Two-part quantile regression models for semi-continuous longitudinal data: A finite mixture approach

Luca Merlo[1], Antonello Maruotti[2,3] and Lea Petrella[4]

[1]Department of Statistical Sciences, Sapienza University of Rome, 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

Abstract: This article develops a two-part finite mixture quantile regression model for semi-continuous longitudinal data. The proposed methodology allows heterogeneity sources that influence the model for the binary response variable to also influence the distribution of the positive outcomes. As is common in the quantile regression literature, estimation and inference on the model parameters are based on the asymmetric Laplace distribution. Maximum likelihood estimates are obtained through the EM algorithm without parametric assumptions on the random effects distribution. In addition, a penalized version of the EM algorithm is presented to tackle the problem of variable selection. The proposed statistical method is applied to the well-known RAND Health Insurance Experiment dataset which gives further insights on its empirical behaviour.

1 Introduction

Semi-continuous data are non-negative and characterized by the coexistence of high kurtosis, positive skewness and an abundance of zeros observed often enough that there are compelling substantive and statistical reasons for special treatment (Belotti et al., 2015). Modelling non-negative data with clumping at zero has a long history in the statistical literature and several competing models have been proposed (see Min and Agresti 2002; Min and Agresti 2005 and references therein). Such data structure arises on many occasions: economics, actuarial sciences, environmental modelling and health services (see Liu, 2009; Belasco and Ghosh, 2012; Neelon et al.,

---

Address for correspondence: Luca Merlo, Department of Statistical Sciences, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy.
E-mail: luca.merlo@uniroma1.it

2016a,b; Farewell et al., 2017; Merlo et al., 2020). Linear regression approaches can be considered as starting points of the analysis (see e.g., Iversen et al., 2015). Nevertheless, parameter estimates are sensitive to extreme values and likely to be inefficient if the underlying distribution is not Gaussian (Zhou, 2002; Basu and Manning, 2009). Thus, other approaches have attracted researchers attention to model semi-continuous data. Among those, two-part models, introduced by Duan et al., (1983) and Mullahy (1986), play an important role. This class of models helps handle excess of zeros and overdispersion: they involve a mixture distribution consisting in a mixing of a discrete point mass, with all mass at zero, and a discrete or continuous random variable. In particular, they are described by two equations: a binary choice model is fitted for the probability of observing a positive-versus-zero outcome. Then, conditional on a positive outcome, an appropriate regression model is fitted for the positive outcome (see Belotti et al., 2015). This statistical method introduces significant modelling flexibility by allowing the zeros and the positive values to be generated by two different processes. For these reasons, they have been deeply implemented especially in biomedical applications and in health economics as they well recover the two-step structure in the health demand process (see, e.g., Diehr et al., 1999; Deb and Trivedi, 2002; Alfò and Maruotti, 2010; Mihaylova et al., 2011; Maruotti and Raponi, 2014).

The common structure of such models assumes that the effect of the covariates influence the mean of the conditional distribution of the response. However, in many real applications, the effect of the covariates can be different on different parts of the response distribution. In these cases, it may be of interest to infer on the entire conditional distribution of the response variable using the quantile regression approach proposed in the seminal article by Koenker and Bassett (1978) which allows for quantile-specific inference, and it is typically used for modelling non-Gaussian outcomes.

Quantile regression methods have become widely used in literature mainly because they are suitable in all those situations where skewness, fat-tails, outliers, truncation, censoring and heteroscedasticity arise. They have been implemented in a wide range of different fields, both in a frequentist paradigm and in a Bayesian setting, spanning from medicine (see Cole and Green, 1992; Royston and Altman, 1994; Alhamzawi et al., 2012; Waldmann, 2018), financial and economic research (see Bassett and Chen, 2002; Bernardi et al., 2015; Petrella et al., 2018; Laporta et al., 2018; Tian et al., 2018; Bernardi et al., 2018; Petrella and Raponi, 2019) and environmental modelling (see, for example, Hendricks and Koenker, 1992; Pandey and Nguyen 1999; Reich et al., 2011 for a discussion. For a detailed review and list of references, Koenker (2005) and Koenker et al. (2017) provide an overview of the most used quantile regression techniques in a classical setting.

In longitudinal studies, quantile methods with random effects have been proposed in order to account for the dependence between serial observations on the same subject (see Marino and Farcomeni, 2015; Alfò et al., 2017; Marino et al., 2018). Alfò et al. (2017), for example, defined a finite mixture of quantile regression models for heterogeneous data. Quantile regression and two-part models have also been positively considered in several studies (e.g, Grilli et al., 2016; Heras et al., 2018;

Sauzet et al.,2019; Biswas et al., 2020). In particular, Biswas et al., (2020) considered a semi-parametric quantile regression approach to zero-inflated and incomplete longitudinal outcomes in a repeated measurements design study.

From an inferential point of view both classical and Bayesian inferential approaches have been used in the literature to estimate the parameters and the quantiles of the models. In the frequentist setting, the inferential approach used to estimate the parameters relies on the minimization of the asymmetric loss function of Koenker and Bassett (1978), while in the Bayesian setting, the asymmetric Laplace (AL) distribution has been introduced as a likelihood inferential tool (see Yu and Moyeed, 2001). The two approaches are well-justified by the relationship between the quantile loss function and the AL density: 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. Therefore, the AL distribution could offer a convenient device to implement a likelihood based inferential approach in a quantile regression analysis.

The main goal of the present article is to extend the two-part quantile regression modelling framework for mixed-type outcomes to longitudinal data using a frequentist approach. In particular, we consider a mixed effect logistic regression for modelling the probability of zero-nonzero outcome and a linear mixed quantile regression model for the continuous positive outcomes. Following Alfò et al. (2017), in order to prevent inconsistent parameter estimates due to misspecification of the random effects distribution, we adopt a nonparametric approach in which the random effect is left unspecified and approximated by using a discrete finite mixture. Within this scheme, our modelling framework reduces to a two-part finite mixture of quantile regressions where the components of the finite mixture represent clusters of units that share homogeneous values of model parameters.

We propose to estimate model parameters through Maximum Likelihood (ML) by using the AL distribution as a working likelihood. Specifically, estimation is carried out through the Expectation-Maximization (EM) algorithm. From a computational perspective, we generalize the work of Tian et al. (2014) and Tian et al. (2016) and provide an efficient version of the EM algorithm with M-step updates in closed form using the well-known location-scale mixture representation of the AL distribution; see Kozumi and Kobayashi (2011).

In statistical modelling, one of the main issue is the identification of the relevant variables to be considered in the model. It is in fact quite common, using real data, that a large number of predictors are concerned in the initial stage of the analysis. In this situation the researcher would be interested in determining a smaller subset that exhibits the strongest effects. Several variable selection methods have been proposed in the literature, one of them is the penalized method which is particularly useful when dealing with high dimensional statistical problems (see Wasserman and Roeder, 2009; Fan and Lv, 2010). To improve estimation, to gain in parsimony and to conduct a variable selection procedure, we consider a Penalized EM (PEM) algorithm by introducing the Least Absolute Shrinkage and Selecting Operator (LASSO) L_1 penalty term of Tibshirani (1996).

The relevance of our approach is also shown empirically by the analysis of a sample taken from the RAND Health Insurance Experiment (RHIE). The RHIE is one of the largest social experiments ever completed in the USA. to study the cost sharing and its effect on service use, quality of care and health expenditures; see Deb and Trivedi (2002). Healthcare data represents a striking example of semi-continuous data because they are non-negative, with substantial positive skewness, heavy tailed and, often, multi-modal, for example, they exhibit a spike-at-zero for non-users. In this article, we adopt the proposed two-part finite mixture of quantile regressions in order to investigate whether the effect of socioeconomic and household's characteristics changes with the increase in conditional health spending. In accordance with the results of Deb and Trivedi (2002), our analysis shows that the two-part model identifies two groups of users: a group of reluctant users and a second group that often uses healthcare services. In addition, the effect of the included covariates is not uniform across quantiles but it changes sign and magnitude as the quantile level varies.

The rest of the article is organized as follows. In Section 2, we introduce the two-part finite mixture quantile regression model. Section 3 illustrates the EM-based maximum likelihood approach, the closed form solutions and the PEM algorithm. Section 4 presents the simulation study. In Section 5 we discuss the main empirical results while Section 6 concludes the article.

2   Methodology

2.1   Two-part quantile regression model

Let y_{it}, $i = 1, ..., N$, $t = 1..., T_i$ be a semi-continuous variable for unit i at time t and let $\mathbf{b}_i = (\mathbf{b}_{i0}, \mathbf{b}_{i1})$ be a time-constant, individual-specific, random effects vector having distribution $f_{\mathbf{b}}(\cdot)$ with support $\mathcal{B}$ where $\mathbb{E}[\mathbf{b}_i] = 0$ is used for parameter identifiability. The role of the random coefficients $\mathbf{b}_i$ is to capture unobserved heterogeneity and within subject dependence. In a two-part model, the probability distribution of the outcome variable y_{it} can be written as:

$$f(y_{it}) = p_{it}^{d_{it}} \Big[(1 - p_{it}) \, g\big(h(y_{it})|y_{it} > 0\big) \Big]^{1-d_{it}} \tag{2.1}$$

with

$$d_{it} = \mathbf{1}(y_{it} = 0), \quad p_{it} = \Pr(y_{it} = 0) = \Pr(d_{it} = 1),$$

where d_{it} denotes the occurrence variable for unit i at time t, $\mathbf{1}(\cdot)$ is the indicator function, $g(\cdot)$ is the density function for the positive outcome given that $y_{it} > 0$ and $h(\cdot)$ is a (monotone) transformation function of y_{it}. The model is completed by defining the linear predictors for the binary and the positive parts of the model. We

assume that the spike-at-zero process is governed by a binary logistic model such that:

$$\text{logit}(p_{it} \mid \mathbf{s}_{it}, \mathbf{b}_{i0}) = \mathbf{s}_{it}'\boldsymbol{\gamma} + \mathbf{c}_{it}'\mathbf{b}_{i0}, \tag{2.2}$$

where $\mathbf{s}_{it} = (s_{it1}, \ldots, s_{itm})$ is the m dimensional set of explanatory variables, $\boldsymbol{\gamma}$ is the parameter vector and $\mathbf{c}_{it}$ is a subset of $\mathbf{s}_{it}$.

As mentioned in the Introduction, in order to determine the effect of explanatory variables on the tails of the distribution of the outcome and make inference at an arbitrary quantile level, we model the positive outcomes using the quantile regression approach. In the quantile regression literature, it is well established that the likelihood approach is based on the AL distribution of Yu and Moyeed (2001), which gives equivalent estimates to the minimization of the loss function in Koenker and Bassett (1978). The functional form of the AL distribution for our model is the following:

$$g(y_{it}; \mu_{it}(\tau), \sigma(\tau)) = \frac{\tau(1 - \tau)}{\sigma(\tau)} \exp\left\{ -\rho_\tau\left(\frac{y_{it} - \mu_{it}(\tau)}{\sigma(\tau)} \right) \right\}, \tag{2.3}$$

where $\mu_{it}(\tau)$ represents the τth quantile, with $\tau \in (0, 1)$, of y_{it}, $\sigma(\tau) > 0$ is the scale parameter and $\rho_\tau(\cdot)$ denotes the quantile loss function of Koenker and Bassett (1978):

$$\rho_\tau(u) = u(\tau - \mathbf{1}(u < 0)). \tag{2.4}$$

Because we are modelling positive values, to match the support of the AL density we consider the logarithmic transformation of the positive values of y_{it}. In particular, we assume that for a given τ, conditionally on $\mathbf{b}_{i1}$ and after log-transforming the outcome variable, $\tilde{y}_{it} = \log(y_{it})$, the conditional density $g(\tilde{y}_{it}|y_{it} > 0, \mathbf{x}_{it}, \mathbf{b}_{i1})$ in (2.1) is an AL distribution as given in (2.3) whose location parameter is defined by the linear model:

$$\mu_{it}(\tau) = \mathbf{x}_{it}'\boldsymbol{\beta}(\tau) + \mathbf{z}_{it}'\mathbf{b}_{i1}(\tau) \tag{2.5}$$

where $\mathbf{z}_{it}$ is a subset of covariates of $\mathbf{x}_{it}$.

For a fixed quantile level τ, responses are assumed to be independent conditional on the random vector $\mathbf{b}_i(\tau)$ and parameter estimates can be obtained by maximizing the likelihood function of the model defined in (2.1)-(2.5):

$$L(\Phi_\tau) = \prod_{i=1}^{N} \left\{ \int_{\mathcal{B}} \prod_{t=1}^{T_i} \left(p_{it}^{d_{it}} \left[(1 - p_{it}) g_{it} \right]^{1 - d_{it}} \right) f_{\mathbf{b}}(\mathbf{b}_i) \mathrm{d}\mathbf{b}_i \right\} \tag{2.6}$$

where $\Phi_\tau = \{\boldsymbol{\gamma}(\tau), \boldsymbol{\beta}(\tau), \sigma(\tau)\}$ denotes the global set of model parameters. The likelihood in (2.6) involves a multidimensional integral over the random coefficients whose corresponding distribution $f_{\mathbf{b}}(\cdot)$ allows to explain differences in the response quantiles across individuals. Hence, the choice of an appropriate distribution should be data driven and resistant to misspecification (Marino and Farcomeni 2015). In

the next section, we will discuss how we may avoid evaluating the integral in (2.6) for ML estimation.

2.2 Finite mixture of quantile regressions

In the literature, typically the Gaussian distribution is a convenient choice for $f_{\mathbf{b}}(\cdot)$ from a computational point of view. In this case, we may approximate the integral in (2.6) using Gaussian quadrature or adaptive Gaussian quadrature schemes (see Winkelmann, 2004; Rabe-Hesketh et al., 2005). A disadvantage of such approaches lies in the required computational effort, which is exponentially increasing with the dimension of the random parameter vector. For these reasons, potential alternatives availed themselves of simulation methods such as Monte Carlo and simulated ML approaches. However, for samples of finite size and short individual sequences, these methods may not provide a good approximation of the true mixing distribution (Alfò et al. 2017). As a robust alternative to the Gaussian choice, a Symmetric Laplace or a multivariate Student t random variable have been considered by Geraci and Bottai (2014) and Farcomeni and Viviani (2015). However, a parametric assumption on the distribution of the random coefficients could be rather restrictive and misspecification of the mixing distribution could lead to biased parameter estimates (see Alfò and Maruotti, 2010). In view of these considerations, in this work we exploit the approach based on the Nonparametric Maximum Likelihood (NPML) estimation of Laird (1978). Instead of specifying parametrically the distribution $f_{\mathbf{b}}(\cdot)$ we approximate it by using a discrete distribution on $G < N$ locations $\mathbf{b}_k(\tau) = (\mathbf{b}_{0k}(\tau), \mathbf{b}_{1k}(\tau))$, that is, $\mathbf{b}_i(\tau) \sim \sum_{k=1}^{G} \pi_k(\tau)\delta_{\mathbf{b}_k(\tau)}$ where the probability $\pi_k(\tau)$ is defined by $\pi_k(\tau) = \Pr(\mathbf{b}_i(\tau) = \mathbf{b}_k(\tau))$, $i = 1, ..., N$ for $k = 1, ..., G$ and $\delta_{\mathbf{b}_k(\tau)}$ is a one-point distribution putting a unit mass at $\mathbf{b}_k(\tau)$. The proposed approach can be thought as an approximation of a fully parametric framework as the discrete support approximates a possibly continuous distribution for the random coefficients. For the interested reader, a detailed survey on the NPML method can be found in Aitkin and Alfó (1998); Alfò and Aitkin (2000); Aitkin and Alfò (2003) and Alfò et al. (2017), for example.

To ease the notation, hereinafter we omit the quantile level τ but all parameters are allowed to depend on it. In this setting, the likelihood in (2.6) reduces to:

$$L(\Phi_\tau) = \prod_{i=1}^{N} \left\{ \sum_{k=1}^{G} \prod_{t=1}^{T_i} \left(p_{itk}^{d_{it}} \left[(1 - p_{itk}) g_{itk} \right]^{1-d_{it}} \right) \pi_k \right\}, \qquad (2.7)$$

where $\Phi_\tau = \{\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma, \mathbf{b}_1, ..., \mathbf{b}_G, \pi_1, ..., \pi_G\}$ is the parameter vector.

The likelihood in (2.7) is similar to the likelihood of a finite mixture of quantile regressions with G clusters. More specifically, in the kth cluster the spike-at-zero process is governed by the binary logistic model, $\text{logit}(p_{itk} \mid \mathbf{s}_{it}, \mathbf{b}_{0k}) = \mathbf{s}'_{it}\boldsymbol{\gamma} + \mathbf{c}'_{it}\mathbf{b}_{0k}$; meanwhile the positive outcomes process is regulated by a linear mixed quantile with AL density in (2.3) having location parameter given by $\mu_{itk} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{b}_{1k}$.

Therefore, our modelling framework reduces to a finite bivariate mixture model for each quantile level where heterogeneity sources that influence the binary decision process, are assumed to influence also the distribution of the positive outcomes through the latent structure defined by discrete multivariate random effects.

3 Estimation

In this section, we propose an ML approach based on the EM algorithm to estimate the parameters of the methodology illustrated in Section 2. Given the finite mixture representation in (2.7), each unit i can be conceptualized as drawn from one of G distinct groups: we denote with w_{ik} the indicator variable that is equal to 1 if the ith unit belongs to the kth component of the finite mixture, and 0 otherwise. The EM algorithm treats as missing data the component membership w_{ik}. Thus, the log-likelihood for the complete data has the following form:

$$\ell_c(\Phi_\tau) = \sum_{i=1}^{N} \sum_{k=1}^{G} w_{ik} \left\{ \sum_{t=1}^{T_i} \log \left(p_{itk}^{d_{it}} \left[(1 - p_{itk}) \, g_{itk} \right]^{1-d_{it}} \right) + \log(\pi_k) \right\}. \qquad (3.1)$$

In the E-step, the presence of the unobserved group-indicator w_{ik} is handled by taking the conditional expectation of w_{ik} given the observed data and the parameter estimates at the rth iteration $\hat{\Phi}_\tau^{(r)} = \{\hat{\boldsymbol{\gamma}}^{(r)}, \hat{\boldsymbol{\beta}}^{(r)}, \hat{\sigma}^{(r)}, \hat{\mathbf{b}}_1^{(r)}, ..., \hat{\mathbf{b}}_G^{(r)}, \hat{\pi}_1^{(r)}, ..., \hat{\pi}_G^{(r)}\}$. At the $(r + 1)$th iteration of the algorithm, we replace w_{ik} by its conditional expectation $\hat{w}_{ik}^{(r+1)}$ using the following update equation:

$$\hat{w}_{ik}^{(r+1)} = \mathbb{E}[w_{ik}|y_{it}, \mathbf{s}_{it}, \mathbf{x}_{it}, \hat{\Phi}_\tau^{(r)}] = \frac{\prod_{t=1}^{T_i} f_{itk}^{(r)} \hat{\pi}_k^{(r)}}{\sum_{l=1}^{G} \prod_{t=1}^{T_i} f_{itl}^{(r)} \hat{\pi}_l^{(r)}}, \qquad (3.2)$$

where $f_{itk}^{(r)} = p_{itk}(\hat{\Phi}_\tau^{(r)})^{d_{it}} \left[\left(1 - p_{itk}(\hat{\Phi}_\tau^{(r)}) \right) g_{itk}(\hat{\Phi}_\tau^{(r)}) \right]^{1-d_{it}}$. Conditionally on the posterior probabilities $\hat{w}_{ik}^{(r+1)}$ in (3.2), the M-step solutions are generally updated by maximizing $\mathbb{E}[\ell_c(\Phi_\tau) \,|\, y_{it}, \mathbf{s}_{it}, \mathbf{x}_{it}, \hat{\Phi}_\tau^{(r)}]$ with respect to Φ_τ using numerical optimization techniques. The E- and M-steps are alternated until convergence, that is when the difference between the likelihood function evaluated at two consecutive iterations is smaller than a predetermined threshold. In this article, we set this convergence criterion equal to 10^{-5}. To avoid convergence to local maxima, for each value of G, we initialize model parameters using a multi-start strategy: we considered 20 different starting points and retained the solution corresponding to the ML value.

A considerable disadvantage of this procedure affects the M-step which could require a high computational effort and it could be very time-consuming especially when the set of explanatory variables is large. Therefore, in the next section we

develop a more efficient approach for estimating two-part quantile regression models, and to obtain a closed form of the ML estimator.

3.1 Closed form EM algorithm solutions

This section develops an efficient estimation method for the two-part quantile regression problem based on the EM algorithm presented in Section 3. We reduce the computational burden of the algorithm compared to direct maximization of the likelihood and extend the work of Tian et al. (2014) and Tian et al. (2016) to two-part mixture models. In particular, we obtain iteratively closed form expressions for the unknown parameters vector $\boldsymbol{\beta}$ and the coefficient vectors $\mathbf{b}_{1k}$ for $k = 1, \ldots, G$.

We use the location-scale mixture representation of the AL density considered in Kozumi and Kobayashi (2011) to specify the positive outcome process of the model in (2.5) as the following hierarchical model:

$$\tilde{y}_{it} \mid (y_{it} > 0, \mathbf{x}_{it}, \mathbf{b}_{1k}, v_{it}) \sim \mathcal{N}(\mu_{itk} + \theta v_{it}, \rho^2 \sigma v_{it}), \quad v_{it} \sim \text{Exp}\left(\frac{1}{\sigma}\right), \qquad (3.3)$$

where $\mu_{itk} = \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_{it}'\mathbf{b}_{1k}$, $\theta = \frac{1-2\tau}{\tau(1-\tau)}$ and $\rho^2 = \frac{2}{\tau(1-\tau)}$. The constraints imposed on θ and ρ must be satisfied in order to guarantee that μ_{itk} coincides with the τth conditional quantile of $\tilde{y}_{it} \mid (y_{it} > 0, \mathbf{x}_{it}, \mathbf{b}_{1k})$.

Due to the independence of the v_{it}'s, one can obtain that the conditional distribution of v_{it} is a Generalized Inverse Gaussian (GIG) distribution (see Tian et al., 2016), namely:

$$f(v_{it}|\tilde{y}_{it}, y_{it} > 0, \mathbf{x}_{it}, \mathbf{b}_{1k}) \sim \text{GIG}\left(\frac{1}{2}, \frac{(\tilde{y}_{it} - \mu_{itk})^2}{\rho^2\sigma}, \frac{2\rho^2 + \theta^2}{\rho^2\sigma}\right). \qquad (3.4)$$

In addition, from (3.3) the joint density of $\tilde{y}_{it}$ and v_{it} is as follows:

$$f(\tilde{y}_{it}, v_{it}|y_{it} > 0, \mathbf{x}_{it}, \mathbf{b}_{1k}) = \frac{1}{\sqrt{2\pi\sigma v_{it}}\sigma\rho} \exp\left(-\frac{(\tilde{y}_{it} - \mu_{itk} - \theta v_{it})^2}{2\rho^2\sigma v_{it}} - \frac{v_{it}}{\sigma}\right). \qquad (3.5)$$

The underlying idea to obtain the updated parameter estimates of $\boldsymbol{\beta}$ and $\mathbf{b}_{1k}$ for $k = 1, \ldots, G$ is to consider v_{it} as an additional latent variable. According to (3.5), after omitting terms which of are not dependent on $\boldsymbol{\beta}$ and $\mathbf{b}_{1k}$, the complete data log-likelihood function is proportional to:

$$\ell_c(\boldsymbol{\beta}, \mathbf{b}_{11}, ..., \mathbf{b}_{1G}) \propto \frac{1}{2} \sum_{i=1}^{N} \sum_{k=1}^{G} \sum_{t=1}^{T_i} w_{ik}(1 - d_{it})v_{it}^{-1}(\tilde{y}_{it} - \mathbf{x}_{it}'\boldsymbol{\beta} - \mathbf{z}_{it}'\mathbf{b}_{1k})^2 \qquad (3.6)$$

$$- \theta \sum_{i=1}^{N} \sum_{k=1}^{G} \sum_{t=1}^{T_i} w_{ik}(1 - d_{it})(\tilde{y}_{it} - \mathbf{x}_{it}'\boldsymbol{\beta} - \mathbf{z}_{it}'\mathbf{b}_{1k}). \qquad (3.7)$$

In the E-step, the conditional expectation of the complete log-likelihood function, given the observed data and the current parameter estimates at the rth iteration, $\hat{\Phi}_\tau^{(r)}$, is given by:

$$\mathbb{E}[\ell_c(\boldsymbol{\beta}, \mathbf{b}_{11}, ..., \mathbf{b}_{1G}) \mid y_{it}, s_{it}, \mathbf{x}_{it}, \hat{\Phi}_\tau^{(r)}] \tag{3.8}$$

$$\propto \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^G \sum_{t=1}^{T_i} \hat{w}_{ik}^{(r+1)}(1 - d_{it})\hat{v}_{it}^{(r+1)}(\tilde{y}_{it} - \mathbf{x}_{it}'\boldsymbol{\beta} - \mathbf{z}_{it}'\mathbf{b}_{1k})^2 \tag{3.9}$$

$$- \theta \sum_{i=1}^N \sum_{k=1}^G \sum_{t=1}^{T_i} \hat{w}_{ik}^{(r+1)}(1 - d_{it})(\tilde{y}_{it} - \mathbf{x}_{it}'\boldsymbol{\beta} - \mathbf{z}_{it}'\mathbf{b}_{1k}), \tag{3.10}$$

where $\hat{w}_{ik}^{(r+1)}$ has been defined in (3.2) and $\hat{v}_{it}^{(r+1)} = \mathbb{E}[v_{it}^{-1} \mid y_{it}, s_{it}, \mathbf{x}_{it}, \hat{\Phi}_\tau^{(r)}]$. To compute $\hat{v}_{it}^{(r+1)}$, we can exploit the moment properties of the GIG distribution in (3.4). Hence, we have that:

$$\hat{v}_{it}^{(r+1)} = \mathbb{E}[v_{it}^{-1} \mid y_{it}, s_{it}, \mathbf{x}_{it}, \hat{\Phi}_\tau^{(r)}] = \frac{\sqrt{\theta^2 + 2\rho^2}}{\mid \tilde{y}_{it} - \mathbf{x}_{it}'\hat{\boldsymbol{\beta}}^{(r)} - \mathbf{z}_{it}'\hat{\mathbf{b}}_{1k}^{(r)} \mid}. \tag{3.11}$$

In the M-step, we determine the update expressions by setting to zero the derivative of (3.9)-(3.10) with respect to $\boldsymbol{\beta}$ and $\mathbf{b}_{11}, ..., \mathbf{b}_{1G}$ and solve the corresponding M-step equations. In conclusion, we obtain the following update expressions:

$$\hat{\boldsymbol{\beta}}^{(r+1)} = \left(\sum_{i=1}^N \sum_{k=1}^G \sum_{t=1}^{T_i} \hat{w}_{ik}^{(r+1)}(1 - d_{it})\hat{v}_{it}^{(r+1)}\mathbf{x}_{it}\mathbf{x}_{it}' \right)^{-1} \left(\sum_{i=1}^N \sum_{k=1}^G \sum_{t=1}^{T_i} \hat{w}_{ik}^{(r+1)}(1 - d_{it})\left(\hat{v}_{it}^{(r+1)}\mathbf{x}_{it}\left(\tilde{y}_{it} - \mathbf{z}_{it}'\hat{\mathbf{b}}_{1k}^{(r)} \right) - \theta\mathbf{x}_{it} \right) \right) \tag{3.12}$$

and

$$\hat{\mathbf{b}}_{1k}^{(r+1)} = \left(\sum_{i=1}^N \sum_{t=1}^{T_i} \hat{w}_{ik}^{(r+1)}(1 - d_{it})\hat{v}_{it}^{(r+1)}\mathbf{z}_{it}\mathbf{z}_{it}' \right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^{T_i} \hat{w}_{ik}^{(r+1)}(1 - d_{it})\left(\hat{v}_{it}^{(r+1)}\mathbf{z}_{it}\left(\tilde{y}_{it} - \mathbf{x}_{it}'\hat{\boldsymbol{\beta}}^{(r)} \right) - \theta\mathbf{z}_{it} \right) \right). \tag{3.13}$$

Equations (3.12) and (3.13) essentially, are equivalent to modified weighted least square estimator expressions where the quantile level τ is used to modify both the weights, $\hat{w}_{ik}^{(r)}\hat{v}_{it}^{(r)}$, and the response variable through $\hat{v}_{it}^{(r)}$ and θ.

3.2 Variable selection and the penalized EM algorithm

When dealing with high-dimensional problems, it may be of interest to reduce the set of explanatory variables using penalized regression methods which allow for sparse modelling and enhance model interpretability. In this section, we introduce a penalized version of the EM algorithm described in Section 3. In particular, we use the PEM algorithm, originally proposed by Green (1990), which leaves the E-step unchanged while it modifies the M-step by introducing a penalty function to achieve simultaneous shrinkage and/or variable selection. Here we consider the LASSO penalty term put forward by Tibshirani (1996) to shrink the coefficients of the positive outcomes, that is, the vector $\boldsymbol{\beta}$, and simultaneously select a smaller subset of variables that exhibits the strongest effects. For a chosen quantile level τ and number of latent classes G, the penalized log-likelihood for the complete data has the following form:

$$\ell_{pen}(\Phi_\tau|\lambda_\tau) = \ell_c(\Phi_\tau) - \lambda_\tau J(\boldsymbol{\beta}), \qquad (3.14)$$

where $\ell_c(\Phi_\tau)$ has been defined in (3.1), $J(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$ is the convex LASSO penalty function and λ_τ is a tuning parameter that regulates the strength of the penalization assigned to the coefficients of the model. For a given value of τ, G and λ_τ, at the $(r + 1)$th iteration the PEM considers the following step:

$$\hat{\Phi}_\tau^{(r+1)} = \underset{\Phi_\tau}{\operatorname{argmax}} \left\{ \mathcal{Q}(\Phi_\tau|\hat{\Phi}_\tau^{(r)}) - \lambda_\tau J(\boldsymbol{\beta}) \right\}, \qquad (3.15)$$

where $\mathcal{Q}(\Phi | \hat{\Phi}_\tau^{(r)}) = \mathbb{E}[\ell_c(\Phi_\tau) | y_{it}, \mathbf{s}_{it}, \mathbf{x}_{it}, \hat{\Phi}_\tau^{(r)}]$. The optimal value of λ_τ is selected via ten fold cross-validation which allows us to consider λ_τ as a data-driven parameter. In particular, following Friedman et al. (2010) and Städler et al. (2010) we compute a sequence of K values of λ_τ equispaced in the log scale. Then, for each value of λ_τ, the PEM method iteratively alternates between the algorithm of Rigby and Stasinopoulos (2005) to obtain the update estimates of $\boldsymbol{\gamma}$ and the optimization algorithm of Koenker et al. (2014) to compute the penalized estimates of $\boldsymbol{\beta}$.

4 Simulation study

In this section, we conduct a simulation study to evaluate the finite sample properties of the proposed two-part finite mixture quantile regression model. This simulation exercise addresses the following issues. First, we consider different distributional choices for the error term to investigate the behaviour of the model in the presence of non-Gaussian errors. Second, we evaluate the robustness of the nonparametric approach to non-Gaussian distributions for the subject-specific, random coefficients. Finally, we assess the performance of the PEM algorithm proposed in Section 3.2.

We consider two sample sizes $N = (100, 200)$ and two longitudinal lengths $T_i = T = (5, 10)$, for all $i = 1, \ldots, N$, and three explanatory variables $x_{1it} \sim \mathcal{N}(0, 1)$,

$x_{2it} \sim \mathcal{N}(0, 1)$ and $x_{3it} \sim \text{Ber}(0.5)$. The observations are generated from the following two-part model:

$$\text{logit}(p_{it}) = \gamma_1 x_{1it} + \gamma_2 x_{2it} + \gamma_3 x_{3it} + b_{i0}, \quad i = 1, ..., N, \quad t = 1, ..., T \tag{4.1}$$

$$\log(y_{it}) \mid y_{it} > 0 = \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + b_{i1} + \epsilon_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T, \tag{4.2}$$

where $\mathbf{b}_i = (b_{i0}, b_{i1})$ are possibly correlated random intercepts that capture individual departures from the fixed effects $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$, respectively. For each sample size, two different simulation scenarios are considered:

1. $(\mathcal{N} - \mathcal{N})$: $\epsilon_{it} \sim \mathcal{N}(0, 1)$ and $\mathbf{b}_i \sim \mathcal{N}_2(0, \Sigma)$,
2. $(\chi^2 - \mathcal{T})$: $\epsilon_{it} \sim \chi^2(2)$ to allow for skewed data and $\mathbf{b}_i \sim \mathcal{T}_2(0, \Sigma)$ with 3 degrees of freedom.

We assumed the following true values for the model parameters:

$$\boldsymbol{\gamma} = (-1.5, 2.5, 1), \quad \boldsymbol{\beta} = (2, 0.5, -1) \quad \text{and} \quad \Sigma = \begin{pmatrix} 0.25 & 0.0625 \\ 0.0625 & 0.25 \end{pmatrix}. \tag{4.3}$$

Because the number of components is unknown a-priori, following Pan and Shen (2007); Städler et al. (2010) and Sottile et al. (2020), to account for shrinkage we select G by employing the modified Bayesian Information Criterion (BIC) (Schwarz et al., 1978):

$$\text{BIC} = -2\log(L) + \nu_f \log(N), \tag{4.4}$$

where ν_f denotes the number of free model parameters and N is the number of individuals. In our proposed model, $\nu_f = \dim(\boldsymbol{\gamma}) + \dim(\boldsymbol{\beta}) + 2G + G - q$, because we have five sets of parameters, $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, σ_τ, the random intercepts $\mathbf{b}_g$'s, π_g's under the constraint $\sum_{k=1}^{G} \pi_k = 1$ and where q is the number of nonzero components in the penalized ML estimate of $\boldsymbol{\beta}$. Naturally, in the unpenalized two-part finite mixture quantile regression model, $q = 0$.

We analyse three quantile levels $\tau = (0.1, 0.5, 0.9)$ for a varying number of mixture components $G = (2, ..., 6)$, retaining the solution corresponding to the lowest BIC value. For each model, we carry out $B = 100$ Monte Carlo replications and report the following indicators. The Average Relative Bias (ARB) defined as:

$$ARB(\hat{\theta}_\tau) = \frac{1}{B} \sum_{b=1}^{B} \frac{(\hat{\theta}_\tau^{(b)} - \theta_\tau)}{\theta_\tau} \times 100, \tag{4.5}$$

where $\hat{\theta}_\tau^{(b)}$ is the estimated parameter at quantile level τ for the bth replication and θ_τ is the corresponding "true" value of the parameter. Second, the Root Mean Square

Table 1 ARB and RMSE (in brackets) of point estimates for γ and β

Scenario	$(\mathcal{N}-\mathcal{N})$			$(\chi^2-\mathcal{T})$		
τth quantile	0.1	0.5	0.9	0.1	0.5	0.9
Panel A: $N=100$, $T=5$						
γ_1	−1.975 (0.136)	1.577 (0.138)	−2.236 (0.139)	−2.642 (0.190)	−2.394 (0.189)	3.645 (0.181)
γ_2	−1.276 (0.196)	−1.500 (0.194)	−1.831 (0.190)	2.618 (0.260)	−1.850 (0.263)	−4.015 (0.260)
γ_3	2.298 (0.248)	1.024 (0.248)	1.184 (0.244)	−3.972 (0.246)	−2.977 (0.253)	4.472 (0.237)
β_1	−0.783 (0.124)	0.615 (0.104)	−0.291 (0.116)	0.306 (0.084)	−1.684 (0.137)	−1.711 (0.225)
β_2	1.283 (0.145)	−0.314 (0.114)	−4.235 (0.136)	−0.518 (0.114)	0.693 (0.146)	4.055 (0.282)
β_3	−2.413 (0.218)	−1.197 (0.155)	−0.347 (0.202)	2.383 (0.181)	2.046 (0.237)	−5.281 (0.602)
Panel B: $N=100$, $T=10$						
γ_1	1.278 (0.114)	−1.474 (0.118)	−1.075 (0.112)	−2.509 (0.135)	−2.876 (0.129)	−3.315 (0.129)
γ_2	−0.722 (0.147)	−0.612 (0.154)	−0.751 (0.145)	2.696 (0.170)	−2.385 (0.153)	−3.518 (0.154)
γ_3	0.337 (0.175)	0.090 (0.179)	0.305 (0.165)	−1.509 (0.171)	−1.979 (0.176)	3.825 (0.178)
β_1	−0.346 (0.078)	0.206 (0.069)	−0.196 (0.089)	0.620 (0.056)	−0.193 (0.092)	1.024 (0.224)
β_2	−2.181 (0.099)	−0.491 (0.074)	0.968 (0.107)	−0.900 (0.076)	−0.978 (0.101)	−2.892 (0.247)
β_3	−1.975 (0.133)	−0.663 (0.120)	−1.451 (0.171)	2.623 (0.114)	−2.600 (0.139)	−2.197 (0.431)
Panel C: $N=200$, $T=5$						
γ_1	−1.265 (0.109)	−1.872 (0.113)	−1.581 (0.108)	3.170 (0.138)	3.415 (0.141)	−4.049 (0.130)
γ_2	−0.758 (0.133)	1.285 (0.137)	−1.120 (0.136)	2.757 (0.201)	−3.305 (0.180)	3.366 (0.167)
γ_3	−1.140 (0.178)	−1.573 (0.180)	1.642 (0.176)	−2.211 (0.171)	1.194 (0.173)	−4.450 (0.178)
β_1	0.701 (0.082)	0.410 (0.064)	1.552 (0.084)	0.229 (0.076)	−0.079 (0.087)	1.314 (0.198)
β_2	2.834 (0.104)	1.444 (0.092)	−0.809 (0.108)	−0.569 (0.084)	−0.524 (0.097)	−3.741 (0.255)
β_3	1.350 (0.150)	−0.540 (0.107)	−0.589 (0.160)	0.428 (0.121)	0.010 (0.138)	2.374 (0.340)
Panel D: $N=200$, $T=10$						
γ_1	−0.480 (0.076)	−0.754 (0.078)	−0.715 (0.075)	−0.589 (0.085)	−1.063 (0.094)	−3.145 (0.095)
γ_2	0.373 (0.107)	1.129 (0.105)	1.277 (0.108)	2.318 (0.112)	−1.486 (0.111)	2.374 (0.131)
γ_3	−1.065 (0.115)	0.560 (0.116)	−0.901 (0.116)	−2.235 (0.127)	2.105 (0.128)	−1.571 (0.133)
β_1	0.264 (0.051)	0.283 (0.037)	0.033 (0.051)	−0.288 (0.043)	−0.344 (0.062)	−0.319 (0.159)
β_2	−0.631 (0.072)	−0.118 (0.056)	−1.547 (0.073)	−0.301 (0.048)	0.645 (0.069)	1.166 (0.187)
β_3	−0.986 (0.105)	−0.335 (0.061)	0.542 (0.101)	0.361 (0.074)	1.714 (0.098)	−0.065 (0.288)

Table 2 Median number of mixture components, iterations and CPU Time (in seconds)

Scenario	$(\mathcal{N}-\mathcal{N})$			$(\chi^2-\mathcal{T})$			$(\mathcal{N}-\mathcal{N})$			$(\chi^2-\mathcal{T})$		
τth quantile	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
Panel A:			$N=100$, $T=5$						$N=100$, $T=10$			
G	3	2	2	3	2	2	3	2	2	3	2	4
Iterations	17	9	14	16	13	19	15	12	15	16	14	17
CPU time	0.997	0.344	0.721	0.695	0.378	1.588	1.490	0.820	1.563	1.730	0.842	2.718
Panel B:			$N=200$, $T=5$						$N=200$, $T=10$			
G	3	2	2	3	2	4	2	2	3	4	3	3
Iterations	19	13	17	18	14	21	17	12	16	21	17	18
CPU time	2.007	0.817	1.793	1.766	0.736	3.803	3.520	1.403	3.555	5.166	2.302	7.036

Error (RMSE) of model parameters averaged across the B simulations:

$$RMSE(\hat{\theta}_\tau) = \sqrt{\frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}_\tau^{(b)} - \theta_\tau)^2}. \tag{4.6}$$

Tables 1 and 2 report the results for the fixed parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ of the unpenalized two-part finite mixture quantile regression model. As can be noted in Table 1, the proposed model under the $(\mathcal{N} - \mathcal{N})$ and $(\chi^2 - \mathcal{T})$ scenarios is able to recover the true parameters for both the binary and positive decision mechanisms. As expected, both the ARB and the RMSE tend to decrease with increasing sample sizes and number of measurement occasions. Also, the parameters of the continuous process are better estimated than those of the logistic model (Maruotti et al., 2016). Concerning the point estimates of the quantile regressions, the bias effect is quite small when we analyse the median levels (see columns 2 and 5). When the quantile levels become more extreme (see columns 1, 3, 4 and 6), the ARB and RMSE increase because of the reduced amount of information in the tails of the distribution, but remain reasonably small. We can identify a similar pattern in the worst scenario $(\chi^2 - \mathcal{T})$, where the skewed errors and the heavier tails of the Student t distribution contribute to increase the ARB and RMSE, especially at the 10th and 90th percentiles. The selected number of mixture components G approximates the random effects distribution in a parsimonious manner in the $(\mathcal{N} - \mathcal{N})$ scenario as $G \leq 3$ in 77% of simulated examples; on the other hand, more support point are needed in the second setting. To evaluate the speed of convergence of the EM algorithm, we report the median number of iterations and CPU Time (in seconds) required by the implemented R code using an Intel Xeon E5-2609 2.40GHz processor. The proposed model reveals decent computational performance as running times range from 0.344 seconds in simple scenarios up to 7.036 seconds for larger sample size.

To evaluate the performance of the PEM algorithm proposed in Section 3.2, we repeat the simulation experiments where we set $\boldsymbol{\beta} = (2, 0, 0)$. In the penalized model, in addition to G, we also have to choose an appropriate value for the tuning parameter λ_τ. Specifically, we construct a sequence of $K = 100$ values of λ_τ as described in Section 3.2. Then, for fixed τ and G, we use a ten fold cross-validation to select the tuning parameter using the PEM and repeat this procedure for a varying number of mixture components $G = (2, ..., 6)$. Eventually, we select the number of mixture components corresponding to the lowest value of the modified BIC in (4.4).

Table 3 reports the True Positive Rate (TPR) for each of the true coefficients initially set equal to zero, the average Number of Covariates (NC) being selected and the median number of chosen mixture components G. We compute the TPR as the ratio between the number of simulations that correctly identify the parameter as a zero value, over the total number of simulations in order to produce a measure of how sensitive a method is at discovering nonzero entries. As one can see, the PEM method performs well, with an average TPR of more than the 85% across the two simulation scenarios. As the sample size increases, the TPR increases and the average number of selected predictors approaches 1. Finally, the number of mixture

Table 3 Performance of the PEM algorithm: True Positive Rate. In each panel, the last two rows show the average number of covariates being selected and the median number of mixture components

Scenario	$(\mathcal{N} - \mathcal{N})$			$(\chi^2 - \mathcal{T})$			$(\mathcal{N} - \mathcal{N})$			$(\chi^2 - \mathcal{T})$		
τth quantile	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
Panel A:			$N = 100, T = 5$						$N = 100, T = 10$			
β_2	0.80	0.94	0.84	0.78	0.78	0.79	0.93	0.94	0.92	0.92	0.94	0.88
β_3	0.71	0.83	0.82	0.69	0.80	0.78	0.78	0.97	0.93	0.79	0.81	0.89
NC	1.32	1.22	1.26	1.46	1.38	1.34	1.22	1.17	1.18	1.25	1.18	1.23
G	3	2	2	3	2	3	3	2	2	3	2	2
Panel B:			$N = 200, T = 5$						$N = 200, T = 10$			
β_2	0.94	0.94	0.94	0.92	0.82	0.93	0.96	0.98	0.99	0.92	0.92	0.90
β_3	0.85	0.96	0.95	0.84	0.89	0.83	0.88	0.99	0.96	0.87	0.90	0.91
NC	1.18	1.12	1.17	1.24	1.28	1.23	1.12	1.01	1.02	1.15	1.04	1.09
G	3	2	2	3	2	3	3	2	2	4	3	3

components selected by the penalized two-part finite mixture model is consistent with the one chosen by the unpenalized two-part model in Table 1, regardless of the distributional assumptions on the random intercepts or on the error terms.

5 Application

5.1 Data description

As stated in the Introduction, in this section we present the application of the proposed methodology to the well-known RHIE dataset. These data have already been discussed by Deb and Trivedi (2002); Duan et al., (1983) and Manning et al. (1987). The RHIE is the most important health insurance study ever conducted to assess how medical care costs affect a patient's use of health services and quality and it is widely regarded as the basis of the most reliable estimates of price sensitivity of demand for medical services. The experiment, conducted by the RAND Corporation from 1974 to 1982, collects data from about 8 000 enrollees in 2 823 families, from six sites across the US. Each family was enrolled in one of 14 different HIS insurance plans for either 3 or 5 years.

Our aim is to understand how available covariates influence healthcare decisions spending in US families at different quantile levels of interest. We consider one measure of utilization: the total spending on health services (MED) defined as the sum of outpatient, inpatient, drugs, supplies and psychotherapy expenses expressed in U.S. dollars. The considered covariates are the same of the work of Deb and Trivedi (2002): they include personal characteristics such as sex (FEMALE), age (XAGE), race (BLACK) and education level (EDUCDEC). There are also socio-economic variables of the household such as income (LINC), number of components (LFAM)

Table 4 Summary statistics of healthcare expenditures

Variable	Mean	SD	Skewness	Kurtosis	No Exp. (%)	Maximum	τ-th quantile				
							0.1	0.25	0.5	0.75	0.9
MED	171.568	698.201	20.189	734.007	22.055	39182.016	0.000	5.503	35.378	104.541	341.201

Table 5 Covariates description and summary statistics. [†]indicates that the baseline individual is in excellent self-rated health

Variable	Description	Mean	SD
LOGC	log(coinsurance + 1), $0 \leq$ coinsurance ≤ 100	2.384	2.042
LFAM	log(family size)	1.248	0.539
LINC	log(family income)	8.708	1.228
XAGE	Age in years	25.718	16.768
FEMALE	If person is female: 1	0.517	0.500
CHILD	If age is less than 18: 1	0.401	0.490
FEMCHILD	FEMALE * CHILD	0.194	0.395
BLACK	If race of household head is black: 1	0.184	0.387
EDUCDEC	Education of the household head in years	11.966	2.806
PHYSLM	If the person has a physical limitation: 1	0.118	0.323
DISEA	Index of chronic diseases	11.244	6.742
HLTHG	If self-rated health is good: 1[†]	0.362	0.481
HLTHF	If self-rated health is fair: 1[†]	0.077	0.267
HLTHP	If self-rated health is poor: 1[†]	0.015	0.121
MHI	Mental Health Index	76.554	12.503

and the presence of a person aged 18 or less (CHILD). An interaction term between the gender and the presence of a child (FEMCHILD) is also included. Quantitative indicators of health condition are measured via an index of chronic conditions (DISEA) and through the existence of a physical limitation (PHYSLM) while the binary coding of self-rated health status (HLTHG, HLTHF, HLTHP) controls for variations in perceived healthcare conditions. The summary statistics are reported in Tables 4 and 5. By looking at Table 4, the proportion of zero expenditures is significant: 22% did not have any medical expenditure. Meanwhile, the right tail of the distribution is very long with the maximum expense being 39182.02 US dollars which indicates the presence of overdispersion and potential outliers in the data. In addition, the dependent variable is severely skewed and presents a high kurtosis. From a graphical point of view, Figure 1 shows that the data is characterized by a substantial zero-mass health expenditure (left graph): zero expenditure indicates no utilization, and it may reflect population' reluctance to spend on health care treatments. By looking at the right plot, also the logarithmic transformation of positive values does not follow a Gaussian distribution. Therefore, a two-part quantile regression model would be appropriate for these data. In particular, the binary decision process would estimate the association between the covariates and the probability of having any health care expenditure. Through the positive part of the model instead, our goal is to see whether the covariates have a different impact on health care expenditures at
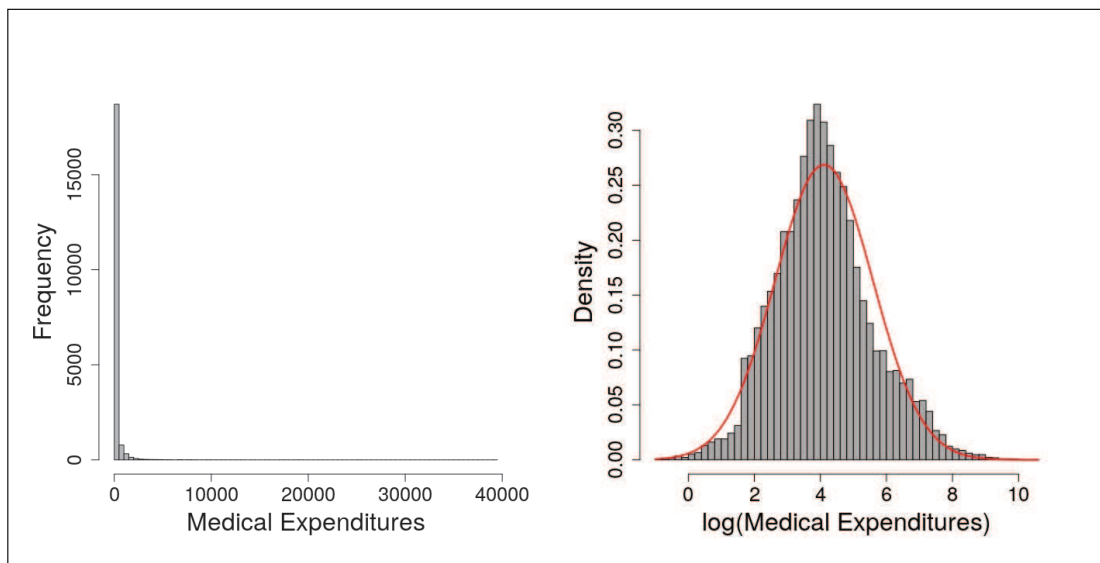
Figure 1 Medical expenditure distribution: Marginal distribution (left) and distribution of the logarithm of positive values (right)

different quantiles. Indeed, by doing so, we can single out the impact of health determinants across extreme quantiles and non extreme quantiles reflecting the association with low-intensity primary health care services, at the 10th and 25th percentiles, and that with high-intensity such as the highly intensive and expensive technology care, at the 75th and 90th percentiles.

5.2 Marginal inferences from the two-part mixture quantile regression model

In this section, we present the results of the application to the RHIE data. We estimated the proposed LASSO penalized two-part finite mixture quantile regression at different quantile levels of interest $\tau = (0.1, 0.25, 0.5, 0.75, 0.9)$, for a varying number of mixture components $G = (1, \ldots, 6)$. We considered a sequence of $K = 100$ values of λ_τ for the ten fold cross-validation procedure in Section 3.2 and select the number of mixture components according to the BIC in (4.4). We used the same covariates for both the binary process and the positive part of the model and we assumed that random intercepts are sufficient to capture heterogeneity between subjects even tough the generalization to multiple random effects is computationally inexpensive and straightforward thanks to the discrete mixture structure. Standard errors are obtained using a parametric bootstrap approach: we refitted the model to 250 bootstrap samples simulated from the estimated model parameters, and approximated the standard error of each parameter with its corresponding standard

deviation computed on bootstrap samples. The implemented R source codes and data are available at http://www.statmod.org/smij/archive.html.

Table 6 reports the estimated penalized model coefficients and standard errors (in parentheses) for the binary (Panel A) and positive (Panel B) part of the model. Parameter estimates are displayed in boldface when significant at the standard 5% level. Table 7 illustrates the estimated locations, masses of the bivariate discrete distribution of the random effects (Panel A) and the model fit summary (Panel B) at each quantile level, respectively.

Consistently with the findings of Deb and Trivedi (2002); Duan et al., (1983) and Bago d'Uva (2006), the selected number of mixture components G is two for all investigated quantiles. Based on Table 7 (Panel A), the two-part model demarcates two distinct subpopulations: a group of reluctant users who rarely accesses health services and a group of users that uses health services with high frequency. This claim is confirmed by the estimated mixing probabilities (π_1, π_2) and locations $\mathbf{b}_1, \mathbf{b}_2$ which can be interpreted as high and low users. Group separation is all the more apparent as the quantile level increases with estimated masses $(0.328, 0.672)$ at the 90th quantile.

By looking at the binary process estimates in Table 6 (Panel A), one can see that all coefficients are statistically different from zero at the 5% significance level except LFAM, MHI and HLTHG and HLTHF at low and high quantile levels, respectively. The marginal inferences from the binary part of the two-part model suggest that the most important determinants of people's willingness to spend on health services are related to socio-cultural and economic factors: income, age, sex, race ethnicity and education level have the expected signs and significance. The presence of a coinsurance health plan acts as a protection factor reducing the probability of consuming health services. Ageing, presumably, involves complications such as heart diseases, cancer, diabetes and many others that require (obligatory) hospitalization. The analysis points out also profound economic and ethnicity inequalities.

It is worth noting that the coefficient estimates vary as the quantile level varies, for example, they generally show an upward or downward trend as the quantile increases. This may be due to the presence of correlation between the binary and the positive processes which is channelled via the latent structure defined by the discrete bivariate random effects.

Moving onto the analysis of the positive outcomes (Table 6, Panel B), our results are generally consistent with those of Deb and Trivedi (2002). The first thing one can notice is that not all variables have the same effect on healthcare consumption, that is, they change in sign and magnitude and some of them have been set to zero at specific quantile levels. This highlights the importance of considering a quantile regression approach because such effects could not be detected by the classical mean regression. As previously described, the presence of health insurance seems to act as a safety ledge: insured people spend less. This may reflect the fact that health care is largely provided by private hospitals and clinics. The results indicate also that individuals in larger families are more likely to be in the low-use group especially up to the 75th quantile. As expected, the economic situation appears to be one of the main barriers that influence healthcare access for low income households. In relation

Table 6 Penalized two-part finite mixture of quantile regressions coefficient estimates. Panel A refers to the binary part while Panel B refers to the positive part of the model

τth quantile	0.1	0.25	0.5	0.75	0.9
Panel A: Binary process					
LOGC	**0.413** (0.022)	**0.414** (0.025)	**0.406** (0.026)	**0.379** (0.022)	**0.377** (0.019)
LFAM	−0.039 (0.025)	−0.016 (0.026)	−0.021 (0.027)	0.014 (0.024)	−0.017 (0.022)
LINC	**−0.190** (0.022)	**−0.196** (0.024)	**−0.186** (0.025)	**−0.175** (0.022)	**−0.097** (0.019)
XAGE	**−0.101** (0.037)	−0.066 (0.039)	−0.059 (0.041)	**−0.082** (0.038)	**−0.120** (0.034)
FEMALE	**−0.841** (0.061)	**−0.974** (0.067)	**−0.972** (0.067)	**−0.860** (0.059)	**−0.689** (0.054)
CHILD	**−0.450** (0.088)	**−0.465** (0.094)	**−0.426** (0.096)	**−0.511** (0.087)	**−0.432** (0.080)
FEMCHILD	**0.882** (0.087)	**0.919** (0.095)	**0.921** (0.094)	**0.873** (0.088)	**0.695** (0.074)
BLACK	**1.422** (0.058)	**1.351** (0.059)	**1.443** (0.066)	**1.229** (0.049)	**1.134** (0.045)
EDUCDEC	**−0.160** (0.023)	**−0.168** (0.024)	**−0.192** (0.025)	**−0.151** (0.021)	**−0.163** (0.020)
PHYSLM	**−0.543** (0.078)	**−0.725** (0.082)	**−0.720** (0.088)	**−0.590** (0.078)	**−0.500** (0.073)
DISEA	**−0.287** (0.028)	**−0.278** (0.028)	**−0.276** (0.029)	**−0.237** (0.027)	**−0.246** (0.024)
HLTHG	**0.123** (0.048)	0.066 (0.053)	0.020 (0.054)	**0.124** (0.047)	**0.117** (0.044)
HLTHF	**−0.207** (0.091)	−0.178 (0.098)	**−0.272** (0.108)	**−0.271** (0.089)	−0.131 (0.083)
HLTHP	**−1.026** (0.246)	**−0.996** (0.258)	**−1.011** (0.259)	**−0.903** (0.231)	**−0.635** (0.209)
MHI	−0.016 (0.023)	−0.037 (0.025)	**−0.057** (0.026)	−0.026 (0.022)	−0.030 (0.021)
Panel B: Positive process					
LOGC	**−0.196** (0.005)	**−0.192** (0.007)	**−0.142** (0.009)	**−0.091** (0.008)	**−0.081** (0.005)
LFAM	**−0.075** (0.006)	**−0.098** (0.009)	**−0.057** (0.010)	**−0.106** (0.009)	–
LINC	**0.103** (0.008)	**0.088** (0.010)	**0.068** (0.012)	**0.122** (0.011)	–
XAGE	**0.219** (0.010)	**0.194** (0.012)	**0.165** (0.015)	**0.154** (0.013)	**0.359** (0.006)
FEMALE	**0.287** (0.014)	**0.351** (0.021)	**0.270** (0.027)	**0.400** (0.023)	–
CHILD	**0.096** (0.026)	–	**−0.204** (0.036)	**−0.097** (0.027)	−0.001 (0.003)
FEMCHILD	**−0.299** (0.025)	**−0.259** (0.032)	**−0.230** (0.041)	**−0.461** (0.034)	**−0.006** (0.001)
BLACK	**−0.547** (0.018)	**−0.379** (0.023)	**−0.436** (0.028)	**−0.243** (0.024)	–
EDUCDEC	**0.054** (0.006)	**0.039** (0.008)	**0.044** (0.010)	−0.017 (0.009)	**0.012** (0.005)
PHYSLM	**0.189** (0.019)	**0.404** (0.025)	**0.371** (0.028)	**0.437** (0.027)	**0.281** (0.017)
DISEA	**0.186** (0.007)	**0.158** (0.008)	**0.139** (0.011)	**0.073** (0.009)	**0.072** (0.005)
HLTHG	0.002 (0.012)	**0.059** (0.017)	**0.100** (0.021)	0.011 (0.017)	–
HLTHF	**0.294** (0.025)	**0.209** (0.034)	**0.358** (0.039)	**0.411** (0.035)	–
HLTHP	**0.801** (0.054)	**0.588** (0.066)	**0.613** (0.079)	**0.700** (0.074)	–
MHI	**−0.039** (0.006)	**−0.043** (0.008)	**−0.029** (0.010)	**−0.060** (0.009)	**−0.053** (0.005)
σ_τ	**0.187** (0.001)	**0.355** (0.003)	**0.471** (0.004)	**0.372** (0.003)	**0.187** (0.001)

to the gender, females have a higher spending pattern up to the 75th percentiles which may be caused by hospitalization being parturition or its complications.

Other covariates also appear to have some explanatory power, notably XAGE, CHILD, FEMCHILD and BLACK. The elderly undergoes frequent checks and require more health services due to ageing phenomena to which they may be exposed to. The interaction term, FEMCHILD, and BLACK are negatively associated with the response indicating that, in the latter case, racial differences in the amount spent on medical care are persistent and confirms that among individuals who spend on care, blacks have lower expenditures than whites. Moreover, highly literate and educated individuals tend to follow a healthy lifestyle, be less reluctant to go through a visit to a physician, and thus spend more to preserve and improve their health. Variables controlling for the health status of the individual such as PHYSLM and DISEA

Table 7 Panel A illustrates the estimated locations and mixing probabilities of the penalized two-part finite mixture of quantile regressions. Panel B reports the optimal values for λ_τ, log-likelihood, number of nonzero model parameters and penalized likelihood criteria (AIC, BIC)

τth quantile	0.1	0.25	0.5	0.75	0.9
Panel A: Finite mixture estimates					
b_{01}	**−2.671** (0.072)	**−2.781** (0.081)	**−2.897** (0.086)	**−2.803** (0.084)	**−2.193** (0.066)
b_{02}	**−0.476** (0.052)	**−0.326** (0.056)	**−0.299** (0.057)	**−0.686** (0.053)	**−0.904** (0.049)
b_{11}	**3.107** (0.015)	**3.590** (0.020)	**4.353** (0.025)	**5.557** (0.020)	**6.771** (0.008)
b_{12}	**1.700** (0.016)	**2.362** (0.020)	**3.285** (0.026)	**4.044** (0.018)	**4.705** (0.007)
π_1	0.495 (0.006)	0.484 (0.008)	0.487 (0.009)	**0.337** (0.008)	**0.328** (0.007)
π_2	0.505 (0.006)	0.516 (0.008)	0.513 (0.009)	**0.663** (0.008)	**0.672** (0.007)
Panel B: Model fit summary					
λ_τ	0.013	0.018	0.018	0.032	0.566
$\log(L)$	−10349.09	−7446.91	−6742.39	−9394.92	−12392.29
# par	36	35	36	36	29
AIC	20770.18	14963.82	13556.77	18861.84	24844.59
BIC	21055.04	15240.77	13841.63	19146.70	25081.97

Table 8 LQMM coefficient estimates for the positive outcomes for the investigated quantile levels. Standard errors are in parentheses. Parameter estimates are displayed in boldface when significant at the standard 5% level

τth quantile	0.1	0.25	0.5	0.75	0.9
LOGC	**−0.202** (0.026)	**−0.159** (0.017)	**−0.147** (0.017)	**−0.100** (0.025)	**−0.102** (0.026)
LFAM	**−0.130** (0.023)	**−0.108** (0.018)	**−0.116** (0.019)	−0.039 (0.030)	**−0.061** (0.025)
LINC	**0.062** (0.028)	**0.066** (0.022)	**0.055** (0.018)	**0.079** (0.035)	0.055 (0.034)
XAGE	**0.232** (0.028)	**0.169** (0.023)	**0.173** (0.024)	**0.115** (0.029)	**0.110** (0.035)
FEMALE	**0.328** (0.036)	**0.244** (0.029)	**0.280** (0.031)	**0.377** (0.041)	**0.368** (0.039)
CHILD	**0.064** (0.033)	−0.020 (0.030)	**−0.083** (0.036)	**−0.279** (0.040)	**−0.242** (0.042)
FEMCHILD	**−0.254** (0.035)	**−0.281** (0.033)	**−0.266** (0.037)	**−0.464** (0.042)	**−0.478** (0.044)
BLACK	**−0.192** (0.035)	**−0.359** (0.039)	**−0.315** (0.042)	**−0.290** (0.050)	**−0.305** (0.055)
EDUCDEC	0.056 (0.046)	**0.051** (0.016)	0.023 (0.017)	−0.016 (0.028)	−0.011 (0.027)
PHYSLM	**0.326** (0.037)	**0.274** (0.042)	**0.342** (0.046)	**0.428** (0.049)	**0.450** (0.045)
DISEA	**0.103** (0.030)	**0.150** (0.022)	**0.115** (0.020)	**0.057** (0.023)	**0.074** (0.024)
HLTHG	0.024 (0.044)	**0.095** (0.030)	**0.073** (0.033)	**0.087** (0.042)	**0.148** (0.044)
HLTHF	**0.268** (0.038)	**0.357** (0.042)	**0.327** (0.049)	**0.332** (0.043)	**0.272** (0.047)
HLTHP	**0.573** (0.026)	**0.631** (0.046)	**0.657** (0.040)	**0.866** (0.038)	**0.698** (0.037)
MHI	−0.013 (0.024)	**−0.072** (0.018)	**−0.037** (0.015)	**−0.080** (0.025)	**−0.077** (0.023)
Intercept	**2.497** (0.036)	**3.153** (0.024)	**3.845** (0.027)	**4.637** (0.034)	**4.833** (0.030)
σ_τ	**0.155** (0.001)	**0.321** (0.003)	**0.430** (0.004)	**0.340** (0.003)	**0.291** (0.003)

are positively associated with healthcare expenses: supposedly, disabilities, physical limitations and diagnosis of chronic illnesses require highly intensive, technology care or newer, more costly treatments. Also, self-perceived health condition indicators are all positively associated with health expenses with a much steeper slope at the lower percentiles meanwhile, their impact seems to be negligible above the 75th quantile.

Moreover, it is possible to see that the impact of several variables varies across quantiles: the effect of LOGC and BLACK is not uniform as the quantile increases but is worse on the left tail of the distribution of health expenditures. Also, variables

such as LINC, FEMALE, CHILD and FEMCHILD exhibit nonlinear effects: in particular, CHILD and FEMCHILD have a U-shaped and an inverted U-shape effect, respectively. EDUCDEC changes sign and magnitude across the quantile levels as well. As regards the estimated random intercepts $\mathbf{b}_{1k}$, from Table 7 (Panel A) we notice that the estimates increase with τ and this is consistent with increasing values of healthcare spending. We repeated the analysis using alternative optimization procedures for the PEM algorithm, including the semismooth Newton coordinate descent in Yi and Huang (2017), the Frisch–Newton interior point method described in Portnoy et al. (1997) and a Nelder–Mead optimization, which produce all similar results and support the main findings of the article.

We conclude the analysis comparing our methodology with the Linear Quantile Mixed Model (LQMM) of Geraci and Bottai (2014). In particular, we fit a LQMM only on the positive values of the dependent variable under the assumption that the distribution of the random intercepts is a bivariate Gaussian while ignoring both the correlation between the binary and positive process and the sparsity induced through the LASSO regularization. By looking at the parameter estimates in Table 8, we notice slight differences with respect to the results obtained using our methodology at low quantiles while, as we move from the left to the right tail of the response distribution, there are considerable differences. Such discrepancies in the two models might be traced back to the fact that the distribution of the random effects is not Gaussian. Indeed, the semiparametric mixture approach and the LQMM perform equivalently only if the random coefficients are Gaussianly distributed otherwise, the semiparametric mixture performs better being more flexible and able to accommodate departure from the Gaussianity assumption (see Alfò et al., 2017).

6 Conclusions

This article introduces a two-part finite mixture of quantile regressions for mixed-type outcomes under a longitudinal setting. Random effect coefficients are added in both the binary and the positive decision mechanisms to account for zero inflation and unobserved heterogeneity. Rather than assuming a parametric distribution on the random coefficients distribution, we approximate it using a multivariate discrete variable defined on a finite number of support points. Estimation of the model parameters is based on a suitable likelihood-based EM algorithm. In addition, a LASSO penalized version of the algorithm is proposed as an automatic data-driven procedure to perform variable selection. The application of the proposed method to the RHIE on health behaviours and attitudes shows consistent results with existing studies.

The proposed approach can be further extended to consider time-varying sources of unobserved heterogeneity via individual-specific coefficients evolving according a hidden Markov chain. Second, we may extend the univariate quantile framework to a multivariate quantile regression setting taking into account for the correlation among the marginals of a multivariate response variable (Petrella and Raponi, 2019).

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

References

Aitkin M and Alfò M (1998) Regression models for binary longitudinal responses. *Statistics and Computing*, **8**, 289–307.

——— (2003) Longitudinal analysis of repeated binary data using autoregressive and random effect modelling. *Statistical Modelling*, **3**, 291–303.

Alfò M and Aitkin M (2000) Random coefficient models for binary longitudinal responses with attrition. *Statistics and Computing*, **10**, 279–87.

Alfò M and Maruotti A (2010) Two-part regression models for longitudinal zero-inflated count data. *Canadian Journal of Statistics*, **38**, 197–216.

Alfò M, Salvati N and Ranallli MG (2017) Finite mixtures of quantile and M-quantile regression models. *Statistics and Computing*, **27**, 547–70.

Alhamzawi R, Yu K and Benoit DF (2012) Bayesian adaptive Lasso quantile regression. *Statistical Modelling*, **12**, 279–97.

Bago d'Uva T (2006) Latent class models for utilisation of health care. *Health Economics*, 15, 329–43.

Bassett GW and Chen H-L (2002) Portfolio style: Return-based attribution using quantile regression, pages 293–305. Heidelberg: Physica-Verlag HD.

Basu A and Manning WG (2009) Issues for the next generation of health care cost analyses. *Medical Care*, **47**, S109–14.

Belasco EJ and Ghosh SK (2012) Modelling semi-continuous data using mixture regression models with an application to cattle production yields. *The Journal of Agricultural Science*, **150**, 109–21.

Belotti F, Deb P, Manning WG and Norton EC (2015) Twopm: Two-part models. *The Stata Journal*, **15**, 3–20.

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

Bernardi M, Bottone M and Petrella L (2018) Bayesian quantile regression using the skew exponential power distribution. *Computational Statistics & Data Analysis*, **126**, 92–111.

Biswas J, Ghosh P and Das K (2020) A semi-parametric quantile regression approach to zero-inflated and incomplete longitudinal outcomes. *AStA Advances in Statistical Analysis*, 1–23. doi: 10.1007/s10182-020-00362-9

Cole TJ and Green PJ (1992) Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine*, **11**, 1305–19.

Deb P and Trivedi PK (2002) The structure of demand for health care: Latent class versus two-part models. *Journal of Health Economics*, **21**, 601–25.

Diehr P, Yanez D, Ash A, Hornbrook M and Lin D (1999) Methods for analyzing health

care utilization and costs. *Annual Review of Public Health*, **20**, 125–44.

Duan N, Manning WG, Morris CN and Newhouse JP (1983) A comparison of alternative models for the demand for medical care. *Journal of Business & Economic Statistics*, **1**, 115–26.

Fan J and Lv J (2010) A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, **20**, 101.

Farcomeni A and Viviani S (2015) Longitudinal quantile regression in the presence of informative dropout through longitudinal-survival joint modelling. *Statistics in Medicine*, **34**, 1199–1213.

Farewell V, Long D, Tom B, Yiu S and Su L (2017) Two-part and related regression models for longitudinal data. *Annual Review of Statistics and Its Application*, **4**, 283–315.

Friedman J, Hastie T and Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, **33**, 1.

Geraci M and Bottai M (2014) Linear quantile mixed models. *Statistics and Computing*, **24**, 461–79.

Green PJ (1990) On use of the EM algorithm for penalized likelihood estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, **52**, 443–52.

Grilli L, Rampichini C and Varriale R (2016) Statistical modelling of gained university credits to evaluate the role of pre-enrolment assessment tests: An approach based on quantile regression for counts. *Statistical Modelling*, **16**, 47–66.

Hendricks W and Koenker R (1992) Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association*, **87**, 58–68.

Heras A, Moreno I and Vilar-Zanon JL (2018) An application of two-stage quantile regression to insurance ratemaking. *Scandinavian Actuarial Journal*, **2018**, 753–69.

Iversen T, Aas E, Rosenqvist G, Hakkinen U and on behalf of the EuroHOPE study group (2015) Comparative analysis of treatment costs in EUROHOPE. *Health Economics*, **24**, 5–22.

Koenker R (2005) *Quantile Regression*. Cambridge: Cambridge University Press.

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

Koenker R, Chernozhukov V, He X and Peng L (2017) *Handbook of Quantile Rregression*. Boca Raton, FL: CRC Press.

Koenker R and Mizera I (2014) Convex optimization in R. *Journal of Statistical Software*, **60**, 1–23.

Kozumi H and Kobayashi G (2011) Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, **81**, 1565–78.

Laird N (1978) Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, **73**, 805–11.

Laporta AG, Merlo L and Petrella L (2018) Selection of value at risk models for energy commodities. *Energy Economics*, **74**, 628–43.

Liu L (2009) Joint modelling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data. *Statistics in Medicine*, **28**, 972–86.

Manning WG, Newhouse JP, Duan N, Keeler EB and Leibowitz A (1987) Health insurance and the demand for medical care: Evidence from a randomized experiment. *The American Economic Review*, **77**, 251–77.

Marino MF and Farcomeni A (2015) Linear quantile regression models for longitudinal experiments: an overview. *Metron*, **73**, 229–47.

Marino MF, Tzavidis N and Alfò M (2018) Mixed hidden Markov quantile regression models for longitudinal data with possibly incomplete sequences. *Statistical Methods in Medical Research*, **27**, 2231–46.

Maruotti A and Raponi V (2014) On baseline conditions for zero-inflated longitudinal count data. *Communications in Statistics: Simulation and Computation*, **43**, 743–60.

Maruotti A, Raponi V and Lagona F (2016) Handling endogeneity and nonnegativity in

correlated random effects models: Evidence from ambulatory expenditure. *Biometrical Journal*, **58**, 280–302.

Merlo L, Petrella L and Raponi V (2020) Sectoral decomposition of CO2 world emissions: A joint quantile regression approach. *International Review of Environmental and Resource Economics*, **14**, 197–239.

Mihaylova B, Briggs A, O'Hagan A and Thompson SG (2011) Review of statistical methods for analysing healthcare resources and costs. *Health Economics*, **20**, 897–916.

Min Y and Agresti A (2002) Modelling nonnegative data with clumping at zero: A survey. *Journal of the Iranian Statistical Society*, **1**, 7–33.

——— (2005) Random effect models for repeated measures of zero-inflated count data. *Statistical Modelling*, **5**, 1–19.

Mullahy J (1986) Specification and testing of some modified count data models. *Journal of Econometrics*, **33**, 341–65.

Neelon B, O'Malley AJ and Smith VA (2016a) Modelling zero-modified count and semicontinuous data in health services research part 1: Background and overview. *Statistics in Medicine*, **35**, 5070–93.

——— (2016b) Modelling zero-modified count and semicontinuous data in health services research part 2: Case studies. *Statistics in Medicine*, **35**, 5094–5112.

Pan W and Shen X (2007) Penalized model-based clustering with application to variable selection. *Journal of Machine Learning Research*, **8**, 1145–64.

Pandey GR and Nguyen V-T-V (1999) A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology*, **225**, 92–101.

Petrella L, Laporta AG and Merlo L (2018) Cross-country assessment of systemic risk in the European Stock Market: Evidence from a CoVaR analysis. *Social Indicators Research*, **146**, 1–18.

Petrella L and Raponi V (2019) Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress. *Journal of Multivariate Analysis*, **173**, 70–84.

Portnoy S and Koenker R (1997) The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimators. *Statistical Science*, **12**, 279–300.

Rabe-Hesketh S, Skrondal A and Pickles A (2005) Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, **128**, 301–23.

Reich BJ, Fuentes M and Dunson DB (2011) Bayesian spatial quantile regression. *Journal of the American Statistical Association*, **106**, 6–20.

Rigby RA and Stasinopoulos DM (2005) Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**, 507–54.

Royston P and Altman DG (1994) Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **43**, 429–53.

Sauzet O, Razum O, Widera T and Brzoska P (2019) Two-part models and quantile regression for the analysis of survey data with a spike. The example of satisfaction with health care. *Frontiers in Public Health*, **7**, 146.

Schwarz G (1978) Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–64.

Sottile G, Frumento P, Chiodi M and Bottai M (2020) A penalized approach to covariate selection through quantile regression coefficient models. *Statistical Modelling*, **20**, 369–85.

Städler N, Bühlmann P and Van De Geer S (2010) ℓ1-penalization for mixture regression models. *Test*, **19**, 209–56.

Tian F, Gao J and Yang K (2018). A quantile regression approach to panel data analysis of health-care expenditure in organisation for economic co-operation and development countries. *Health Economics*, **27**, 1921–44.

Tian Y, Tang M and Tian M (2016) A class of finite mixture of quantile regressions with its

applications. *Journal of Applied Statistics*, **43**, 1240–52.

Tian Y, Tian M and Zhu Q (2014) Linear quantile regression based on EM algorithm. *Communications in Statistics-Theory and Methods*, **43**, 3464–84.

Tibshirani R (1996) Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B*, **58**, 267–88.

Waldmann E (2018) Quantile regression: A short story on how and why. *Statistical Modelling*, **18**, 203–18.

Wasserman L and Roeder K (2009) High dimensional variable selection. *Annals of Statistics*, **37**, 2178.

Winkelmann R (2004) Health care reform and the number of doctor visits: An econometric analysis. *Journal of Applied Econometrics*, **19**, 455–72.

Yi C and Huang J (2017). Semismooth Newton coordinate descent algorithm for elastic-net penalized Huber loss regression and quantile regression. *Journal of Computational and Graphical Statistics*, **26**, 547–57.

Yu K and Moyeed RA (2001) Bayesian quantile regression. *Statistics & Probability Letters*, **54**, 437–47.

Zhou X-H (2002) Inferences about population means of health care costs. *Statistical Methods in Medical Research*, **11**, 327–39.