels of concentrations of chemicals. As a first step of the analysis, we fit the proposed QHSMM for a sequence of states K from 1 to 6 and a K-dimensional grid, $\mathcal{U} \subset \mathbb{R}^K_{\geq 0}$, of maximum sojourn times, $(10, 15, \ldots, 60) \times \cdots \times (10, 15, \ldots, 60)$. To select the optimal combination of K and $\mathbf{U}$, with $\mathbf{U} \in \mathcal{U}$, and avoid the computational cost of a full grid search over $\mathcal{U}$, we employ the greedy search algorithm described in Sect. 3.2 with 200 starting points. Table 8 reports the log-likelihood, AIC, BIC and ICL values for the fitted models at the investigated quantile levels. The AIC selects $K = 5$ states for all quantile levels meanwhile the BIC and ICL are more aligned in the choice of K as they identify 3, 5 and 4 states for $\boldsymbol{\tau} = (0.50, 0.50, 0.50)$, $\boldsymbol{\tau} = (0.75, 0.75, 0.75)$ and $\boldsymbol{\tau} = (0.90, 0.90, 0.90)$. This is not surprising since the AIC tends to overestimate the number of hidden states and, for this reason, we will not consider it hereafter. We can also see that the ICL values associated to $K = 3$ and $K = 5$ are extremely similar at $\boldsymbol{\tau} = (0.75, 0.75, 0.75)$. Following these considerations and looking at Table 8, we select the best fitted models with K equal to 3, 3 and 4 according to both the BIC and ICL criteria at $\boldsymbol{\tau} = (0.50, 0.50, 0.50)$, $\boldsymbol{\tau} = (0.75, 0.75, 0.75)$ and $\boldsymbol{\tau} = (0.90, 0.90, 0.90)$, respectively.

Figures 4, 5 and 6 report the classification results according to the selected models at the investigated quantile levels. Each plot shows the data points colored according to the estimated posterior probability of class membership, $\widehat{\gamma}_{tk}$, with the vertical lines separating blocks of four months. In our study, the latent components can be associated to specific exposure regimes characterized by seasonal weather conditions. Specifically, blue points (state 2) tend to cluster days in late autumn, winter, and early spring, while green ones (state 3) are generally inferred during late spring and summer. At $\boldsymbol{\tau} = (0.90, 0.90, 0.90)$ (see Figure 6), the four state QHSMM identifies a similar classification pattern with violet points (state 4) occurring mainly from spring until the end of

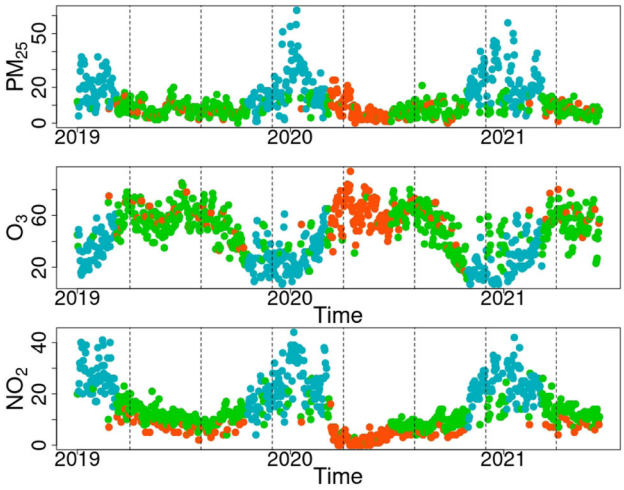

Fig. 4 Time series classified according to their estimated posterior probability of class membership at $\boldsymbol{\tau} = (0.50, 0.50, 0.50)$. Vertical lines separate blocks of four months

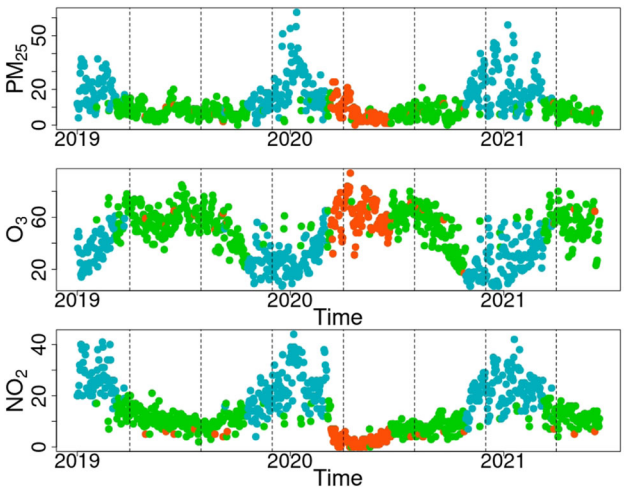

Fig. 5 Time series classified according to their estimated posterior probability of class membership at $\boldsymbol{\tau} = (0.75, 0.75, 0.75)$. Vertical lines separate blocks of four months

summer, but also sporadically during the rest of the year. The overall results reflect the seasonal variation in atmospheric pollutants, with high concentrations of PM_{25} and NO_2 in winter and high O_3 concentrations in summer and warm months. It is also worth noting that the minimum values of particulate matter and nitrogen dioxide, and the maximum level of ozone were reached in state 1 (orange dots) during the period March-June 2020. These sudden changes may be possibly due to the implementation of lockdown measures to contain the COVID-19 outbreak in Italy (Bassani et al. 2021; Putaud et al. 2021). Indeed, in the first weeks of March, significant PM_{25} and NO_2 declines were observed which led to O_3 peaks resulting from reduced titration with nitrogen oxides. As restrictions were lifted in May-June, chemical concentrations settled again around the 2019 and early 2021 levels.

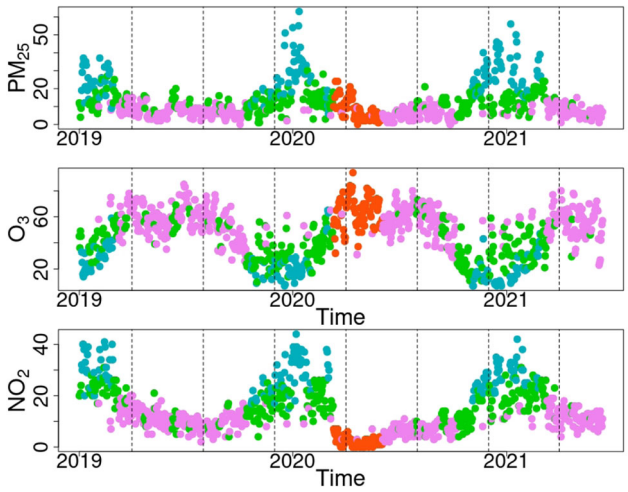

Fig. 6 Time series classified according to their estimated posterior probability of class membership at $\boldsymbol{\tau} = (0.90, 0.90, 0.90)$. Vertical lines separate blocks of four months

The estimated transition probability matrices of the latent semi-Markov chain (see Table 9) confirm that pollutant concentrations alternate between good and poor air conditions. The off-diagonal elements demonstrate that the hidden process stays in state 3 (low levels of particulate matter, nitrogen dioxide and moderate levels of ozone), and 4 in the case $\boldsymbol{\tau} = (0.90, 0.90, 0.90)$, with temporary changes towards more severe PM_{25} and NO_2 concentrations (state 2) or higher ozone episodes (state 1). Meanwhile, direct transitions between states 1 and 2 are very unlikely at all three quantile levels.

Taking the effect of covariates into account, Table 10 shows the estimated state-specific regression parameters, $\boldsymbol{\beta}_k, k = 1, \ldots, K$, at each quantile level, respectively. Standard errors are computed via parametric bootstrap using $H = 1000$ resamples as illustrated in Section 3 and point estimates are displayed in boldface when significant at the standard 5% level.

The estimated effects of the included covariates tend to be nonlinear, state-specific and are generally more pronounced in the upper end of the distribution of the responses. States 1 and 3 yield similar estimates as they are both associated with the best air conditions, especially when looking at effect of pressure and humidity. This is in sharp contrast with the point estimates in state 2 which is characterized by hazardous air quality. Among the four factors, wind speed and temperature exert the strongest influence on the considered pollutants in all seasons. In particular, PM_{25} and NO_2 are negatively associated with both variables and such association increases during the coldest months of the year. Wind intensity, therefore, contributes considerably to the reduction of pollution, in particular in at-risk situations as identified by state 2. On the other hand, O_3 concentration is positively associated with wind speed meanwhile the effect of temperature is positive in late-spring and hot weather, but negative throughout the rest of the year. Humidity can also help to decrease ozone pollution because the moisture in the air could enhance the condensation of water and slow down ozone production, but also reduce PM_{25} and NO_2. Further, the concentrations of PM_{25} and NO_2 are positively associated with atmospheric pressure and negatively associated with O_3.

We conclude the analysis by reporting the estimated correlation matrices, $\boldsymbol{\Psi}_k, k = 1, \ldots, K$, (see Table 11) in order to provide an indirect measure of tail dependence between the outcomes. Firstly, regardless of the state, the correlation coefficients between the air pollutants are generally significant, indicating that in this case fitting univariate quantile regressions separately would be inappropriate. Secondly, the correlation coefficients are increasing somewhat with $\boldsymbol{\tau}$. Finally, the estimates depict a data correlation structure that varies among the latent groups as the correlation between

Table 9 Estimated transition probabilities for different quantile levels

states	1	2	3	4
Panel A: (0.50, 0.50, 0.50)				
1	0	0.028 (0.009)	0.972 (0.009)	
2	0.000 (0.001)	0	1.000 (0.001)	
3	0.836 (0.020)	0.164 (0.020)	0	
Panel B: (0.75, 0.75, 0.75)				
1	0	0.000 (0.007)	1.000 (0.007)	
2	0.000 (0.000)	0	1.000 (0.000)	
3	0.543 (0.030)	0.457 (0.030)	0	
Panel C: (0.90, 0.90, 0.90)				
1	0	0.000 (0.000)	0.000 (0.009)	1.000 (0.009)
2	0.000 (0.000)	0	0.966 (0.014)	0.034 (0.014)
3	0.016 (0.007)	0.482 (0.029)	0	0.502 (0.030)
4	0.088 (0.025)	0.097 (0.021)	0.815 (0.032)	0

Table 10 Regression parameters estimates. Point estimates are displayed in boldface when significant at the standard 5% level

States	1			2			3			4		
	PM_{25}	O_3	NO_2	PM_{25}	O_3	NO_2	PM_{25}	O_3	NO_2	PM_{25}	O_3	NO_2
Panel A: (0.50, 0.50, 0.50)												
Intercept	**−192.006**	**408.316**	**−166.952**	**−488.373**	149.664	−110.794	**−163.261**	**690.359**	**−52.286**			
	(28.810)	(58.965)	(20.840)	(130.720)	(157.563)	(101.024)	(20.776)	(55.920)	(17.086)			
Wind Speed	**−0.323**	**0.320**	**−0.112**	**−0.538**	**1.869**	−0.227	−0.059	**0.704**	**−0.196**			
	(0.045)	(0.093)	(0.034)	(0.209)	(0.256)	(0.160)	(0.035)	(0.086)	(0.028)			
Temperature	**−0.236**	**−0.346**	**0.042**	**−1.095**	0.237	**−1.075**	**−0.100**	**0.314**	**−0.425**			
	(0.025)	(0.050)	(0.019)	(0.117)	(0.145)	(0.089)	(0.020)	(0.048)	(0.015)			
Pressure	**0.209**	**−0.309**	**0.167**	**0.515**	−0.118	0.156	**0.173**	**−0.610**	**0.076**			
	(0.028)	(0.057)	(0.020)	(0.126)	(0.152)	(0.097)	(0.020)	(0.054)	(0.016)			
Humidity	**−0.105**	**−0.619**	**0.042**	−0.006	**−0.248**	**−0.155**	**−0.030**	**−0.603**	**−0.059**			
	(0.012)	(0.023)	(0.008)	(0.053)	(0.066)	(0.041)	(0.009)	(0.022)	(0.007)			
Panel B: (0.75, 0.75, 0.75)												
Intercept	**−100.301**	**151.191**	**−119.282**	**−482.680**	**174.791**	**−108.564**	**−176.381**	**778.126**	**−44.040**			
	(37.049)	(75.844)	(17.254)	(75.045)	(80.288)	(53.198)	(21.221)	(50.143)	(15.392)			
Wind Speed	**−0.816**	−0.022	**−0.104**	**−0.584**	**1.832**	**−0.270**	**−0.069**	**0.554**	**−0.229**			
	(0.059)	(0.115)	(0.027)	(0.124)	(0.126)	(0.085)	(0.033)	(0.079)	(0.024)			
Temperature	**−0.621**	**−0.723**	**0.083**	**−1.170**	**0.212**	**−1.118**	**−0.105**	**−0.153**	**−0.412**			
	(0.032)	(0.062)	(0.015)	(0.067)	(0.071)	(0.046)	(0.018)	(0.045)	(0.014)			
Pressure	**0.137**	−0.039	**0.119**	**0.518**	−0.132	**0.158**	**0.190**	**−0.675**	**0.070**			
	(0.036)	(0.073)	(0.017)	(0.072)	(0.077)	(0.051)	(0.020)	(0.048)	(0.015)			
Humidity	**−0.151**	**−0.599**	**0.058**	−0.053	**−0.294**	**−0.170**	**−0.053**	**−0.646**	**−0.061**			
	(0.014)	(0.029)	(0.007)	(0.029)	(0.032)	(0.021)	(0.009)	(0.021)	(0.007)			
Panel C: (0.90, 0.90, 0.90)												
Intercept	−124.840	161.731	−72.439	**−476.669**	**−202.209**	**−82.551**	**−209.689**	**491.864**	34.514	**−148.541**	**575.923**	**−52.309**
	(157.253)	(257.786)	(59.375)	(76.346)	(47.414)	(32.034)	(24.663)	(43.603)	(22.358)	(22.284)	(59.029)	(19.866)
Wind Speed	**−1.137**	−0.369	**−0.366**	**0.744**	**3.277**	**0.562**	**−0.441**	**0.704**	**−0.221**	−0.055	**0.492**	**−0.278**
	(0.260)	(0.415)	(0.151)	(0.121)	(0.074)	(0.053)	(0.039)	(0.066)	(0.035)	(0.034)	(0.091)	(0.031)
Temperature	**−0.725**	0.178	**−0.191**	**−1.844**	−0.009	**−1.244**	**−0.261**	**0.494**	**−0.795**	**−0.149**	**−0.176**	**−0.428**
	(0.137)	(0.235)	(0.053)	(0.069)	(0.041)	(0.029)	(0.022)	(0.039)	(0.019)	(0.019)	(0.049)	(0.017)
Pressure	0.170	−0.055	0.084	**0.528**	**0.228**	**0.139**	**0.237**	**−0.423**	0.009	**0.165**	**−0.470**	**0.081**
	(0.152)	(0.247)	(0.057)	(0.074)	(0.046)	(0.031)	(0.024)	(0.042)	(0.021)	(0.021)	(0.057)	(0.019)
Humidity	**−0.153**	**−0.551**	−0.016	**−0.083**	**−0.212**	**−0.217**	**−0.056**	**−0.365**	**−0.105**	**−0.063**	**−0.582**	**−0.069**
	(0.064)	(0.107)	(0.025)	(0.033)	(0.019)	(0.013)	(0.010)	(0.018)	(0.009)	(0.009)	(0.024)	(0.008)

Table 11 Estimated state-dependent correlation matrices for different quantile levels. Point estimates are displayed in boldface when significant at the standard 5% level

States	$k=1$			$k=2$			$k=3$			$k=4$		
	PM_{25}	O_3	NO_2	PM_{25}	O_3	NO_2	PM_{25}	O_3	NO_2	PM_{25}	O_3	NO_2
Panel A: (0.50, 0.50, 0.50)												
PM_{25}	1			1			1					
O_3	0.098	1		−0.226	1		0.031	1				
	(0.062)			(0.154)			(0.057)					
NO_2	**0.224**	**−0.246**	1	**0.512**	**−0.344**	1	**0.391**	−0.020	1			
	(0.063)	(0.061)		(0.116)	(0.137)		(0.047)	(0.056)				
Panel B: (0.75, 0.75, 0.75)												
PM_{25}	1			1			1					
O_3	**0.206**	1		−0.024	1		**0.170**	1				
	(0.077)			(0.095)			(0.056)					
NO_2	**0.199**	**−0.422**	1	**0.586**	−0.142	1	**0.385**	−0.021	1			
	(0.084)	(0.082)		(0.061)	(0.097)		(0.050)	(0.061)				
Panel C: (0.90, 0.90, 0.90)												
PM_{25}	1			1			1			1		
O_3	0.151	1		**−0.361**	1		0.133	1		−0.018	1	
	(0.193)			(0.121)			(0.076)			(0.094)		
NO_2	**0.385**	**−0.591**	1	**0.202**	**−0.749**	1	0.078	**−0.230**	1	**0.186**	**−0.395**	1
	(0.182)	(0.203)		(0.077)	(0.190)		(0.077)	(0.095)		(0.089)	(0.114)	

PM_{25}, O_3 and NO_2 in the wintertime is substantially higher than that in spring and summer.

6 Conclusion

The study of pollution exposure is at the heart of policy attention for health and economics welfare analysis. Motivated the necessity to develop sound policies for controlling air contaminants emissions, this paper extends the joint quantile regression of Petrella and Raponi (2019) by introducing a hidden semi-Markov quantile regression for the analysis of multiple pollutants time series. The proposed model allows to capture quantile-specific effects across the entire distribution of several outcomes in one step and infer cluster-specific covariate effects at various quantile levels of interest. In order to avoid making biased inference related to incorrect distributional assumptions about the SDs, we adopt the approach of Guédon (2003) where the latent sojourn densities are approximated by using nonparametric discrete distributions and estimated directly from the data. Using simulation exercises, the proposed approach reveals promising results in reducing state misclassification rates with respect to HMMs and identifying the correct number of hidden states. In the empirical application, we employ our methodology to jointly model daily PM_{25}, O_3 and NO_2 concentrations recorded in Rieti (Italy) as a function of wind speed, temperature, pressure and humidity. We find that seasonal changes in air pollutant concentrations are greatly affected by meteorological conditions, whose effects are generally amplified at the top end of the distribution of the responses. Moreover, the latent regimes capture seasonal variations in air pollution particles that characterize low and hazardous contamination levels.

This work could be extended further in the following directions. In particular, the method proposed could be positively applied to financial time series modeling. Indeed, financial returns often exhibit empirical characteristics, such as skewness, leptokurtosis, heteroscedasticity and clustering behavior over time, which are heavily influenced by hidden variables (e.g., the state of the market) during tranquil and crisis periods (Maruotti et al. 2021). In this context, time dependence can be further included in the regression model, allowing the quantile to vary over time according to an autoregressive process (Engle and Manganelli 2004).

The approach introduced might also be extended to other settings and data structures. Firstly, our QHSMM can be generalized to multivariate longitudinal data, extending the recent quantile mixed HMM in Merlo et al. (2022) by using more flexible sojourn distributions and, possibly, allowing the semi-Markov chain parameters to depend on observable covariates. Secondly, another extension would deal with the inclusion of spatial heterogeneity into the modeling approach to account for spatial-temporal variations of air pollutants from different monitoring sites. Lastly, while in the

application we focused on daily concentrations of three air pollutants, in high-dimensional settings with a larger number of response variables and/or number of hidden states, the introduced QHSMM can be easily over-parameterized. This often occurs because of the large number of unique parameters in the covariance matrices to be estimated, implying a loss in terms of interpretability as well as numerically ill-conditioned estimators. In these cases, following Maruotti et al. (2017), we may consider a class of parsimonious HSMMs by imposing a factor decomposition on the state-specific covariance matrices. Not only would this modeling strategy provide information about the dependence between pollutants, but also a clear interpretation of the latent association structure among them.

Supplementary Materials

The Supplementary Materials include additional simulations that are used to support the results in the manuscript when the number of analyzed response variables p increases.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11222-022-10130-1.

Acknowledgements We would like to warmly thank the Associate Editor and two anonymous reviewers for their thoughtful comments and efforts towards improving our manuscript. This work was partially supported by the Finance Market Fund, Norway, project number 309218; "Statistical modelling and inference for (high-dimensional) financial data".

Appendix

Proof of Proposition 1 Firstly, we note that to ensure identifiability of the MAL density in (5) it suffices to apply Proposition 1 of Petrella and Raponi (2019). Secondly, for a general HSMM, identifiability has been proven up to label switching (Leroux 1992). Thus, to ensure identifiability, all one needs to prove is the identifiability of the marginal mixtures (Dannemann et al. 2014), which in our case are represented by the finite mixtures of MAL distributions. Based on the work of Holzmann et al. (2006), Browne and McNicholas (2015) prove identifiability of finite mixtures of multivariate generalized hyperbolic distributions. Since the MAL in (5) is a limiting case of the multivariate generalized hyperbolic distribution (see (3) and (4) in Browne and McNicholas (2015) with $\lambda = 1$, $\psi = 2$ and $\chi \to 0$), model identifiability follows by applying Corollary 2 of Browne and McNicholas (2015).

Proof of Proposition 2 The E-step of the EM algorithm considers the conditional expectation of the complete log-likelihood function in (8) given the observed data and the current parameter estimates $\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(r-1)}$. At first, we recall that under the constraints imposed on $\tilde{\boldsymbol{\xi}}$ and $\boldsymbol{\Lambda}$, the representation in (7) implies that:

$$\mathbf{Y} \mid \tilde{l} = \tilde{c} \sim \mathcal{N}_p(\boldsymbol{\mu} + \mathbf{D}\tilde{\boldsymbol{\xi}}\tilde{c}\,,\ \tilde{c}\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}), \qquad \tilde{C} \sim \text{Exp}(1). \quad (35)$$

This means that the joint density function of $\mathbf{Y}$ and $\tilde{C}$ is:

$$f_{\mathbf{Y},\tilde{C}}(\mathbf{y},\tilde{c}) = \frac{\exp\left\{(\mathbf{y}-\boldsymbol{\mu})'\mathbf{D}^{-1}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\xi}}\right\}}{(2\pi)^{p/2} \mid \mathbf{D}\boldsymbol{\Sigma}\mathbf{D} \mid^{1/2}} \left(\tilde{c}^{-p/2}\exp\left\{-\frac{1}{2}\frac{\tilde{m}}{\tilde{c}} - \frac{1}{2}\tilde{c}(\tilde{d}+2)\right\}\right). \quad (36)$$

By substituting (36) in (8) and taking the conditional expectation of the logarithm of (8), we obtain the expected complete log-likelihood function in (13).

To compute the conditional expectation of $\tilde{c}_{tk}$ and $\tilde{z}_{tk}$ in (13), $\tilde{C}$ is treated as an additional latent variable. Using the joint distribution of $\mathbf{Y}$ and $\tilde{C}$ derived in (36) and the MAL density of $\mathbf{Y}$ given in (5), we have that:

$$f_{\tilde{C}}(\tilde{C} \mid \mathbf{Y} = \mathbf{y}) = \frac{f_{\tilde{C},\mathbf{Y}}(\tilde{c},\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})} = \frac{\tilde{c}^{-p/2}\left(\frac{2+\tilde{d}}{\tilde{m}}\right)^{\nu/2}\exp\left\{-\frac{\tilde{m}}{2\tilde{c}} - \frac{\tilde{c}(2+\tilde{d})}{2}\right\}}{2K_\nu\left(\sqrt{(2+\tilde{d})\tilde{m}}\right)}, \quad (37)$$

which corresponds to a Generalized Inverse Gaussian (GIG) distribution with parameters $\nu, 2+\tilde{d}, \tilde{m}_i$, i.e.[1]

$$f_{\tilde{C}}(\tilde{C} \mid \mathbf{Y} = \mathbf{y}) \sim \text{GIG}\left(\nu, \tilde{d}+2, \tilde{m}\right). \quad (38)$$

Then, it follows that

$$\mathbb{E}[\tilde{C} \mid \cdot] = \left(\frac{\hat{\tilde{m}}}{2+\hat{\tilde{d}}}\right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}}\right)}{K_\nu\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}}\right)} \quad (39)$$

and

$$\mathbb{E}[\tilde{C}^{-1} \mid \cdot] = \left(\frac{2+\hat{\tilde{d}}}{\hat{\tilde{m}}}\right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}}\right)}{K_\nu\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}}\right)} - \frac{2\nu}{\hat{\tilde{m}}}. \quad (40)$$

[1] The pdf of a GIG(p,a,b) distribution is defined as $f_{GIG}(x;p,a,b) = \frac{\left(\frac{a}{b}\right)^{p/2}}{2K_p(\sqrt{ab})}x^{p-1}e^{-\frac{1}{2}\left(ax+bx^{-1}\right)}$, with $a > 0$, $b > 0$ and $p \in \mathcal{R}$.

Denoting the two conditional expectations in (39) and (40) by $\hat{\tilde{c}}$ and $\hat{\tilde{z}}$ respectively, concludes the proof.

Proof of Proposition 3 Imposing the first order conditions on (13) with respect to each component of the set $\boldsymbol{\Phi}_{\boldsymbol{\tau}}$, gives the update estimates in (16), (17), (18) and (19). However, there is not closed formula solution to update the elements of the scale matrix $\mathbf{D}_j$; hence, the M-step update requires using numerical optimization techniques to maximize (13). A considerable disadvantage of this procedure is the necessary high computational effort which could be very time-consuming. For this reason, we utilize a simpler estimator for the scale parameters $\delta_{jk}, k = 1, \ldots, p$, which follows directly from the fact that all marginals of the MAL distribution are univariate AL distributions (see Yu and Zhang 2005):

$$\hat{\delta}_{jk} = \frac{1}{T}\sum_{t=1}^{T} \hat{\gamma}_{tj}\rho_{\tau}(Y_t^{(k)} - \hat{\mu}_{tjk}), \tag{41}$$

where $\hat{\mu}_{tjk}$ is the k-th element of the vector $\hat{\boldsymbol{\mu}}_{tj}$.

References

Adam, T., Langrock, R., Weiß, C.H.: Penalized estimation of flexible hidden Markov models for time series of counts. Metron **77**(2), 87–104 (2019)

Akaike, H.: Information theory and an extension of the maximum likelihood principle, *in* 'Selected papers of Hirotugu Akaike', Springer, pp. 199–213, (1998)

Barbu, V. S. and Limnios, N.: *Semi-Markov chains and hidden semi-Markov models toward applications: their use in reliability and DNA analysis*, Vol. 191, Springer Science & Business Media (2009)

Bartolucci, F., Farcomeni, A. and Pennoni, F.: *Latent Markov models for longitudinal data*, CRC Press, (2012)

Bassani, C., Vichi, F., Esposito, G., Montagnoli, M., Giusto, M., Ianniello, A.: Nitrogen dioxide reductions from satellite and surface observations during COVID-19 mitigation in Rome (Italy). Environmental Science and Pollution Research **28**(18), 22981–23004 (2021)

Bernardi, M., Gayraud, G., Petrella, L., et al.: Bayesian tail risk interdependence using quantile regression. Bayesian Analysis **10**(3), 553–603 (2015)

Biernacki, C., Celeux, G., Govaert, G.: Assessing a mixture model for clustering with the integrated completed likelihood. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(7), 719–725 (2000)

Browne, R.P., McNicholas, P.D.: A mixture of generalized hyperbolic distributions. Canadian Journal of Statistics **43**(2), 176–198 (2015)

Bulla, J., Bulla, I.: Stylized facts of financial time series and hidden semi-Markov models. Computational Statistics & Data Analysis **51**(4), 2192–2209 (2006)

Bulla, J., Bulla, I., Nenadić, O.: hsmm-an R package for analyzing hidden semi-Markov models. Computational Statistics & Data Analysis **54**(3), 611–619 (2010)

Cappé, O., Moulines, E. and Rydén, T. (2006), *Inference in hidden Markov models*, Springer Science & Business Media

Charlier, I., Paindaveine, D., Saracco, J.: Multiple-output quantile regression through optimal quantization. Scandinavian Journal of Statistics **47**(1), 250–278 (2020)

Chavas, J.-P.: On multivariate quantile regression analysis. Statistical Methods & Applications **27**(3), 365–384 (2018)

Dannemann, J., Holzmann, H., Leister, A.: Semiparametric hidden Markov models: identifiability and estimation. Wiley Interdisciplinary Reviews: Computational Statistics **6**(6), 418–425 (2014)

Dempster, A. P., Laird, NM and Rubin, DB.: 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society Series B (Methodological)* pp. 1–38 (1977)

El Ghouch, A., Genton, M.G.: Local polynomial quantile regression with parametric features. J.Am. Stat. Assoc. **104**(488), 1416–1429 (2009)

Engle, R.F., Manganelli, S.: CAViaR: conditional autoregressive value at risk by regression quantiles. J. Bus. Econ. Stat. **22**(4), 367–381 (2004)

Ephraim, Y., Merhav, N.: Hidden Markov processes. IEEE Trans. Inf. Theor. **48**(6), 1518–1569 (2002)

Farcomeni, A.: Quantile regression for longitudinal data based on latent Markov subject-specific parameters. Stat. Comput. **22**(1), 141–152 (2012)

Geraci, M.: Modelling and estimation of nonlinear quantile regression with clustered data. Comput. Stat. Data Anal. **136**, 30–46 (2019)

Guédon, Y.: Estimating hidden semi-Markov chains from discrete sequences. J. Comput. Graph. Stat. **12**(3), 604–639 (2003)

Hamilton, J. D.:'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica: Journal of the Econometric Society* pp. 357–384, (1989)

Holzmann, H., Munk, A., Gneiting, T.: Identifiability of finite mixtures of elliptical distributions. Scandinav. J. Stat. **33**(4), 753–763 (2006)

Hubert, L., Arabie, P.: Comparing partitions. J. Classif. **2**(1), 193–218 (1985)

Koenker, R.: Quantile regression. Cambridge University Press (2005)

Koenker, R. and Bassett, G.:'Regression Quantiles', *Econometrica: Journal of the Econometric Society* **46**(1), 33–50, (1978)

Koenker, R., Chernozhukov, V., He, X. and Peng, L.: *Handbook of quantile regression*, CRC press, (2017)

Kong, L. and Mizera, I.: 'Quantile tomography: using quantiles with multivariate data', *Statistica Sinica* pp. 1589–1610, (2012)

Kotz, S., Kozubowski, T. and Podgorski, K.:*The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*, Springer Science & Business Media, (2012)

Langrock, R., Kneib, T., Sohn, A., DeRuiter, S.L.: Nonparametric inference in hidden Markov models using P-splines. Biometrics **71**(2), 520–528 (2015)

Langrock, R., Zucchini, W.: Hidden Markov models with arbitrary state dwell-time distributions. Comput. Stat. Data Anal. **55**(1), 715–724 (2011)

Leroux, B.G.: Maximum-likelihood estimation for hidden Markov models. Stochastic Process. Appl. **40**(1), 127–143 (1992)

Levinson, S.E., Rabiner, L.R., Sondhi, M.M.: 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 (1983)

Luo, Y., Lian, H., Tian, M.: Bayesian quantile regression for longitudinal data models. J. Stat. Comput. Simul. **82**(11), 1635–1649 (2012)

MacDonald, I. L. and Zucchini, W.:*Hidden Markov and other models for discrete-valued time series*, Vol. 110, CRC Press (1997)

Marino, M.F., Tzavidis, N., Alfò, M.: Mixed hidden Markov quantile regression models for longitudinal data with possibly incomplete sequences. Stat. Methods Med. Res **27**(7), 2231–2246 (2018)

Maruotti, A.: Mixed hidden Markov models for longitudinal data: an overview. Int. Stat. Rev. **79**(3), 427–454 (2011)

Maruotti, A., Bulla, J., Lagona, F., Picone, M. and Martella, F.: 'Dynamic mixtures of factor analyzers to characterize multivariate air pollutant exposures', *The Annals of Applied Statistics* pp. 1617–1648,(2017)

Maruotti, A., Petrella, L., Sposito, L.: Hidden semi-Markov-switching quantile regression for time series. Compu. Stati. Data Anal. **159**, 107208 (2021)

Maruotti, A., Punzo, A.: Initialization of hidden Markov and semi-Markov models: a critical evaluation of several strategies. Int. Stat. Rev. **89**(3), 447–480 (2021)

Maruotti, A., Punzo, A., Bagnato, L.: Hidden Markov and semi-Markov models with multivariate leptokurtic-normal components for robust modeling of daily returns series. J. Financ. Econom. **17**(1), 91–117 (2019)

Merlo, L., Petrella, L., Raponi, V.: Forecasting VaR and ES using a joint quantile regression and its implications in portfolio allocation. J. Bank. Finan. **133**, 106248 (2021)

Merlo, L., Petrella, L., Salvati, N., Tzavidis, N.: Marginal M-quantile regression for multivariate dependent data. Comput. Stat. Data Anal. **173**, 107500 (2022)

Merlo, L., Petrella, L., Tzavidis, N.: 'Quantile mixed hidden Markov models for multivariate longitudinal data: an application to children's Strengths and Difficulties Questionnaire scores', Journal of the Royal Statistical Society. Ser. C Appl. Stat. **71**(2), 417–448 (2022)

O'Connell, J., Højsgaard, S., et al.: Hidden semi Markov models for multiple observation sequences: the mhsmm package for R. J. Stat. Softw. **39**(4), 1–22 (2011)

Petrella, L., Raponi, V.: Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress. J. Multivar. Anal. **173**, 70–84 (2019)

Pohle, J., Adam, T., Beumer, L.T.: Flexible estimation of the state dwell-time distribution in hidden semi-Markov models. Comput. Stat. Data Anal. **172**, 107479 (2022)

Pohle, J., Langrock, R., van Beest, F.M., Schmidt, N.M.: Selecting the number of states in hidden Markov models: pragmatic solutions illustrated using animal movement. J. Agric Biol. Environ. Stat. **22**(3), 270–293 (2017)

Putaud, J.-P., Pozzoli, L., Pisoni, E., Martins Dos Santos, S., Lagler, F., Lanzani, G., Dal Santo, U., Colette, A.: Impacts of the COVID-19 lockdown on air pollution at regional and urban background sites in northern Italy. Atmosp. Chem. Phys. **21**(10), 7597–7609 (2021)

Sansom, J. and Thomson, P. (2001), 'Fitting hidden semi-Markov models to breakpoint rainfall data', *Journal of Applied Probability* **38**(A), 142–157

Schwarz, G., et al.: Estimating the dimension of a model. Anna. Stat. **6**(2), 461–464 (1978)

Serfling, R.: Quantile functions for multivariate analysis: approaches and applications. Statist. Neerlandica **56**(2), 214–232 (2002)

Stolfi, P., Bernardi, M. and Petrella, L.: 'The sparse method of simulated quantiles: An application to portfolio optimization', *Statistica Neerlandica* (2018),

Visser, I., Raijmakers, M.E., Molenaar, P.C.: Confidence intervals for hidden Markov model parameters. Br. J. Math. Statist. Psychol. **53**(2), 317–327 (2000)

Ye, W., Zhu, Y., Wu, Y. and Miao, B.: 'Markov regime-switching quantile regression models and financial contagion detection', *Insurance: Mathematics and Economics* **67**, 21–26,(2016)

Yu, K., Moyeed, R.A.: Bayesian quantile regression. Stat. Probab. Lett. **54**(4), 437–447 (2001)

Yu, K., Zhang, J.: A three-parameter asymmetric Laplace distribution and its extension. Commun. Statist. Theory Methods **34**(9–10), 1867–1879 (2005)

Yu, S.-Z.: Hidden Semi-Markov models: theory, algorithms and applications. Morgan Kaufmann (2015)

Zucchini, W., MacDonald, I. L. and Langrock, R.: *Hidden Markov models for time series: an introduction using R*, Chapman and Hall CRC (2016)